ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/369764699

# Emerging Trends in Applied Mathematics And Statistics

**Book** · December 2022

CITATIONS
0

READS
92

**7 authors**, including:

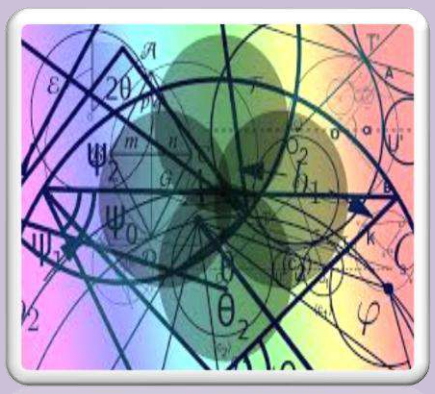
Prasad Gavali
Balwant College
**3** PUBLICATIONS **0** CITATIONS
SEE PROFILE

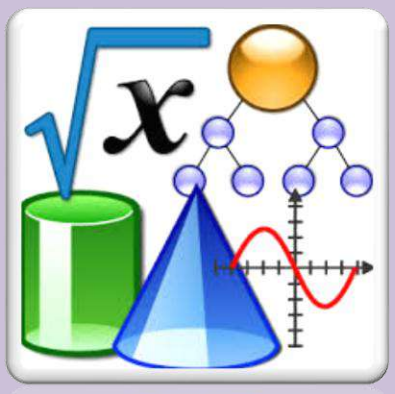
Annasaheb Suryawanshi
Balwant College Vita
**14** PUBLICATIONS **8** CITATIONS
SEE PROFILE

All content following this page was uploaded by Annasaheb Suryawanshi on 04 April 2023.

The user has requested enhancement of the downloaded file.

# EMERGING TRENDS IN APPLIED MATHEMATICS AND STATISTICS






EDITORIAL BOARD

Prin. Dr. Rajendra S. More

Mr. Nitin M. Jagtap

Mr. Prasad A. Gavali

Mr. Shrikant V. Gimawekar

Mr. Mahesh B. Pawar

Mrs. Amruta B. Hasabe

Dr. Annasaheb M. Suryawanshi



PRARUP PUBLICATION, KOLHAPUR

# Emerging Trends in Applied Mathematics and Statistics

**(ISBN: 978-81-956739-5-7)**

## *Editors*

**Prin. Dr. Rajendra S. More**
*Principal,*
*Balwant College, Vita*
*(M S), India-415 311*

**Mr. Nitin. M. Jagtap**
*Head & Assistant Professor,*
*Department of Mathematics,*
*Balwant College, Vita*
*(M S), India-415 311*

**Mr. Prasad. A. Gavali**
*Head &Assistant Professor,*
*Department of Statistics,*
*Balwant College, Vita*
*(M S), India-415 311*

**Mr. Shrikant. V. Gimawekar**
*Assistant Professor,*
*Department of Statistics,*
*Balwant College, Vita*
*(M S), India-415 311*

**Mr. Mahesh. B. Pawar**
*Assistant Professor,*
*Department of Mathematics,*
*Balwant College, Vita*
*(M S), India-415 311*

**Mrs. Amruta. B. Hasabe**
*Assistant Professor,*
*Department of Mathematics,*
*Balwant College, Vita*
*(M S), India-415 311*

**Dr. Annasaheb. M. Suryawanshi**
*Assistant Professor,*
*Department of Statistics,*
*Balwant College, Vita*
*(M S), India-415 311*

**Prarup Publication**

**2022**

*First Edition: December 2022*

*ISBN: 978-81-956739-5-7*


© **Copyright reserved by the Editors:** Publication, Distribution and Promotion Rights reserved by Prarup Publication, Kolhapur. No part of this publication may be reproduced in any form or by any means, electronically, mechanically, by photocopying, recording or otherwise, without the prior permission of the publishers. The views and results expressed in various articles are those of the authors and not of editors or publisher of the book. Balwant College Vita, Editors and Publishers remains neutral with regard to jurisdictional claims in published maps, institutional affiliations and other data.

**Published By:** Prarup Publication, Kolhapur, (M.S.), India

# PREFACE

We are delighted to publish our book entitled "Emerging Trends in Applied Mathematics and Statistics". This book is the compilation of esteemed articles of acknowledged experts in the various fields of Mathematics, Statistics and applied mathematics and Statistics providing a sufficient depth of the subject to satisfy the need of a level which will be comprehensive and interesting. It is an assemblage of variety of information about emerging trends in applied Mathematics and Statistics with its application oriented and interdisciplinary approach. We hope that the students, teachers, researchers and industrialists makers in India and abroad will find this book much more useful. Our special thanks and appreciation goes to experts and research scholars whose contributions have enriched this book. We thank our publisher, Prarup Publication, Kolhapur (M.S.), India for taking pains in bringing out the book. Finally, we will always remain a debtor to all our well-wishers for their blessings, without which this book would not have come into existence.

**- Editors**

# Table of Content

Contents available at www.balwantcollege.edu.in

**Emerging Trends in Applied Mathematics and Statistics**

***(Balwant College, Vita (M.S.) India 415 311)***

# Confidence Interval through MLE of Poisson type Length Biased Exponential class model

Rajesh Singh[1], Pritee Singh[2], and Preeti A. Badge[3*]

[1]*Department of Statistics, S. G. B. Amravati University, Amravati, India.*

[2] *Department of Statistics, S. G. B. Amravati University, Amravati, India.*

[3]*Department of Statistics, Institute of Science, Nagpur, India.*

**Abstract:** In this research paper, software reliability growth model (SRGM) is studied considering Poisson process of occurrence of software failures. Here, Length biased exponential class model is considered as a function with parameters i.e. total number of failures $\theta_0$ and scale parameter $\theta_1$. Assuming that no information is available about both the parameters, so using non informative priors, Confidence interval through MLE have been obtained for the parameters $\theta_0$ and $\theta_1$. To study the performance of confidence interval, average length and coverage probability are calculated using Monte Carlo simulation technique. From the obtained intervals it is concluded that Confidence interval of $\theta_0$ and $\theta_1$ perform better for appropriate choice of execution time and certain values of parameters.

**Key Words:** Length biased exponential distribution, Non informative prior, Software reliability growth model, Maximum likelihood estimator (MLE), Average length and coverage probability.

**1. Introduction:** This paper considered Poisson type length biased exponential class model as per Musa and Okumoto (1984) classification scheme. The exponential exponentiated distribution proposed by Gupta and Kundu (1999) which is special case of the exponentiated weibull family. Mudholkar and Shrivastava (1993) proposed the exponentiated weibull distribution as an extension of the weibull family obtained by adding the second shape parameter. Gupta and Keating (1986) developed relationship between the survival function, the failure rate and mean residual life of exponential distribution and its length biased form. Mudholkar et al (1996) applied exponentiated weibull distribution to serve survival data and showed those hazard rates are increasing, decreasing bathtub shape and unimodal. Seenoi et al (2014) proposed length biased exponentiated invented weibull distribution including some probability functions and moments of this distribution. Das and Kundu (2009) proposed weighted exponential distribution. Mir et al (2013) introduced a length biased Beta distribution and also given a test for detection of length biasedness of beta distribution. Neppala et al (2011) proposed Pareto type II based software reliability growth model with interval domain data using maximum likelihood estimation to estimate the parameter. Singh et al (2016) proposed Bayes estimators for length biased distribution compared with ML estimators.Singh and Singh (2012) have obtained Bayes estimates for the parameters of Poisson Exponential models considering the non-informative and informative prior.

***Corresponding Author:** Miss. Preeti A. Badge, Research Scholar, Department of Statistics, S.G.B Amravati University, Amravati, Maharashtra. **Email:** preetibadge10@gmail.com*

In the field of software reliability most of the work done on point estimation which give single guess value. Interval estimation with confidence interval gives more information than a point estimate. Confidence interval will be derived for both finite and infinite failure type models. By interval estimation we can give accurate range of values of the parameter. In this paper, the work of Singh and Singh (2016) regarding Bayes estimates for the parameters of Poisson Type Length Biased Exponential Class model have been applied for the case of Poisson Type of SRGMs to study the performance of length biased exponential model.

**2. Model Formulation and Evaluation:**

Consider that software is tested for its performance and observed the time of failure occurs during software system performance. Let the number of failures present in software be $\theta_0$,and $t_e$ be the execution time i.e. time during which CPU is busy and $m_e$ be the number of failures observed up to execution time $t_e$ .Consider time between the failures $t_i$ (i=1,2,……$m_e$ ) follows the exponential distribution with parameter $\theta_1$. The length biased exponential distribution is given as

$$f^*(t) = \begin{cases} t\theta_1^2 e^{-\theta_1 t} & ,t > 0, \theta_1 > 0, E[t] \neq 0 \\ 0 & otherwise \end{cases} \quad \ldots(1)$$

Where $f^*(t)$ denotes the length biased exponential distribution.

The failure intensity function is obtained by using equation (1)

$$\lambda(t) = \theta_0 t\theta_1^2 e^{-\theta_1 t} \qquad ,t > 0, \theta_0 > 0 \quad \ldots(2)$$

Where $\theta_0$ express the number of failures and $\theta_1$ express the for failure rate.

The mean failure function i.e. expected number of failures at time $t_e$ can be obtained by using equations (2)

$$\mu(t_e) = \theta_0 \theta_1^2 I_1 \quad \ldots(3)$$

Where, $I_1 = \int_0^{t_e} t_i e^{-\theta_1 t_i}\, dt$ and by solving (see Gradshteyn and Ryzhik (1994, p. 357)) we get,

$$\mu(t_e) = \theta_0 \left[1 - (1 + \theta_1 t_e) e^{-\theta_1 t_e}\right] \quad ,t > 0, \theta_0 > 0, \theta_1 > 0 \quad \ldots(4)$$

The study of behavior of failure intensity and expected number of failure of length biased exponential class model has been done by Singh et al (2016). They have compared the MLE's and Bayesian estimators on the basis of risk efficiencies.

**3. Maximum Likelihood Estimation:**

Maximum likelihood estimation is most preferable because of its easy computation. It requires likelihood function for estimation. The likelihood function of $(\theta_0, \theta_1)$ is obtained with the help of failure intensity (2) and expected number of failures (4) (see for details Lamino and Okumoto (1987)) given by:

$$L(\theta_0, \theta_1) = \theta_0^{m_e} \theta_1^{2m_e} \left[\prod_{i=1}^{m_e} t_i\right] e^{-T\theta_1} e^{-\theta_0 \left[1-(1+\theta_1 t_e)e^{-\theta_1 t_e}\right]} \quad \ldots(5)$$

Where,

$$\sum_{i=1}^{m_e} t_i = T$$

After taking the logarithm of both sides of above equation and applying the procedure of obtaining the MLE's for parameters $\theta_0$ and $\theta_1$, the MLE's are

$$\hat{\theta}_{m0} = \left[\frac{m_e}{\left(1-(1+\hat{\theta}_{m1} t_e)e^{-\hat{\theta}_{m1} t_e}\right)}\right] \quad \ldots(6)$$

and

$$\hat{\theta}_{m1} = \left[\frac{(2m_e - T\theta_{m1})e^{\hat{\theta}_{m1} t_e}}{\hat{\theta}_{m0} t_e^2}\right]^{1/2} \quad \ldots(7)$$

respectively.

The values of $\hat{\theta}_{m0}$ and $\hat{\theta}_{m1}$ can be obtained by solving simultaneous equations (6) and (7) using any available standard numerical method viz. Bisection Method. Singh et al (2016) obtained maximum likelihood estimates for parameters of length biased exponential model. They compared maximum likelihood estimates and Bayes estimates on the basis of risk efficiencies and concluded that Bayes estimates preferred over maximum likelihood estimates.

Now to obtain confidence interval for both the parameter, it requires variance-covariance matrix for Σ all the MLE. Variance-covariance matrix is derived using Fisher information matrix. For asymptotic variance we can calculate Fisher information matrix which is negative second partial derivative of log likelihood function.

$$\text{Var}(\hat{\theta}_0) = \frac{\hat{\theta}_0^2}{me} \quad \ldots.(8)$$

$$\text{Var}(\hat{\theta}_1) = [(1/2m_e + \theta_0\theta_1^2 e^{-\hat{\theta}_1 t_e} t_e^2 - \theta_0\theta_1^3 e^{-\hat{\theta}_1 t_e} t_e^3)] \quad \ldots.(9)$$

Using equation (8) confidence limits for parameter $\theta_0$ is given by:

$$\hat{\theta}_{0L} = \hat{\theta}_0 + z_\alpha (\hat{\theta}_0/(m_e)^{1/2}) \quad \ldots.(10)$$

$$\hat{\theta}_{0U} = \hat{\theta}_0 - z_\alpha (\hat{\theta}_0/(m_e)^{1/2}) \quad \ldots.(11)$$

For parameter $\theta_1$ confidence limits using equation (9) are given by:

$$\hat{\theta}_{1L} = \hat{\theta}_1 - Z_\alpha [(1/2m_e + \theta_0\theta_1^2 e^{-\hat{\theta}_1 t_e} t_e^2 - \theta_0\theta_1^3 e^{-\hat{\theta}_1 t_e} t_e^3)]^{1/2} \quad \ldots.(12)$$

$$\hat{\theta}_{1U} = \hat{\theta}_1 + Z_\alpha [(1/2m_e + \theta_0\theta_1^2 e^{-\hat{\theta}_1 t_e} t_e^2 - \theta_0\theta_1^3 e^{-\hat{\theta}_1 t_e} t_e^3)]^{1/2} \quad \ldots.(13)$$

By substituting tabulated values of $Z_\alpha$, 95% confidence interval can be obtained.

## 4. Discussion and Results:

Here, 95% confidence interval through MLE is obtained for the parameters $\theta_0$ i.e. total number of failures and $\theta_1$. Confidence interval for the proposed length biased exponential model is calculated as defined equation (1). To study the performance of confidence interval,

sample size $m_e$ was generated up to execution time $t_e$ and it was repeated 1000 times from the length biased exponential distribution for distinct values of $\theta_0$ and $\theta_1$. Using Monte Carlo simulation technique Average length and coverage probability is calculated for the respective confidence interval and presented in tabular form.

A number of values for the different parameters $\theta_0$ and $\theta_1$ involved in confidence interval were considered, but only for some of the considered values results have been reported in this chapter. Confidence intervals have been obtained for the 95% confidence level. The values of average length and coverage probability have been obtained by assuming execution time $t_e(= 15,20,25)$, and parameters$\theta_0(= 24(2)32)$, and$\theta_1(= 0.02(0.02)0.1)$. Average length and coverage probability obtained for confidence interval has been summarized in table (1) to (6).

Table (1) to (3) represents the 95% confidence interval for the parameter $\theta_0$. From the table, it is seen that as values of parameter $\theta_0$ increases the calculated average length decreases and there is slight decrease in average length as values of parameter $\theta_1$ increases. From table it can be seen that values of coverage probability decreases as $\theta_0$ increases and it remain constant as $\theta_1$ increases. As average length decreases it affects on coverage probability. Coverage probability decreases as average length decreases.

Table (4) to (6) represents the 95% confidence interval for the parameter $\theta_1$. It is seen that the values of average length increases as $\theta_0$ increases and average length increases slightly as $\theta_1$ increases. From table it can be seen that values of coverage probability increases as $\theta_0$ and $\theta_1$ increases.

**5. Conclusions:** Confidence interval through MLE is derived in this paper, the confidence interval suggested for the parameters of Poisson type length biased exponential class as SRGM. From the computation and above discussion it is concluded that confidence interval for the parameters $\theta_0$ and $\theta_1$, gives shorter average length for particular values of both the parameters for fixed execution time. And it is also concluded that proposed confidence interval maintained coverage probability for different values of parameters for fixed execution time. With increases in execution time confidence interval gives shortest average length and but coverage probability decreases. Confidence interval can prefer for both the parameters.

**Table 1: Average length and coverage probability of 95% confidence interval $\widehat{\theta}_{M0}$ calculated for the parameters $\theta_0$ = (24:2:32) and $\theta_1$ = (0.02:0.02:0.1) and $t_e$ = 15**

| $\theta_1$ \ $\theta_0$ | 24 | 26 | 28 | 30 | 32 |
|---|---|---|---|---|---|
| 0.02 | 57.222518 (0.995) | 56.50778 (0.995) | 55.71389 (0.994) | 54.93576 0.994) | 54.27069 (0.993) |
| 0.04 | 57.243652 (0.995) | 56.45414 (0.995) | 55.67368 (0.994) | 54.92776 (0.994) | 54.23347 (0.993) |
| 0.06 | 57.233046 (0.995) | 56.42863 (0.995) | 55.65794 (0.994) | 54.90752 (0.994) | 54.16633 (0.993) |
| 0.08 | 57.233615 (0.995) | 56.39313 (0.995) | 55.63409 (0.994) | 54.70139 (0.994) | 54.12473 (0.993) |
| 0.1 | 57.16894 (0.995) | 56.38039 (0.995) | 55.55027 (0.994) | 54.65604 (0.994) | 53.83833 (0.993) |

**Table 2: Average length and coverage probability of 95% confidence interval $\widehat{\theta}_{M0}$ calculated for the parameters $\theta_0$ = (24:2:32) and $\theta_1$ = (0.02:0.02:0.1) and $t_e$ = 20**

| $\theta_1$ \ $\theta_0$ | 24 | 26 | 28 | 30 | 32 |
|---|---|---|---|---|---|
| 0.02 | 57.728669 (0.996) | 57.40079 (0.995) | 57.11992 (0.995) | 56.78598 (0.994) | 56.50296 (0.993) |
| 0.04 | 57.725421 (0.996) | 57.38959 (0.995) | 57.09107 (0.995) | 56.77883 (0.994) | 56.48683 (0.993) |
| 0.06 | 57.70429 (0.996) | 57.38709 (0.995) | 57.07483 (0.995) | 56.77586 (0.994) | 56.47019 (0.993) |
| 0.08 | 57.694764 (0.996) | 57.38601 (0.995) | 57.05328 (0.995) | 56.73697 (0.994) | 56.42135 (0.993) |
| 0.1 | 57.687913 (0.996) | 57.35914 (0.995) | 57.05003 (0.995) | 56.72293 (0.994) | 56.39379 (0.992) |

**Table 3: Average length and coverage probability of 95% confidence interval $\widehat{\theta}_{M0}$ calculated for the parameters $\theta_0$ = (24:2:32) and $\theta_1$ = (0.02:0.02:0.1) and $t_e$ = 25**

| $\theta_1$ \ $\theta_0$ | 24 | 26 | 28 | 30 | 32 |
|---|---|---|---|---|---|
| **0.02** | 57.633128 (0.995) | 57.26667 (0.995) | 56.90905 (0.995) | 56.53445 (0.994) | 56.18467 (0.993) |
| **0.04** | 57.621798 (0.995) | 57.2444 (0.995) | 56.89678 (0.995) | 56.46993 (0.994) | 56.09547 (0.993) |
| **0.06** | 57.619242 (0.995) | 57.23135 (0.995) | 56.85961 (0.995) | 56.44942 (0.994) | 56.07487 (0.993) |
| **0.08** | 57.612651 (0.995) | 57.21467 (0.995) | 56.84414 (0.995) | 56.41711 (0.994) | 56.02504 (0.992) |
| **0.1** | 57.605773 (0.995) | 57.18747 (0.995) | 56.78515 (0.995) | 56.38402 (0.994) | 55.94543 (0.992) |

**Table 4: Average length and coverage probability of 95% confidence interval $\widehat{\theta}_{M1}$ calculated for the parameters $\theta_0$ = (24:2:32) and $\theta_1$ = (0.02:0.02:0.1) and $t_e$ = 15**

| $\theta_1$ \ $\theta_0$ | 24 | 26 | 28 | 30 | 32 |
|---|---|---|---|---|---|
| **0.02** | 0.023096 (0.997) | 0.023270 (0.998) | 0.023425 (0.999) | 0.023611 (0.999) | 0.023803 (0.999) |
| **0.04** | 0.023100 (0.997) | 0.023273 (0.998) | 0.023426 (0.999) | 0.023648 (0.999) | 0.023840 (0.999) |
| **0.06** | 0.023103 (0.997) | 0.023279 (0.998) | 0.023448 (0.999) | 0.023651 (0.999) | 0.023856 (0.999) |
| **0.08** | 0.023103 (0.997) | 0.023297 (0.998) | 0.023461 (0.999) | 0.023671 (0.999) | 0.023863 (0.999) |
| **0.1** | 0.023107 (0.997) | 0.023298 (0.999) | 0.023488 (0.999) | 0.023691 (0.999) | 0.023871 (0.999) |

**Table 5: Average length and coverage probability of 95% confidence interval $\widehat{\theta}_{M1}$ calculated for the parameters $\theta_0$ = (24:2:32) and $\theta_1$ = (0.02:0.02:0.1) and $t_e$ = 20**

| $\theta_1$ \ $\theta_0$ | 24 | 26 | 28 | 30 | 32 |
|---|---|---|---|---|---|
| 0.02 | 0.019317 (0.996) | 0.019425 (0.996) | 0.019531 (0.997) | 0.019644 (0.997) | 0.019751 (0.997) |
| 0.04 | 0.019321 (0.996) | 0.019431 (0.996) | 0.019535 (0.997) | 0.019664 (0.997) | 0.019779 (0.997) |
| 0.06 | 0.019322 (0.996) | 0.019435 (0.996) | 0.019546 (0.997) | 0.019668 (0.997) | 0.019786 (0.997) |
| 0.08 | 0.019323 (0.996) | 0.01944 (0.996) | 0.019551 (0.997) | 0.019669 (0.997) | 0.019802 (0.997) |
| 0.1 | 0.019325 (0.996) | 0.019448 (0.996) | 0.019568 (0.997) | 0.019670 (0.997) | 0.019826 (0.997) |

**Table 6: Average length and coverage probability of 95% confidence interval $\widehat{\theta}_{M1}$ calculated for the parameters $\theta_0$ = (24:2:32) and $\theta_1$ = (0.02:0.02:0.1) and $t_e$ = 25**

| $\theta_1$ \ $\theta_0$ | 24 | 26 | 28 | 30 | 32 |
|---|---|---|---|---|---|
| 0.02 | 0.016918 (0.995) | 0.017003 (0.996) | 0.017076 (0.996) | 0.017164 (0.996) | 0.017241 (0.997) |
| 0.04 | 0.016919 (0.995) | 0.017005 (0.996) | 0.017083 (0.996) | 0.017166 (0.996) | 0.017243 (0.997) |
| 0.06 | 0.016924 (0.995) | 0.017006 (0.996) | 0.017088 (0.996) | 0.017167 (0.996) | 0.017245 (0.997) |
| 0.08 | 0.016927 (0.995) | 0.017007 (0.996) | 0.017093 (0.996) | 0.017177 (0.996) | 0.017249 (0.997) |
| 0.1 | 0.016928 (0.995) | 0.017013 (0.996) | 0.017094 (0.996) | 0.017181 (0.996) | 0.017262 (0.997) |

**References:**


1. Chen M. H., Ibrahim J.G. and Kim S.(2008) "Properties and Implementation of Jeffrey's Prior in Binomial Regression Models", Journal of the American Statistical Association, Vol.103, No. 484,pp.1959-1664.
2. Fitrilia A., Fithriani I. and Nurrohmah1 S. (2018). "Parameter estimation for the Lomax distribution using the E Bayesian method", Journal of Physics,1108,https://doi:10.1088/1742-6596/1108/1/012018
3. Fisher R. A. (1934). " The effects of methods of ascertainment upon the estimation of frequencies, Ann. Eugenics, 6, p.13-25.
4. Gradshteyn I. S. and Ryzhik I. M. (1994). " Table of Integrals, Series, and Products", Alan Jeffrey (editor) 5th Ed., New York, Academic Press.
5. Gupta R. C. and Keating J. P. (1986). "Relations for reliability measures under length biased sampling'', Scand Journal of Statistics, 13, p. 49-56.
6. Martz H.F.and Waller R.A.(1982). "Bayesian Reliability Analysis," New York Wiley.
7. Musa J.D.and Okumoto K.(1984) "A logarithmic Poisson execution time model for software reliability measurement," Proc. 7th International Conference on Software Engineering, Orlando, Florida, p. 230–238.
8. Musa J. D., Iannino A. and Okumoto K. (1987). Software Reliability: Measurement, Prediction, Application, New York, McGraw-Hill.
9. Patil G.P. And Rao C.R.(1978) "Weighted distributions and Size biased Sampling with Applications to Wildlife Populations and Human Families" Biometrics 34, p.179-189.
10. Rabie Abdalla and Junping Li (2019) "E-Bayesian Estimation Based on Burr-X Generalized Type-II Hybrid Censored Data" Symmetry 2019, 11, 626; https://doi.org/10.3390/sym11050626.
11. Rao K.A. and D'cunha J.G (2014). "Bayesian Inference for Mean of the Lognormal Distribution" International Journal of Scientific and Research Publications, Volume 4, Issue 10, p. 195-203.
12. Shreshtha S.K. and Kumar V. (2014) "Bayesian Analysis for the Generalized Rayleigh Distribution" International Journal of Statistika and Mathematika, Vol. 9, Issue 3,p. 118-131.
13. Singh R. Singh P. and Kale K.(2016) "Bayes Estimators for the parameters of Poisson type length biased exponential class model using non- informative priors" Journal of Reliability and Statistical Studies, Vol. 9, Issue 1, p. 21-28.
14. Tamak J.(2013) "Use of software reliability of growth model to estimate the reliability of web applications" International Journal of Advanced Research in Computer and Software Engineering vol.3,Iss.6,p.53-59.
15. Xie M. Hong and Wohlin C. (1997). "A practical method for the Estimation of Software Reliability Growth in the Early Stage of Testing," Proc. IEEE 7th International Symposium on software Reliability Engineering, Albuquerque, p.116-123.

Contents available at www.balwantcollege.edu.in
**Emerging Trends in Applied Mathematics and Statistics**
***(Balwant College, Vita (M.S.) India 415 311)***

# Performance Evaluation of selected Banking Stocks during Covid-19 Crises

Prakash Rajaram Chavan[1*]
[1]*Department of Statistics, Smt. Kasturbai Walchand College, Sangli*

**Abstract:** Stock market plays an important role in consolidation in national economy and also helps in the development in the industrial sector. It helps in mobilizing the saving and ensures safety. As India is a developing economy, the study of Indian stock market and its analysis find it prominence in national development. In this paper we discuss about basics of stock market, investment, various aspects of stock market, why people in India hesitate to investing stock market, basic terminologies used in the stock market, finance sector and Indian banking system. Finally this paper analyzes the data by using Auto Regressive Integrated Moving Average (ARIMA) method and concluding remarks.

**Key Words:** Stock, investment, banks, statistical analysis.

## 1. Introduction:

### 1.1 Investments

An investment is an asset or item acquired with the goal of generating income or appreciation. Appreciation refers to an increase in the value of an asset over time. When an individual purchases a good as an investment, the intent is not to consume the good but rather to use it in the future to create wealth. Investments are generally bucketed into three major categories: stocks, bonds and cash equivalents.

The different types of investments that exist are

**Stocks:** A stock is an investment in a specific company. When you purchase a stock, you're buying a share a small piece of that company's earnings and assets. Companies sell shares of stock in their businesses to raise cash; investors can then buy and sell those shares among themselves.

### 1.2 Stock Market

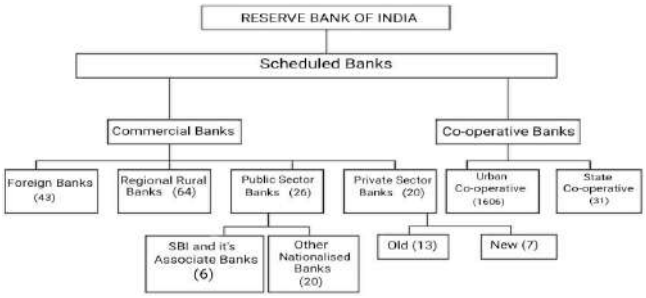
**Definition:**

The "stock market" is the term given to activity that takes place on stock exchanges. These are the physical locations where corporate shares trade hands. You usually refer to the stock market when you address the combined activity on many exchanges. For example, in the United States, the "stock market" is a combination of the New York Stock Exchange, the Nasdaq, the Chicago Stock Exchange and others. Some people refer to the "global markets" or the "world stock market." These refer to the overall activity on stock exchanges throughout the world.

***Corresponding Author:** *Dr. Prakash Rajaram Chavan, Head & Assistant Professor, Department of Statistics, Smt. Kasturbai Walchand College, Sangli, (Maharashtra), India-416 416*
***Email:** prchavan83@gmail.com*

## 2. Indian Banking System:

The Structure of Banking in India instigated in the last decades of 18th century. Indian Banking Structure is very different from that of other countries. Banking can be considered as the backbone of the country, its business, economic growth, and development because finance is the vital element for a country's trade, commerce, and industry. The role that banking system plays in a country can be quite but it is the most significant part of the country's economy.




## 3. Aims and Objective:

We aim in carrying out a study on analysis of Indian Stock market based on the data of share prices of 5 banks in India. Some of the objectives of our study are

1. to check the effect of COVID-19 on Stock market,
2. to forecast monthly high price,
3. to predict the future trend, and
4. to examine volatility associated with the stock price.

## 4. Data collection:

Collection of data required for the analysis. Also, we furnish the information regarding the sources where we can easily get the data. We have collected the data of leading 5 banks in India.

### 4.1 Sources of data

The pertained data on high price and closing prices at daily time step were collected from the secondary source viz. NSE official website (https://www1.nseindia.com). The daily share prices of abovementioned banks were taken for a period between January 2016 to January 2020. For the study we have taken 5 banks on National Stock Exchange namely, IDBI, HDFC, Panjab National bank, Axis bank, Bank f Baroda. The five banks were selected on the basis of their performance in the NSE.

### 4.2 Data on Stock market

Now we present the sample data which we used for analysis purpose. The below table contains daily data of 'high price' and 'close price' of 5 banks.

| | IDBI Bank | | Punjab National Bank | | HDFC bank | | Axis Bank | | Bank of Baroda | |
|---|---|---|---|---|---|---|---|---|---|---|
| Date | High Price (Rs.) | Close Price (Rs.) | High Price (Rs.) | Close Price (Rs.) | High Price (Rs.) | Close Price (Rs.) | High Price (Rs.) | Close Price (Rs.) | High Price (Rs.) | Close Price (Rs.) |
| 1-Jan-16 | 91.2 | 89.9 | 117.9 | 117.6 | 1266.9 | 1258.45 | 452.9 | 449.9 | 159.9 | 159.05 |
| 4-Jan-16 | 89.6 | 84.85 | 117 | 112.8 | 1253.9 | 1216.7 | 448.9 | 438.4 | 158.8 | 151.2 |
| 5-Jan-16 | 85.8 | 83.9 | 113.35 | 110.6 | 1233.45 | 1209.4 | 439.9 | 436.45 | 152.7 | 148.1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 6-Jan-16 | 85.05 | 82.9 | 111.8 | 109.6 | 1220.75 | 1209.3 | 439.3 | 430.7 | 150.95 | 148.7 |
| 7-Jan-16 | 81.35 | 77.7 | 108.45 | 105.05 | 1203.55 | 1179.45 | 425.15 | 409.25 | 147 | 139.5 |
| 8-Jan-16 | 79.85 | 78.05 | 106.75 | 105.1 | 1230 | 1174.4 | 418.75 | 413.7 | 143.45 | 140.45 |
| 11-Jan-16 | 77.8 | 76.55 | 104.25 | 102.9 | 1171 | 1161.55 | 423.4 | 417.2 | 140.2 | 139.1 |
| 12-Jan-16 | 77.4 | 73.75 | 103.5 | 99.65 | 1169.4 | 1151.85 | 420.65 | 406.1 | 140.4 | 134.5 |
| 13-Jan-16 | 75.3 | 70.3 | 101.2 | 98 | 1173.65 | 1167.65 | 411.45 | 406.7 | 137.4 | 132.1 |
| 14-Jan-16 | 68.4 | 66.1 | 98.3 | 97.8 | 1167 | 1159.3 | 402 | 390.6 | 131.85 | 130.9 |
| 15-Jan-16 | 66.6 | 60.8 | 98.8 | 92.2 | 1167.3 | 1149.8 | 395.85 | 374.25 | 133 | 126.15 |
| 18-Jan-16 | 61.05 | 55.55 | 94.2 | 90.85 | 1157.7 | 1131.8 | 383.35 | 373.2 | 131.5 | 125.7 |
| 19-Jan-16 | 58.5 | 57.9 | 93.4 | 92.45 | 1157.15 | 1153 | 396 | 393 | 129.2 | 125.15 |
| 20-Jan-16 | 56.5 | 55.25 | 90.5 | 87.25 | 1198.1 | 1136.65 | 393.4 | 387.6 | 123 | 121.15 |
| 21-Jan-16 | 58.3 | 57.7 | 90.35 | 88.6 | 1147.7 | 1131.1 | 412.45 | 409.15 | 126.5 | 124.25 |
| 22-Jan-16 | 60.95 | 60.1 | 93 | 92.5 | 1163.2 | 1158.45 | 427.65 | 424.15 | 131.6 | 130.2 |
| 25-Jan-16 | 61.4 | 59.1 | 94.25 | 91.85 | 1183.4 | 1174.6 | 429.75 | 420.85 | 133.15 | 129.5 |

We can read the above data as, for 15$^{th}$ January, 2016, the high price of the day was 66.6 rupees and the close price was 60.8 rupees when market closed at 3 PM for IDBI bank, similarly for HDFC bank, high price of the day was 98.8 rupees and the close price was 92.2 rupees, for Punjab National Bank, high price of the day was 1167.3 rupees and the close price was 1149.8 rupees, for Axis bank, high price of the day was 395.85 rupees and the close price was 374.25 rupees and for Bank of Baroda, high price of the day was 133 rupees and the close price was 126.15 rupees.

Here we are presenting the latest data close and high price of stock market for all 5 banks namely, IDBI, HDFC, Punjab national bank, Axis bank, Bank of Baroda.

| | IDBI | | HDFC | | Punjab national bank | | Axis bank | | Bank of Baroda | |
|---|---|---|---|---|---|---|---|---|---|---|
| Date | High Price | Close Price | High Price | Close Price | High Price | Close Price | High Price | Close Price | High Price | Close Price |
| 1-Jul-21 | 38.25 | 37.75 | 2491.35 | 2459.4 | 42.75 | 42.35 | 754.4 | 746.45 | 87 | 86.05 |
| 2-Jul-21 | 39.2 | 38.25 | 2480 | 2475.7 | 42.7 | 42.1 | 752.85 | 750.65 | 86.45 | 85.2 |
| 5-Jul-21 | 38.65 | 37.95 | 2496.8 | 2494.9 | 42.45 | 42.3 | 762 | 760.35 | 86.8 | 86.35 |
| 6-Jul-21 | 38.1 | 37.7 | 2518.9 | 2496.4 | 42.85 | 41.85 | 767 | 758.35 | 87.2 | 85.5 |
| 7-Jul-21 | 37.95 | 37.75 | 2545.8 | 2529.2 | 42.05 | 41.7 | 764 | 762.9 | 86.3 | 86.05 |
| 8-Jul-21 | 38.45 | 37.6 | 2543.85 | 2512.05 | 42.25 | 41.05 | 767.15 | 754.4 | 87 | 83.5 |
| 9-Jul-21 | 39.35 | 38.8 | 2509 | 2496.45 | 41.2 | 41 | 752.9 | 747.35 | 84.15 | 83.65 |
| 12-Jul-21 | 40 | 38.75 | 2511 | 2476.95 | 41.45 | 40.9 | 757.8 | 753.6 | 84.65 | 83.3 |
| 13-Jul-21 | 39.1 | 38.55 | 2549.95 | 2544.9 | 41.25 | 40.9 | 773 | 770.75 | 84.9 | 84.7 |
| 14-Jul-21 | 38.7 | 38.2 | 2544.9 | 2531.1 | 41.1 | 40.7 | 773.45 | 770.05 | 85.15 | 84.25 |
| 15-Jul-21 | 38.4 | 37.95 | 2555 | 2540.1 | 40.9 | 40.35 | 778.25 | 774.15 | 84.85 | 83.7 |

## 5. Statistical Analysis:

### 5.1 Auto Regressive Integrated Moving Average (ARIMA)

ARIMA, short for 'Auto Regressive Integrated Moving Average' is a Statistical tool which is actually a class of models that explains a given time series based on its own past values, that is its own lags so that equation can be used to forecast future values.
An ARIMA models contains 3 parameters: p, d, q where,
p is the order of the AR term (Auto Regressive).
q is the order of the MA term (Moving Average).
d is the number of differencing required to make a time series stationary.
Term Auto Regressive means it is a linear regression model that uses its own lags as predictors. It assumes that future will resemble the past. A pure Auto Regressive (AR only) model is one where $Y_t$depends only on its own lags. That is, $Y_t$is the function of the lags of $Y_t$.

$Y_t = \alpha+\beta_1 Y_{t-1}+\beta_2 Y_{t-2}+\ldots\ldots+\beta_p Y_{t-p}+ \varepsilon_1$

An ARIMA model is one where the time series was differenced once to make it stationary and you combine the AR and MA terms. So the equation becomes:

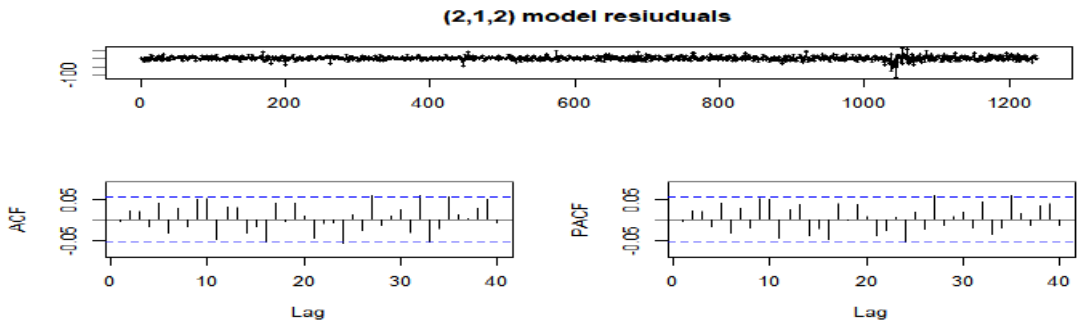
$Y_t = \alpha+\beta_1 Y_{t-1}+\beta_2 Y_{t-2}+\ldots\ldots+\beta_p Y_{t-p}\varepsilon_1+ \varphi_1 \varepsilon_{t-1}+ \varphi_2 \varepsilon_{t-2}+\ldots\ldots+ \varphi_q \varepsilon_{t-q}$

The ARIMA model assume that the time series is stationary.

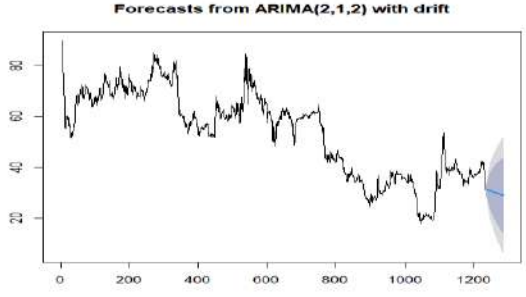
### 5.2 ARIMA forecasting

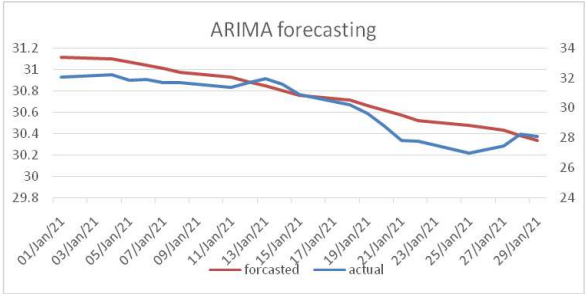
In this section, we use the ARIMA model to forecast future stock prices using various function available in R programming language. We also cross checked our forecasted results with actual crossing prices by plotting graph.

### a) IDBI bank:



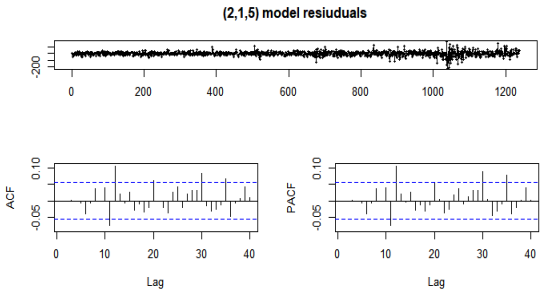

We use Augmented Dickey-Fuller Test to check the stationarity of the data.
Null hypoothesis: not stationary Level of significance(α)=0.05
Dickey-Fuller = -3.0987, Lag order = 10, p-value = 0.1132
alternative hypothesis: stationary p-value >α
so we do not reject the null hypothesis.The series is not stationary

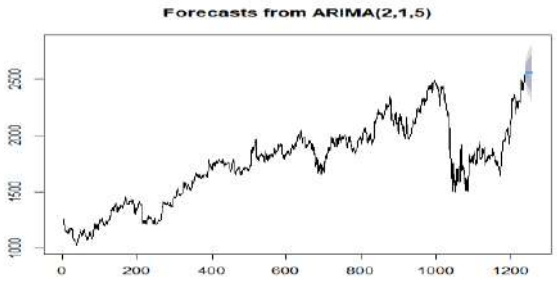


**Interpretation:** from both forecasted and actual values we can see that from 1st of January to till end of the month the stock price values is decreasing.

**b) HDFC bank:**



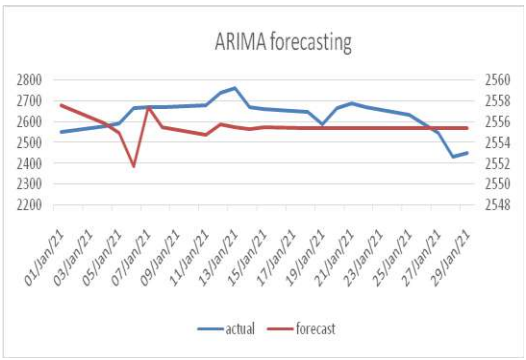



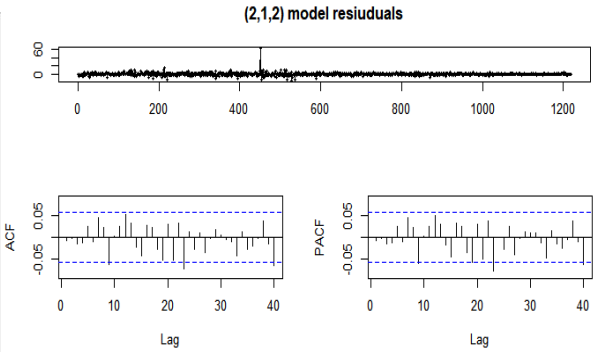

We use Augmented Dickey-Fuller Test to check the stationarity of the data.

Null hypoothesis: not stationary

Dickey-Fuller = -2.4387, Lag order = 10, p-value = 0.3926 alternative hypothesis: stationary

p-value >α so we do not reject the null hypothesis. The series is not stationary




**Interpretation:** From the above plot we can see that the prices are slightly increased in the beginning of the month but after 20th of January the prices are decreasing.

**c) Panjab national bank**

We use Augmented Dickey-Fuller Test to check the stationarity of the data.

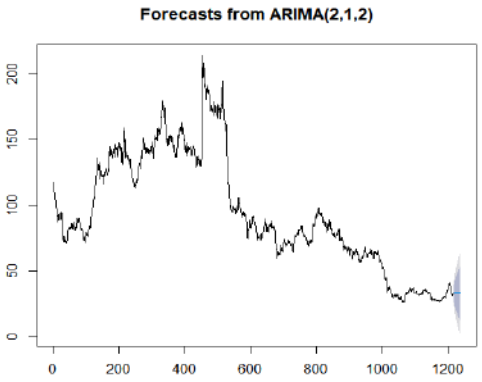
Null hypoothesis: not stationary

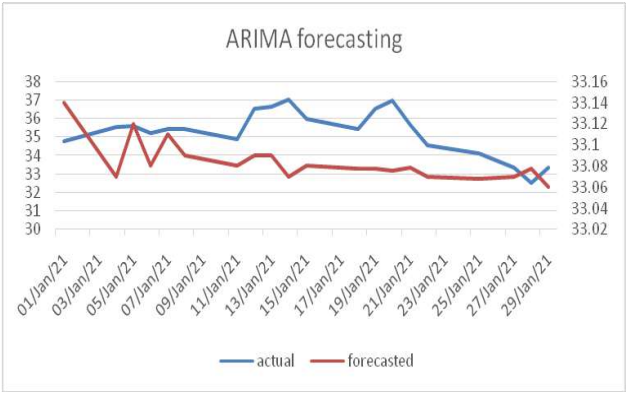
Level of significance(α)=0.05

Dickey-Fuller = -2.3906, Lag order = 10, p-value = 0.413

alternative hypothesis: stationary p-value >α

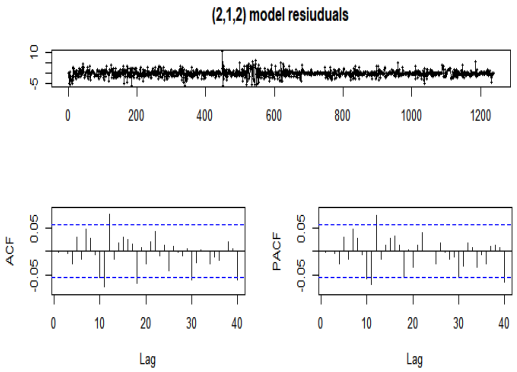
so we do not reject the null hypothesis. The series is not stationary

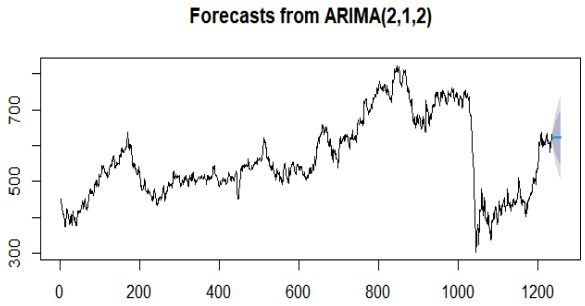


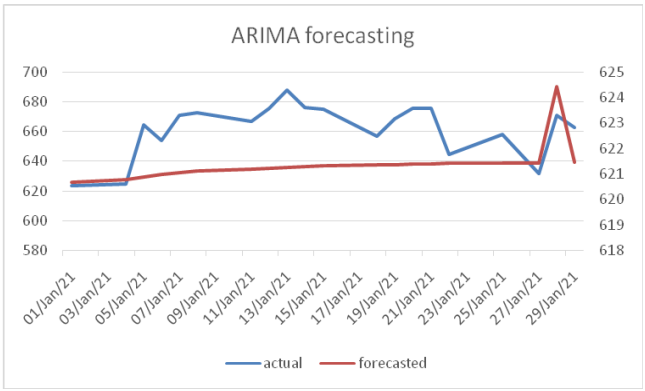
**Interpretation:** from the above plot we can see that the prices are not either increasing nor decreasing.

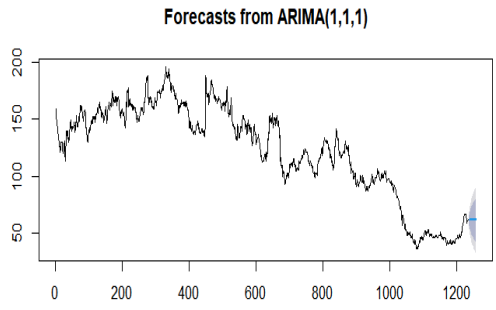
### d) Axis bank:



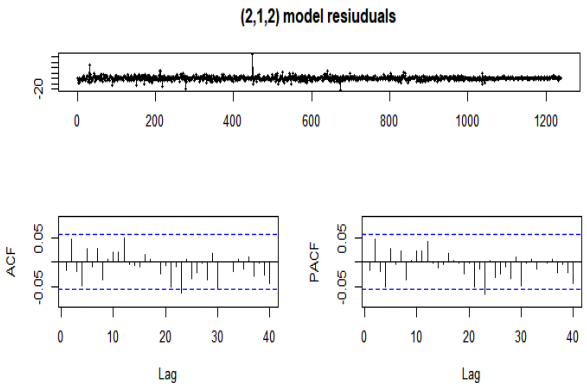

We use Augmented Dickey-Fuller Test to check the stationarity of the data.
Null hypoothesis: not stationary Level of significance(α)=0.05
Dickey-Fuller = -2.52, Lag order = 10, p-value = 0.3582 alternative hypothesis: stationary
p-value >α so we do not reject the null hypothesis. The series is not stationary

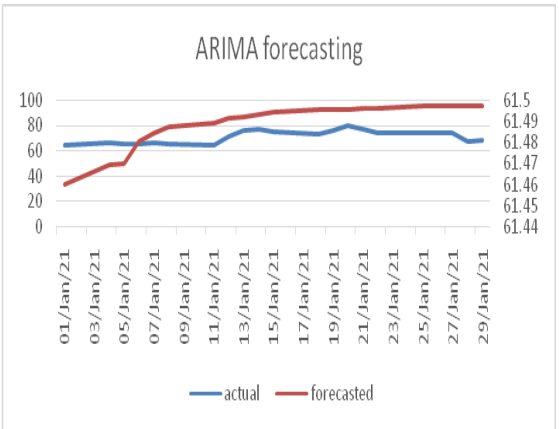



**Interpretation:** above plot shows that the stock prices are increased in the beginning of the month but decreased at the end of the month.

### e) Bank of Baroda:




We use Augmented Dickey-Fuller Test to check the stationarity of the data.
Null hypoothesis: not stationary Level of significance(α) = 0.05
Dickey-Fuller = -3.2474, Lag order = 10, p-value = 0.07975

alternative hypothesis: stationary p-value >α
so we do not reject the null hypothesis. The series is not stationary

**Interpretation:** The stock prices increased in the middle of the month and then gradually decreased at the end of the month.

## 6. Concluding Remarks:

We record our conclusions based on analysis carried out in this paper.

- We have seen that, for all the 5 banks considered namely, IDBI bank, HDFC bank, Punjab National Bank, Axis bank and Bank of Baroda, COVID-19 has not significantly affected the stock prices.
- In the above analysis, we obtained best ARIMA model, demonstrated the potential of ARIMA models to predict stock prices satisfactorily on short time basis.
- ARIMA helps the investors to make profitable investment decisions.

## References:

1. B.Sc. Books (Nirali and Phadke publication)
2. Fundamentals of Mathematical Statistics (S.C Gupta, V. K Kapoor)
3. Ahmed, S. (2008), "Aggregate Economic Variables and Stock Markets in India", International Research Journal of Finance and Economics, 14(1):141-164.
4. Naved, Mohd (2015), "The profitability of five popular variations of moving averages on Indian market index S&P CNX Nifty 50 During January 2004-December. 2014", Advances in Economics and Business Management (AEBM), 2, 27-32.
5. Arize, A., Kallianotis, L. N., Liu, S., Malindretos, J. & Maruffi, B. L. (2014), "The Preponderance of Stock Picking Techniques: The Practice of Applied Money Managers", Accounting and Finance Research, 3(2): 87.
6. Han, Y., Yang, K., & Zhou, G. (2013), "A New Anomaly: The Cross Sectional Profitability of Technical Analysis", Journal of Financial and Quantitative Analysis, 48(1), 1433-1461.
7. Kakani, R., K. & Sundhar, S. (2006), "Profiting from Technical Analysis in Indian Equity Markets: Using Moving Averages", XLRI Jamshedpur School of Business Working Paper No. 06-02.

Contents available at www.balwantcollege.edu.in

**Emerging Trends in Applied Mathematics and Statistics**

***(Balwant College, Vita (M.S.) India 415 311)***

# Applications of Survival Analysis to study the waiting time to get first Employment of Degree Students

Abhijeet Jadhav[1*], Aniruddh Bidkar[2], Pallavi Sakhare[3], Purva Babar[4]

*[1,2,3,4] PG Student, Department of Statistics, Balwant College, Vita*

**Abstract:** The goal of present study was to compare survival probabilities of waiting time to get first employment of degree students from engineering with respect to categorical covariates gender and trade of engineering students and to fit Cox Proportional regression model. The Kaplan Meier method was used to compare the survival probabilities. The Kaplan Meier result shows that, survival distribution of waiting time to get first employment are same in male and female but different in ten categories of tread of engineering degree. The covariates age, trade of engineering, years to complete degree, CGPA were significant in Cox Proportional hazard model.

**Keywords:** Unemployment, Kaplan Meier, Cox-Proportional, Log-Rank Test, Hazard.

**1. Introduction:** Immediately after degree getting a job is a challenging task for Fresher's in engineering field. Year by year the number of degree holders in engineering from universities has been increasing due to the large number of students joining higher in education and the highly expansion programs of higher education. Delayed employment or unemployment is attributed to the lack of soft or nontechnical skills of graduates in engineering filed & shortage of capital to create their own business. Besides factors such as the reputation of higher education institutions. The capacity of higher education to provide consultancy services, mismatch of skills between fresher's & employer's demands affect fresher's employment. Individual factors like, discipline type, achievements, gender, residence, family background, and fresher job hunting skills influences on engineering students' employment.

Unemployment affects not only the unemployed but also family members and society at large. The social and political consequences of large unemployment, especially among youth educated be serious person can. As a result, the issue of fresher's unemployment is becoming a fundamental issue that draws the attention of scholars. Nevertheless, the aforementioned studies were either attributed to fresher's in a specific field of study, which lacked inclusiveness or used inadequate statistical method to explore the factors for fresher's unemployment. G. Ayaneh et al. (2020) developed survival models for the analysis of waiting time to first employment of new graduates. Their study was related to new graduates belong to Debre Markos University Graduates, Northwest Ethiopia. I. T. Jayamanne et al. (2017) studied on the waiting time for the first employment of arts graduates who are from Sri Lanka. M. Salas-Velasc (2007) has analyzed the transition from higher education to employment in Europe. Their analysis is also related to the time to obtain the first job.

***Corresponding Author:** *Mr. Abhijeet Jadhav, PG Student, Department of Statistics, Balwant College, Vita (M.S.) India 415 311. Email: abhijeetjadhav807@gmail.com*

M. A. Pourhoseingholi et al. (2007) compared the Cox Regression and parametric models for survival of patients with gastric carcinoma.

In present study we want to fit Cox Proportional regression model on waiting time to get first employment of degree students from engineering with respect to various covariates like age, gender, trade of engineering, years to complete degree, CGPA, Extra courses completed during the degree etc. We also want to compare the survival probabilities of waiting time to get first employment of degree students from engineering with respect to categorical covariates gender and trade of engineering students.

## 2. Methods and material:

The primary data collected from engineering degree students who have completed their graduate degree from different colleges of Maharashtra State. The data collection has been done through the GOOGLE form. The survival techniques, namely, Kaplan Meier and Cox Proportional regression were used to compare the survival probabilities of waiting time to get first employment of degree students between categorical covariates such as gender, trade of engineering students. The statistical software SPSS was used for data analysis.

## 3. Statistical analysis:

To study the survival analysis of waiting time to get first employment of degree students from engineering. We selected 163 graduate students from different treads of engineering, namely, Civil, Electrical, Mechanical, Chemical, Electronics, Automobile, ENTC, Textile, IT and Computer Science. Out 163 degree students, there are 110 male and 53 female students.

### 3.1. Kaplan Meier:

A descriptive procedure used to examining the distribution of time-to-event variables is known as Kaplan-Meier survival analysis. In Kaplan-Meier Survival Analysis we can compare the survival distribution by levels of a factor variable.

**(i) To compare waiting time vs gender:**

The null and alternative hypothesis for Log-Rank test:

*$H_0$: Survival distributions of waiting time to get first employment are same in the male and female students.*

*$H_1$: Survival distributions of waiting time to get first employment are different in the male and female students.*

The means and medians for survival time (Table -3.1.1) gives a numerical comparison of the waiting time to get first employment for male and female students. It includes estimate of mean waiting survival time and their standard error with 95% confidence interval for male and female. It also includes, estimate of median waiting survival time and their Standard error with 95% confidence interval for male and female.

**Table-3.1.1: Means and Medians for Survival Time**

| Gender | Mean | | | | Median | | | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | Std. Error | 95% Confidence Interval | | Estimate | Std. Error | 95% Confidence Interval | |
| | | | Lower Bound | Upper Bound | | | Lower Bound | Upper Bound |
| Female | 29.000 | 3.843 | 21.467 | 36.533 | 19.267 | 1.090 | 17.130 | 21.403 |
| Male | 31.907 | 2.898 | 26.228 | 37.586 | 19.200 | 1.615 | 16.036 | 22.364 |
| Overall | 30.846 | 2.318 | 26.303 | 35.389 | 19.267 | 1.148 | 17.016 | 21.517 |

The log rank test is used to test the null hypothesis of no difference in survival between two or more independent groups. The test compares the entire survival experience between groups and can be thought of as a test of whether the survival curves are identical (overlapping) or not.

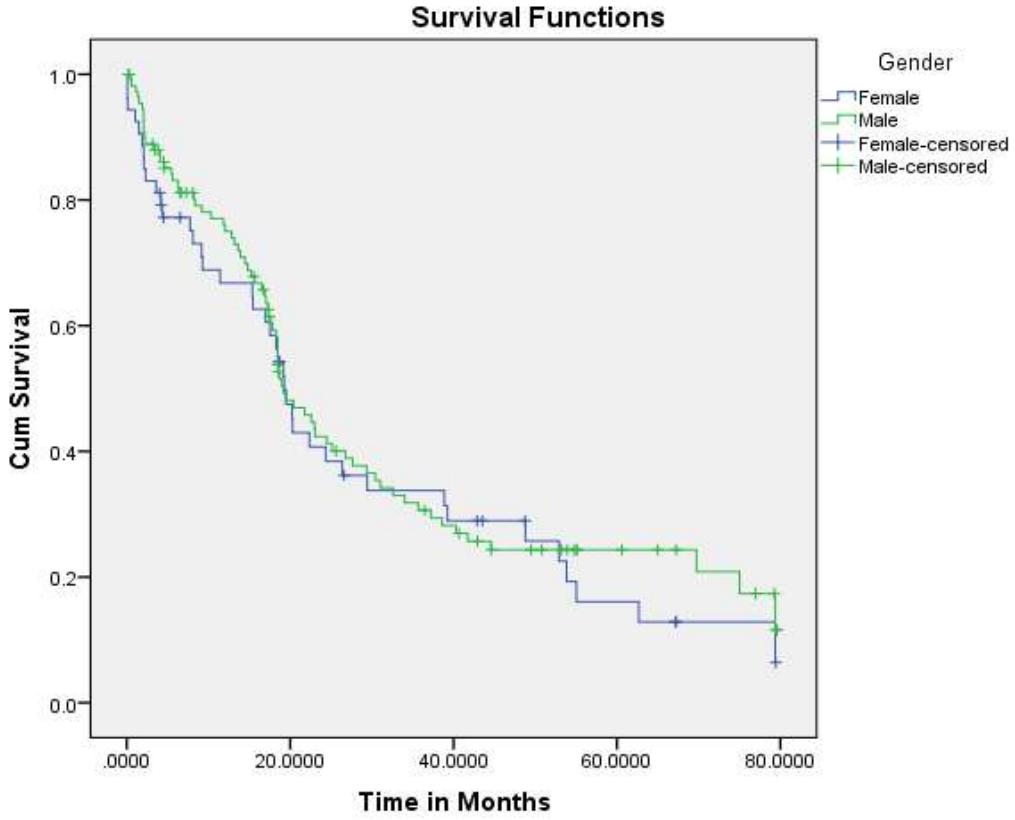
**Table-3.1.2: Overall Comparisons**

| | Chi-Square | df | Sig. |
|---|---|---|---|
| Log Rank (Mantel-Cox) | .371 | 1 | .542 |

Test of equality of survival distributions for the different levels of Gender.

The log rank test calculates a $\chi^2$-statistic in Overall Comparisons (Table-3.1.2). The value of $\chi^2$-statistic is 0.371. The p-value of $\chi^2$-statistic is 0.542 which is greater than 0.05 (level of significance). Hence we fails to reject the null hypothesis at 5% level of significance. That is we do not have a statistically significant evidence to conclude that the survival distributions of waiting time to get first employment are different in the male and female students. (i.e. Survival distributions of waiting time to get first employment are same in the male and female students).

The plot of the cumulative survival proportion against waiting time for male and female is shown below. The horizontal axis shows the waiting time to get first employment. In this plot, drops in the survival curve occur whenever the engineering students get first job. The vertical axis shows the probability of survival.




The plot above will help to understand how the survival distributions compare gender (male, female). When considering the gender, there is no difference between two survival curves. The waiting time to get first job after completion of degree is same for both male and female.

**(ii) To compare waiting time vs trade:**

The null and alternative hypothesis for Log-Rank test:

*$H_0$: Survival distributions of waiting time to get first employment are same in ten categories of tread.*

*$H_1$: Survival distributions of waiting time to get first employment are different in ten categories of tread.*

Means and Medians for Survival Time (Table-3.1.3) include that estimate of mean waiting survival time and their standard error with 95% confidence interval for each category of tread. It also includes, estimate of median waiting survival time and their Standard error with 95% confidence interval for each category of tread.

**Table-3.1.3: Means and Medians for Survival Time**

| Trade | Mean | | | | Median | | | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | Std. Error | 95% Confidence Interval | | Estimate | Std. Error | 95% Confidence Interval | |
| | | | Lower Bound | Upper Bound | | | Lower Bound | Upper Bound |
| Civil | 43.915 | 6.128 | 31.904 | 55.926 | 35.667 | 11.016 | 14.075 | 57.259 |
| Electrical | 30.018 | 7.253 | 15.802 | 44.235 | 23.000 | 2.626 | 17.854 | 28.146 |
| Mechanical | 20.277 | 3.969 | 12.498 | 28.055 | 14.800 | 5.017 | 4.967 | 24.633 |
| Chemical | 47.964 | 11.633 | 25.163 | 70.764 | 52.900 | 26.058 | 1.826 | 103.974 |
| Electronics | 28.588 | 8.407 | 12.110 | 45.066 | 19.267 | 2.161 | 15.030 | 23.503 |
| Automobile | 36.738 | 11.989 | 13.239 | 60.236 | 30.400 | 8.977 | 12.804 | 47.996 |
| ENTC | 20.786 | 3.795 | 13.348 | 28.224 | 20.300 | 3.365 | 13.705 | 26.895 |
| Textile | 12.679 | 2.441 | 7.894 | 17.464 | 13.167 | 3.441 | 6.422 | 19.912 |
| IT | 35.633 | 8.381 | 19.207 | 52.059 | 20.433 | 16.593 | .000 | 52.956 |
| Computer Science | 29.885 | 5.017 | 20.052 | 39.718 | 18.567 | 1.079 | 16.452 | 20.681 |
| Overall | 30.846 | 2.318 | 26.303 | 35.389 | 19.267 | 1.148 | 17.016 | 21.517 |

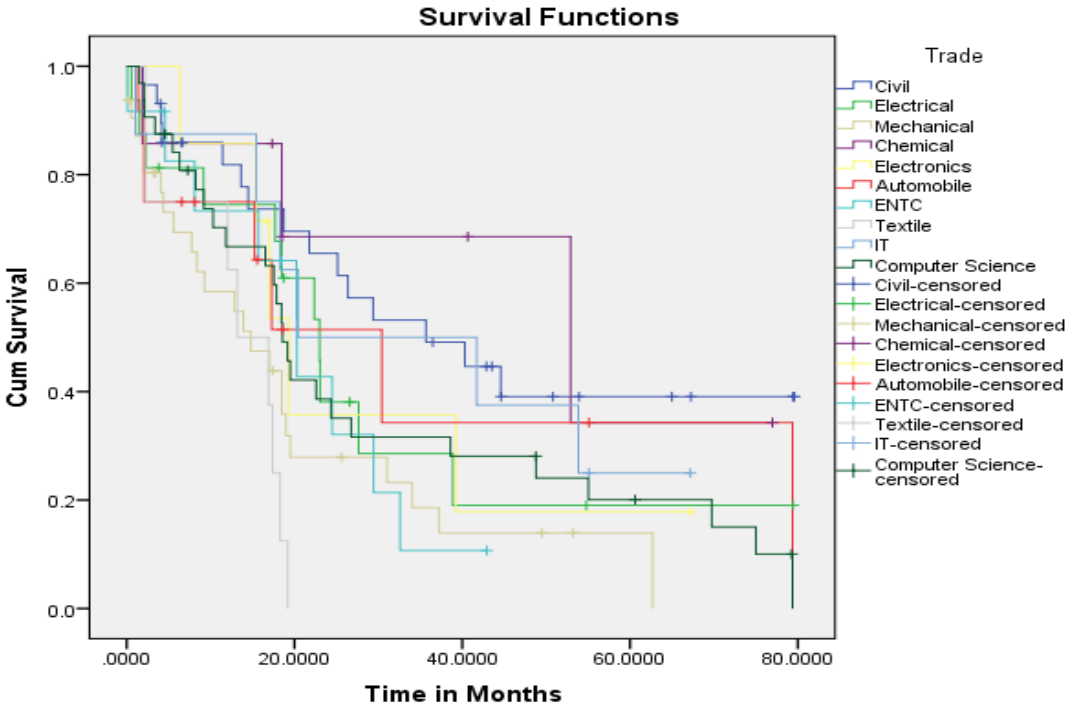
Overall Comparisons (Table-3.1.4) gives the log rank test which is used to test the null hypothesis of no difference in survival between two or more independent groups.

**Table -3.1.4: Overall Comparisons**

| | Chi-Square | df | Sig. |
|---|---|---|---|
| Log Rank (Mantel-Cox) | 21.037 | 9 | .012 |
| Test of equality of survival distributions for the different levels of Trade. | | | |

The value of $\chi^2$-statistic is 21.037. The p-value of $\chi^2$-statistic is 0.012 which is less than 0.05 (level of significance). Hence we reject the null hypothesis at 5% level of significance. That is we have a statistically significant evidence to conclude that the survival distributions of waiting time to get first employment are different in ten categories of tread.

The plot of the cumulative survival proportion against waiting time for each category of tread is shown below. The horizontal axis shows the waiting time to get first employment. In this plot, drops in the survival curve occur whenever the engineering students get first job. The vertical axis shows the probability of survival.




**3.2. Cox-Proportional Regression:** The Cox Regression used to modeling the time to a specified event, based upon the values of covariates. The basic model offered by the Cox Regression procedure is the proportional hazards model which is given below:

$$h(t) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + - - - + \beta_k x_k)$$

Where $x_1$, $x_2$, -----, $x_k$ are the covariates and $\beta'$s are regression coefficients of the covariates.

We are interested in modeling the waiting time to get first job in order to determine the factors that are associated with engineering students who are completed their graduate degree and ready to get first job. To this end, a random sample of 163 degree students is selected and their waiting time to get first job is considered as time variable, if the students are still searching for the first job is said to censored (0) and have a first job it is considered as event (1). We Use Cox Regression model to determine which variables are associated with waiting time to get first job for engineering students who are completed their graduate degree. The case processing summary (Table 3.2.1) shows that 163 cases available in analysis in which 114 are event and 49 are censored. It indicate that 114 students get first job after completion of graduate degree. However, 49 students are still searching for job.

**Table 3.2.1: Case Processing Summary**

| | | N | Percent |
|---|---|---|---|
| Cases available in analysis | Event | 114 | 69.9% |
| | Censored | 49 | 30.1% |
| | Total | 163 | 100.0% |
| Cases dropped | Cases with missing values | 0 | 0.0% |
| | Cases with negative time | 0 | 0.0% |
| | Censored cases before the earliest event in a stratum | 0 | 0.0% |
| | Total | 0 | 0.0% |
| Total | | 163 | 100.0% |

Categorical variable Coding (Table-3.2.2) shows that the male student have variable values of 1 in the data file, they are coded as 0 for the purposes of the Cox Regression.

**Table-3.2.2: Categorical Variable Codings**

| | | Frequency | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Gender | 0=Female | 53 | 1 | | | | | | | | |
| | 1=Male | 110 | 0 | | | | | | | | |
| Trade | 1=Civil | 29 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 2=Electrical | 16 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 3=Mechanical | 32 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 4=Chemical | 7 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | 5=Electronics | 7 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| | 6=Automobile | 12 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| | 7=ENTC | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | 8=Textile | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | 9=IT | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | 10=Computer Science | 32 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

To build Cox-Proportional Regression model, a Backward Stepwise (Likelihood Ratio) method used with the through the SPSS software. To check the model performance, the omnibus test is used. The chi-square change from previous step is the difference between the –2 log-likelihood of the model at the previous step and the current step.

**Table 3.2.3: Omnibus Tests of Model Coefficients**

| Step | -2 Log Likelihood | Overall (score) | | | Change From Previous Step | | | Change From Previous Block | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Chi-square | df | Sig. | Chi-square | df | Sig. | Chi-square | df | Sig. |
| 1[a] | 935.689 | 46.042 | 15 | .000 | 47.858 | 15 | .000 | 47.858 | 15 | .000 |
| 2[b] | 935.829 | 45.992 | 14 | .000 | .140 | 1 | .708 | 47.718 | 14 | .000 |
| 3[c] | 936.153 | 45.648 | 13 | .000 | .324 | 1 | .569 | 47.394 | 13 | .000 |
| 4[d] | 937.725 | 44.064 | 12 | .000 | 1.572 | 1 | .210 | 45.822 | 12 | .000 |

a. Variable(s) Entered at Step Number 1: Age , Gender, Trade, No. of year, CGPA, Extra courses, No. of courses

b. Variable Removed at Step Number 2: Gender

c. Variable Removed at Step Number 3: No. of Courses

d. Variable Removed at Step Number 4: Extra courses

e. Beginning Block Number 1. Method = Backward Stepwise (Likelihood Ratio)

| Table-3.2.4: Variables in the Equation | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | B | SE | Wald | df | Sig. | Exp(B) |
| Step 1 | Age | -.280 | .066 | 17.896 | 1 | .000 | .756 |
| | Gender | .082 | .217 | .141 | 1 | .707 | 1.085 |
| | Trade | | | 25.241 | 9 | .003 | |
| | Trade(1) | -.456 | .365 | 1.563 | 1 | .211 | .634 |
| | Trade(2) | -.130 | .380 | .117 | 1 | .732 | .878 |
| | Trade(3) | .491 | .323 | 2.315 | 1 | .128 | 1.634 |
| | Trade(4) | -.897 | .639 | 1.969 | 1 | .161 | .408 |
| | Trade(5) | .235 | .511 | .211 | 1 | .646 | 1.265 |
| | Trade(6) | -.107 | .452 | .056 | 1 | .813 | .898 |
| | Trade(7) | .086 | .412 | .044 | 1 | .834 | 1.090 |
| | Trade(8) | 1.518 | .453 | 11.248 | 1 | .001 | 4.564 |
| | Trade(9) | -.394 | .484 | .662 | 1 | .416 | .674 |
| | No. of year | .264 | .112 | 5.547 | 1 | .019 | 1.302 |
| | CGPA | .182 | .102 | 3.156 | 1 | .076 | 1.199 |
| | Extra courses | .212 | .279 | .577 | 1 | .447 | 1.236 |
| | No. of courses | .063 | .109 | .334 | 1 | .563 | 1.065 |
| Step 2 | Age | -.281 | .066 | 18.129 | 1 | .000 | .755 |
| | Trade | | | 25.339 | 9 | .003 | |
| | Trade(1) | -.461 | .364 | 1.603 | 1 | .206 | .630 |
| | Trade(2) | -.121 | .379 | .101 | 1 | .750 | .886 |
| | Trade(3) | .474 | .319 | 2.208 | 1 | .137 | 1.606 |
| | Trade(4) | -.900 | .639 | 1.983 | 1 | .159 | .407 |
| | Trade(5) | .260 | .506 | .264 | 1 | .607 | 1.297 |
| | Trade(6) | -.141 | .443 | .101 | 1 | .750 | .868 |
| | Trade(7) | .090 | .412 | .047 | 1 | .828 | 1.094 |
| | Trade(8) | 1.489 | .446 | 11.172 | 1 | .001 | 4.435 |
| | Trade(9) | -.388 | .484 | .643 | 1 | .423 | .678 |
| | No. of year | .264 | .112 | 5.532 | 1 | .019 | 1.302 |
| | CGPA | .186 | .102 | 3.356 | 1 | .067 | 1.205 |
| | Extra courses | .221 | .277 | .635 | 1 | .426 | 1.247 |
| | No. of courses | .062 | .108 | .329 | 1 | .566 | 1.064 |
| Step 3 | Age | -.280 | .066 | 17.876 | 1 | .000 | .756 |
| | Trade | | | 26.158 | 9 | .002 | |
| | Trade(1) | -.543 | .333 | 2.659 | 1 | .103 | .581 |
| | Trade(2) | -.180 | .363 | .245 | 1 | .621 | .836 |
| | Trade(3) | .412 | .299 | 1.903 | 1 | .168 | 1.510 |
| | Trade(4) | -.994 | .616 | 2.600 | 1 | .107 | .370 |
| | Trade(5) | .197 | .493 | .159 | 1 | .690 | 1.217 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Trade(6) | -.190 | .434 | .191 | 1 | .662 | .827 |
| | Trade(7) | .029 | .396 | .005 | 1 | .942 | 1.029 |
| | Trade(8) | 1.438 | .435 | 10.921 | 1 | .001 | 4.213 |
| | Trade(9) | -.475 | .459 | 1.069 | 1 | .301 | .622 |
| | No. of year | .235 | .100 | 5.513 | 1 | .019 | 1.265 |
| | CGPA | .192 | .101 | 3.642 | 1 | .056 | 1.212 |
| | Extra courses | .297 | .242 | 1.504 | 1 | .220 | 1.346 |
| Step 4 | Age | -.283 | .065 | 18.632 | 1 | .000 | .754 |
| | Trade | | | 26.247 | 9 | .002 | |
| | Trade(1) | -.534 | .333 | 2.566 | 1 | .109 | .586 |
| | Trade(2) | -.214 | .363 | .348 | 1 | .555 | .807 |
| | Trade(3) | .400 | .300 | 1.776 | 1 | .183 | 1.491 |
| | Trade(4) | -.995 | .617 | 2.596 | 1 | .107 | .370 |
| | Trade(5) | .180 | .493 | .133 | 1 | .715 | 1.197 |
| | Trade(6) | -.132 | .433 | .093 | 1 | .761 | .876 |
| | Trade(7) | -.014 | .395 | .001 | 1 | .972 | .986 |
| | Trade(8) | 1.482 | .434 | 11.652 | 1 | .001 | 4.401 |
| | Trade(9) | -.405 | .456 | .785 | 1 | .375 | .667 |
| | No. of year | .231 | .101 | 5.261 | 1 | .022 | 1.260 |
| | CGPA | .217 | .099 | 4.842 | 1 | .028 | 1.243 |

The final model (Step 4) includes Age, Tread, No. of years, and CGPA of Graduate students. To understand the effects of individual predictors, we can see the Exp(B) in table 3.2.4 step 4, which can be interpreted as the predicted change in the hazard for a unit increase in the predictor.

## 4. Discussion and Conclusions:

i. Kaplan-Meier analysis shows that, survival distributions of waiting time to get first employment are same between male and female students. However, the survival distributions of waiting time to get first employment are different in ten categories of tread of engineering degree.

ii. Cox-Proportional Hazard regression analysis suggest that, the Exp(B) value 0.754 for the covariate age indicate that the waiting time hazard is reduced by 24.6 % with increase in age of student each year. The regression coefficients for the nine levels of Trade [Trade (1) – Trade (9)] are relative to the reference category, named as Computer Science. The regression coefficient for the Civil Trade, corresponding to Computer Science Trade, suggests that the hazard for Civil Trade students is 0.586 times that of Computer Science Trade students. However, the p-value for this coefficient is greater than 0.05, which means Civil Trade is not statistically different from the Computer Science Trade. The Exp(B) value 1.260 for covariate number of years required to complete degree indicate that the waiting time hazard is increases by 26 % for each year delay in completion of degree. The Exp(B) value 1.243 for

covariate CGPA obtained in degree indicate that the waiting time hazard is increases by 24.3 % for each unit increase in CGPA.

**References:**


1) M. G. Ayaneh, A. A. Dessie, A. W. Ayele (2020). "Survival Models for the Analysis of Waiting Time to First Employment of New Graduates: A Case of 2018 Debre Markos University Graduates, Northwest Ethiopia", *Education Research International*, Volume 2020, Article ID 8877504, 1-10.
2) I. T. Jayamanne, K. A. Ramanayake (2017). "A Study on the Waiting Time for the First Employment of Arts Graduates in Sri Lanka" *International Journal of Computer and Information Engineering*, Vol: 11, No:12, pp. 1153-1161.
3) M. Salas-Velasc (2007). "The transition from higher education to employment in Europe: the analysis of the time to obtain the first job", *Springer, Higher Education*, 333–360.
4) M. A. Pourhoseingholi, E. Hajizadeh, B. Moghimi Dehkordi, A. Safaee, A. Abadi, and M. Reza Zali, (2007). "Comparing cox regression and parametric models for survival of patients with gastric carcinoma," *Asian Pacific Journal of Cancer Preve*ntion, vol. 8, no. 3, pp. 412–416.
*5)* Kleinbaum, D. G. (1996). "Survival Analysis: A Self-Learning Text". *New York: Springer-Verlag.*

Contents available at www.balwantcollege.edu.in

**Emerging Trends in Applied Mathematics and Statistics**

***(Balwant College, Vita (M.S.) India 415 311)***

# A Comparative Study of Employees Job Satisfaction in Government and Private Sector: Special Reference to Khanapur Tahsil

Pragati Shinde[1*], Nikita Velapure[2]

*[1,2]PG Student, Department of Statistics, Balwant College, Vita*

**Abstract:** Employees Job Satisfaction is one of the most widely discussed issues in any Government and Privet Organization. It is a very important attribute which is frequently measured by organizations. There is a direct relation between employee job satisfaction and financial results of organization. The more satisfied employees are the more motivated and committed they will be leads to organization's success. The main purpose of the present study is to compare the job satisfaction of employees working in government and private sector. The Job satisfaction was measured on 29 government and 58 privet employees. Result indicates that there was no significant difference between job satisfaction score of government and private sector employees. However there is significant difference in job satisfaction score among the specialization of employees.

**Keywords:** Government, Private, ANOVA, T-test, Job Satisfaction.

## 1. Introduction:

The Employees Job satisfaction is a set of favorable or unfavorable feelings with which employees view their work. It is a worker's sense of achievement and success. If the people are happier with their job then they are more satisfied. Job satisfaction implies doing a job one enjoys, doing it well, and being suitably rewarded for one's efforts. The job satisfaction affected by many factors such as security in job, working hours, Skill utilization, salary etc. Organizations can identify the factors affecting on job issues and create solutions for improvements. Employee's satisfaction increase when an employee knows that their issues are being addressed. To measure the job satisfaction of employee, we use of rating scales where employees report their reactions on their jobs.

The main objective of research paper is to compare the employee's job satisfaction score between government and privet employees. Shanthi C. et al. (2019) compared job satisfaction in public and private school teachers working in Thirupalathurai.

Kaushik Bhakta (2016) studied the job satisfaction of primary school teachers in Howrah district. Ajay Chauhan et al. (2014) compared job satisfaction in government and private employees in different service sectors of Anand District in Gujarat state. There result showed that there is no significant difference among government and private employees in job satisfaction. Pankaj Arora et al. (2014) compared job satisfaction between private and government sector and found the basic reasons of dissatisfaction in job.

***Corresponding Author:** *Miss. Pragati Shinde, PG Student, Department of Statistics, Balwant College, Vita (M.S.) India 415311.* **Email:** pragatis883@gmail.com*

## 2. Material and Methods:

The employees selected for this study from different service sectors of Khanapur Tahsil. In this study the random sampling method is adopted. Total of 87 employees, 47 are male and 40 are female. The primary data collected from the respondents who are working in government and private organization through the GOOGLE form. The independent two sample t test and one way ANOVA were used to compare the employee's job satisfaction. The statistical software, SPSS was used for data analysis.

## Objectives:

1) To compare the job satisfaction score between government and private employees.
2) To compare the job satisfaction score between male and female employees.
3) To compare the job satisfaction score between employees living in urban and rural area.
4) To compare the job satisfaction score between Permanent and Contractual Employees.
5) To compare the job satisfaction score among education qualification of employees.
6) To compare the job satisfaction score among working time of employees.
7) To compare the job satisfaction score among specialization of employees.

## Hypotheses:

| Null Hypothesis ($H_0$) | Alternative Hypothesis ($H_1$) |
|---|---|
| i) There is no significant difference between job satisfaction score of government and private employees. | i) There is significant difference between job Satisfaction score of government and private Employees. |
| ii) There is no significant difference between job satisfaction score of male and female employees. | ii) There is significant difference between job satisfaction score of male and female employees. |
| iii) There is no significant difference between job satisfaction score of employees living in urban and rural area. | iii) There is significant difference between job satisfaction score of employees living in urban and rural area. |
| iv) There is no significant difference between job satisfaction score of permanent and contractual employees. | iv) There is significant difference between job satisfaction score of permanent and contractual employees. |
| v) There is no significant difference between job satisfaction score in five categories of employee's education qualification. | v) At least one category of employee's education qualification differs significantly in job satisfaction score. |
| vi) There is no significant difference between job satisfaction score in three categories of employee's working time. | vi) At least one category of employee's working time differs significantly in job satisfaction score. |
| vii) There is no significant difference between job satisfaction score in five categories of employee's specialization. | vii) At least one category of employee's specialization differs significantly in job satisfaction score. |

## 3. Statistical Analysis:

In present study employees job satisfaction score is dependent variable. We have given fourteen categorical questions to 87 government and privet employees in order to calculate the job satisfaction score. Each Question of about Job satisfaction is coded as:

Highly Dissatisfied (1), Dissatisfied (2), Neutral (3), Satisfied (4), Highly Satisfied (5). We have summarized score of these fourteen questions about Job Satisfaction to form dependent variable named as Job Satisfaction Score.

We have taken seven independent variables, namely, Gender, Area of employee, Job Type, Nature of Job, Education Qualification, Working Time and Specialization of employees. Each categorical variable, the Frequency (N), Mean, Standard deviation and Standard error are summarized in table-3.1. The Descriptive Statistics table shows, for the categorical variables gender, out of 87 employees 47 are male, 40 are female. The mean job satisfaction score for male and female are 50.89 and 49.80 respectively. The standard deviations for job satisfaction score for male and female are 14.095 and 10.254 respectively. However, the standard error for job satisfaction scores for male and female are 2.056 and 1.621 respectively. The Frequency (N), Mean, Standard deviation and Standard error of job satisfaction Score for other six categorical variables are shown in following Table-3.1.

**Table-3.1: Descriptive Statistics**

| Independent Variable | Categories | N | Mean | Std. Deviation | Std. Error |
|---|---|---|---|---|---|
| Gender | Female | 40 | 49.80 | 10.254 | 1.621 |
| | Male | 47 | 50.89 | 14.095 | 2.056 |
| Area of Employee | Rural | 59 | 50.80 | 11.409 | 1.485 |
| | Urban | 28 | 49.54 | 14.510 | 2.742 |
| Job Type | Government | 29 | 51.34 | 12.442 | 2.310 |
| | Private | 58 | 49.91 | 12.489 | 1.640 |
| Nature of Job | Permanent | 48 | 51.77 | 11.999 | 1.732 |
| | Contractual | 39 | 48.69 | 12.870 | 2.061 |
| Education Qualification | HSC | 5 | 41.60 | 15.931 | 7.125 |
| | Graduation | 42 | 50.60 | 11.886 | 1.834 |
| | Post Graduation | 37 | 52.43 | 11.278 | 1.854 |
| | Ph.D. | 3 | 37.00 | 19.975 | 11.533 |
| Working Time | 5-7 hrs | 23 | 51.48 | 8.826 | 1.840 |
| | 7-9 hrs | 54 | 51.22 | 12.021 | 1.636 |
| | above 9 hrs | 10 | 43.40 | 19.260 | 6.090 |
| Specialization | Arts | 8 | 46.50 | 12.862 | 4.547 |
| | Commerce | 15 | 41.67 | 16.482 | 4.256 |
| | Science | 48 | 53.75 | 9.141 | 1.319 |
| | Engineer | 13 | 52.00 | 14.095 | 3.909 |
| | Agriculture | 3 | 43.67 | 3.055 | 1.764 |
| Total | | 87 | 50.39 | 12.419 | 1.331 |

Independent Sample t test (table -3.2) showing some dependent variable, Independent variable and Degrees of freedom required for t test. It also includes the value of test statistic

and p-value for each binary categorical variable. The value of t-test statistics for variable gender is -0.407 and corresponding p-value is 0.685.

**Table-3.2: Independent Sample t test**

| Dependent Variable | Independent Variable | Degrees of Freedom | Value of t-test statistic | p-value |
|---|---|---|---|---|
| Satisfaction Score | Gender | 85 | -0.407 | 0.685 |
| | Area | 85 | 0.440 | 0.661 |
| | Job Type | 85 | 0.504 | 0.615 |
| | Nature of Job | 85 | 1.152 | 0.253 |

To conduct ANOVA we used dependent variable as job satisfaction score and independent variables like education qualification, working time, and specialization. The ANOVA table -3.3 includes source of variation, sum of squares, degrees of freedom, mean square, F-test statistic and corresponding significance value (p-value) for each independent variable calculated separately.

**Table-3.3: ANOVA**

| **Dependent Variable: Satisfaction Score** | | | | | | |
|---|---|---|---|---|---|---|
| **Independent Variable** | **Source of variation** | **Sum of Squares** | **df** | **Mean Square** | **F** | **Sig.** |
| Education Qualification | Between Groups | 1080.313 | 3 | 360.104 | 2.453 | .069 |
| | Within Groups | 12184.400 | 83 | 146.800 | | |
| | Total | 13264.713 | 86 | | | |
| Working Time | Between Groups | 553.240 | 2 | 276.620 | 1.828 | .167 |
| | Within Groups | 12711.472 | 84 | 151.327 | | |
| | Total | 13264.713 | 86 | | | |
| Specialization | Between Groups | 1973.713 | 4 | 493.428 | 3.583 | .010 |
| | Within Groups | 11291.000 | 82 | 137.695 | | |
| | Total | 13264.713 | 86 | | | |

If we reject the null hypothesis for ANOVA, then we have to do multiple comparison test which is known as Tukey HSD. The multiple comparison Table-3.4 includes pairwise comparison of each category of specialization. The mean difference, standard error, significance value and 95% confidence interval for mean difference is calculated for each pair of specialization category.

**Table-3.4: Multiple Comparisons**

| Dependent Variable: Satisfaction Score | | Tukey HSD | | | | |
|---|---|---|---|---|---|---|
| (I) Specialization | (J) Specialization | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
| | | | | | Lower Bound | Upper Bound |
| | Commerce | 4.833 | 5.137 | .880 | -9.50 | 19.16 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Arts | Science | -7.250 | 4.481 | .490 | -19.75 | 5.25 |
| | Engineer | -5.500 | 5.273 | .835 | -20.21 | 9.21 |
| | Agriculture | 2.833 | 7.944 | .996 | -19.33 | 24.99 |
| Commerce | Arts | -4.833 | 5.137 | .880 | -19.16 | 9.50 |
| | Science | -12.083* | 3.471 | .007 | -21.77 | -2.40 |
| | Engineer | -10.333 | 4.447 | .148 | -22.74 | 2.07 |
| | Agriculture | -2.000 | 7.421 | .999 | -22.70 | 18.70 |
| Science | Arts | 7.250 | 4.481 | .490 | -5.25 | 19.75 |
| | Commerce | 12.083* | 3.471 | .007 | 2.40 | 21.77 |
| | Engineer | 1.750 | 3.669 | .989 | -8.48 | 11.98 |
| | Agriculture | 10.083 | 6.983 | .601 | -9.40 | 29.56 |
| Engineer | Arts | 5.500 | 5.273 | .835 | -9.21 | 20.21 |
| | Commerce | 10.333 | 4.447 | .148 | -2.07 | 22.74 |
| | Science | -1.750 | 3.669 | .989 | -11.98 | 8.48 |
| | Agriculture | 8.333 | 7.516 | .801 | -12.63 | 29.30 |
| Agriculture | Arts | -2.833 | 7.944 | .996 | -24.99 | 19.33 |
| | Commerce | 2.000 | 7.421 | .999 | -18.70 | 22.70 |
| | Science | -10.083 | 6.983 | .601 | -29.56 | 9.40 |
| | Engineer | -8.333 | 7.516 | .801 | -29.30 | 12.63 |
| *. The mean difference is significant at the 0.05 level. | | | | | | |

**4. Result and Discussions:** The results of Independent Sample t test **(**Table-3.2) shows that,

i. For the variable gender, the p-value of t test statistic is 0.685 which is greater than 0.05 (level of significance). Hence we fail to reject null hypothesis at 5% level of significance. Therefore, there is no significant difference between job satisfaction score of male and female employees.

ii. For the variable area, the p-value of t test statistic is 0.661 which is greater than 0.05 (level of significance). Hence we fail to reject null hypothesis at 5% level of significance. Therefore, there is no significant difference between job satisfaction score of employees living in urban and rural area.

iii. For the variable job type, the p-value of t test statistic is 0.615 which is greater than 0.05 (level of significance). Hence we fail to reject null hypothesis at 5% level of significance. Therefore, there is no significant difference between job satisfaction score of government and private employees.

iv. For the variable nature of job, the p-value of t test statistic is 0.253 which is greater than 0.05 (level of significance). Hence we fail to reject null hypothesis at 5% level of significance. Therefore, there is no significant difference between job satisfaction score of permanent and contractual employees.

The output of ANOVA Table-3.3 shows that:

i. The p-value of F-test statistic for the variable education qualification is 0.069 which is greater than 0.05 (level of significance). Hence we fail to reject the null hypothesis at

5% level of significance. Therefor there is no significant difference between job satisfactions score in five categories of employee's education qualification.

ii. The p-value of F-test statistic for the variable working time is 0.167 which is greater than 0.05 (level of significance). Hence we fail to reject the null hypothesis at 5% level of significance. Therefore there is no significant difference between job satisfactions score in five categories of employee's working time.

iii. The p-value of F-test statistic for the variable specialization is 0.010 which is less than 0.05 (level of significance). Hence we reject the null hypothesis at 5% level of significance. Therefore, at least one category of employee's specialization differs significantly in job satisfaction score. The next task is to identify which pair of employee's specialization differs significantly with the help of Tukey Post-Hoc multiple comparison test.

iv. The multiple comparisons Table-3.4, shows which groups differed from each other. We can see that there is significant difference in job satisfaction score between the specialization Commerce & Science (P-value = 0.007). However there is no significant differences in job satisfaction between specialization groups which has P-value greater than 0.05.

**5. Conclusions:**

There is not enough evidence to conclude that the job satisfaction score of government and private employees are different. The specialization of employees significantly effect on the employee's job satisfaction score. However, employee who completed their specialization in commerce has job satisfaction score differ significantly as compared to employees who completed their specialization in Science. The factors gender, education qualification, job type, nature of job, working time and area of employee living doesn't affect on employee job satisfaction score.

**References:**

1) Shanthi C, Rajandran KVR (2019*). "*A Comparative Study Of Job Satisfaction In Public And Private School Teachers In Thirupalathurai Town*", JETIR*, Volume 6, Issue 5.
2) Kaushik Bhakta (2016), "Job Satisfaction of Primary School Teachers in Howrah Dt", *International Journal of Interdisciplinary Research*, Vol 2(10), pp 124 – 129.
3) Ajay Chauhan, Dr. Pravin M. Solanki (2014). *"*A Comparative Study of Job Satisfaction in Government and Private Employees*", The International Journal of Indian Psychology*, Volume 2, Issue 1, pp. 17-22.
4) Pankaj Arora, Shevya Rawal, Dolly Sethi (2014). "A Comparative Study of Job Satisfaction in Private and Public Sector", *IJITKM*. Special Issue, pp. 193-196.
5) Shobhna Gupta J, Hartesh Pannu K (2013). "A Comparative Study of Job Satisfaction in Public and Private Sector", *Indian Journal of Arts*, V-1, N-1.

Contents available at *www.balwantcollege.edu.in*
**Emerging Trends in Applied Mathematics and Statistics**
*(Balwant College, Vita (M.S.) India 415 311)*

# Calories Burn Prediction using Ridge Regression and Random Forest Regression algorithms

Purva Bhosale[1*], Snehal Gidde[2]
[1]*Assistant Professor, Department of Statistics, Balwant College, Vita*
[2]*P.D.V.P. College, Tasgaon*

**Abstract:** The overarching idea of this research study is to make a comparative study of machine learning algorithms to predict the calories burn during the workout. In this research study we first build a machine learning systems that can predict the amount of calories burnt during exercise. In current century many people are doing workout according to weight loss plan that they have taken and calculates how much calorie do they burn once they workout. To solve this problem we use ML algorithms such as Random forest regression and Ridge Regression.

**Key Words:** Prediction, Calories, Random Forest, Ridge Regression, Weight.

**1. Introduction:** In this research study, let's predict the calories burn using Machine learning & let's have a healthy & a happier life. This research study is about calorie prediction with machine learning using python. We will predict calorie based on some features. We eat foods to provide energy so that our bodies can function. This means that we need to eat a certain amount of calories just to sustain life. The risk gaining weight is increases if we take in too many calories. So there is need to burn Calories, for burning calories we doing exercises and more. For know how much calories we have burn during exercise we are going to build a machine learning model that predict calories.

Based on the same data. When we exercise, the body temperature and the heartbeat will be rise. The variables time scale is taken for which the individual carrying out the workout training, the average beats per minute and temperature. Then we additionally take the height, weight, gender and age of the person to predict how tons energy the person may be burning. Suvarna S R. et al. (2022) predicted calorie burn using Machine Learning.

In present research paper, we want to predict calories burned using machine learning random forest regression algorithm and ridge regression algorithms on the independent variables, namely, duration, body temperature, height, weight and age of the person.

## 2. Methods and Materials:

This research study is all about the collection of appropriate set to teach our machine learning models in order that it will find out what is the amount of calories that the individual

**Corresponding Author:* *Miss. Purva Bhosale, Assistant Professor, Department of Statistics, Balwant College, Vita (M.S.) India 415 311.* ***E-mail:*** *bhosalepurva09@gmail.com*

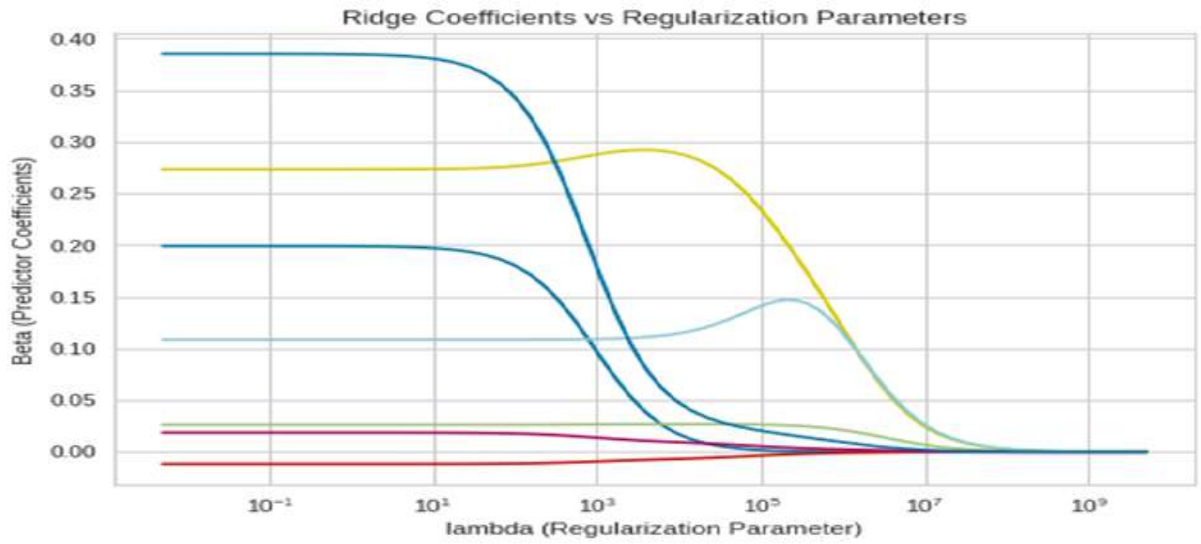
goes to burn. Before feeding procedure the statistics via records pre-processing need to be done. After that data analysis is carry out where we use some visualization techniques to arrange the data in plots and graphs. Afterwards divide the data set into training and test set. Here we use Random forest regression and Ridge regression as machine learning models for comparison and then evaluate MS-Excel and python.

## 3. Statistical Analysis:

### 3.1 Ridge Regression:

Ridge regression is the method used to analysis the multi co-linearity in multiple regression. When the data contains more number of independent variables then it is most suitable.

It is also known as, L2 Regression adds a penalty to the existing model. Ridge regression adds penalty to loss function which makes the model have a smaller coefficients value. i.e, it shrinks coefficients of the variables of the model that don't contribute much to the model. Based on the **Sum of Square Error** it penalizes the model**.** But it prevents from being excluded from model by letting them have towards zero as coefficients value




**Interpretation:** From the plot as alpha increases the coefficients convert to smaller values of their original. The power of ridge regression is to make the coefficients smaller to limit the co-linearity between predictors.

Ridge regression can be framed as follows:

**Ridge = loss + (lambda *L2_penalty)**

**Ridge = loss+ ($\lambda * \Sigma\beta_j^2$)**

**Loss = $\Sigma(y_i - \hat{y}_i)2$**

- **$\Sigma$** : A Greek symbol that means sum
- **$y_i$** : The actual response value for the $i^{th}$ observation
- **$\hat{y}_i$** : The predicted response value based on the multiple linear regression model
- where *j* ranges from 1 to P predictor variables and $\lambda \geq 0$.

- This second term in the equation is known as a shrinkage penalty. we select the value of λ that produces the lowest possible test mean squared error.

**3.2 Random forest Regression:** Random forest is a type of supervised machine learning algorithm based on ensemble learning. The Random forest algorithm combines multiple algorithm of the same type. i.e. The multiple decision trees resulting in forest of trees called as Random Forest. For both the regression and classification tasks, the random forest algorithm can be used.

**Algorithm for Random forest:**

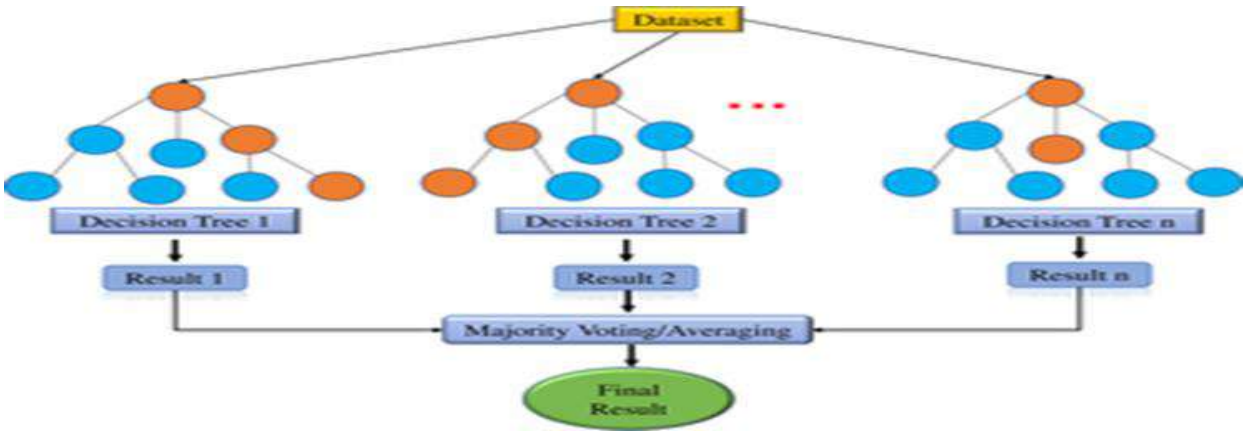
The following are the basic steps involved in performing the random forest algorithm:

1) From the dataset, Pick N random records.

2) Based on these N records build a decision tree.

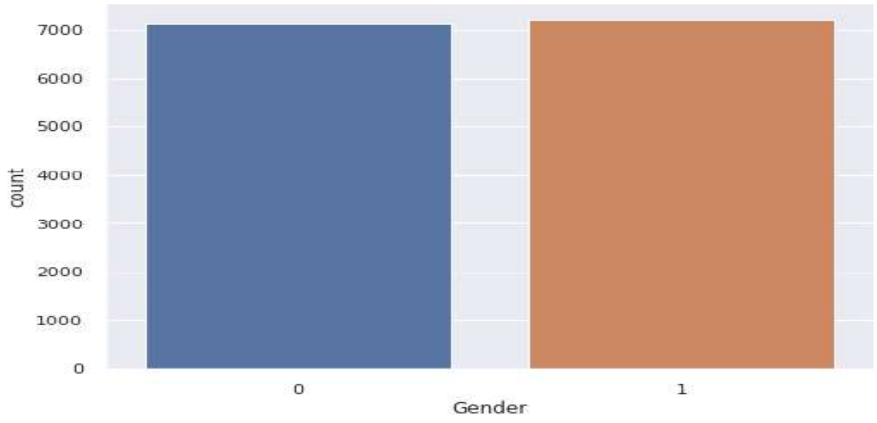
3) Select the number of trees that you want in algorithm and repeat steps-1 and 2.

4) In case of a regression problem, for a new record, each tree in the forest predicts a value for Y (output). By taking the average of all values predicted by all the trees in forest, the final value can be calculated. In case of a classification problem, each tree in forest predicts the category to which the new record belongs. Finally, the new record is assigned to the category that wins the majority vote.




**Data visualization:**

sns.countplot(combined new['Gender'])

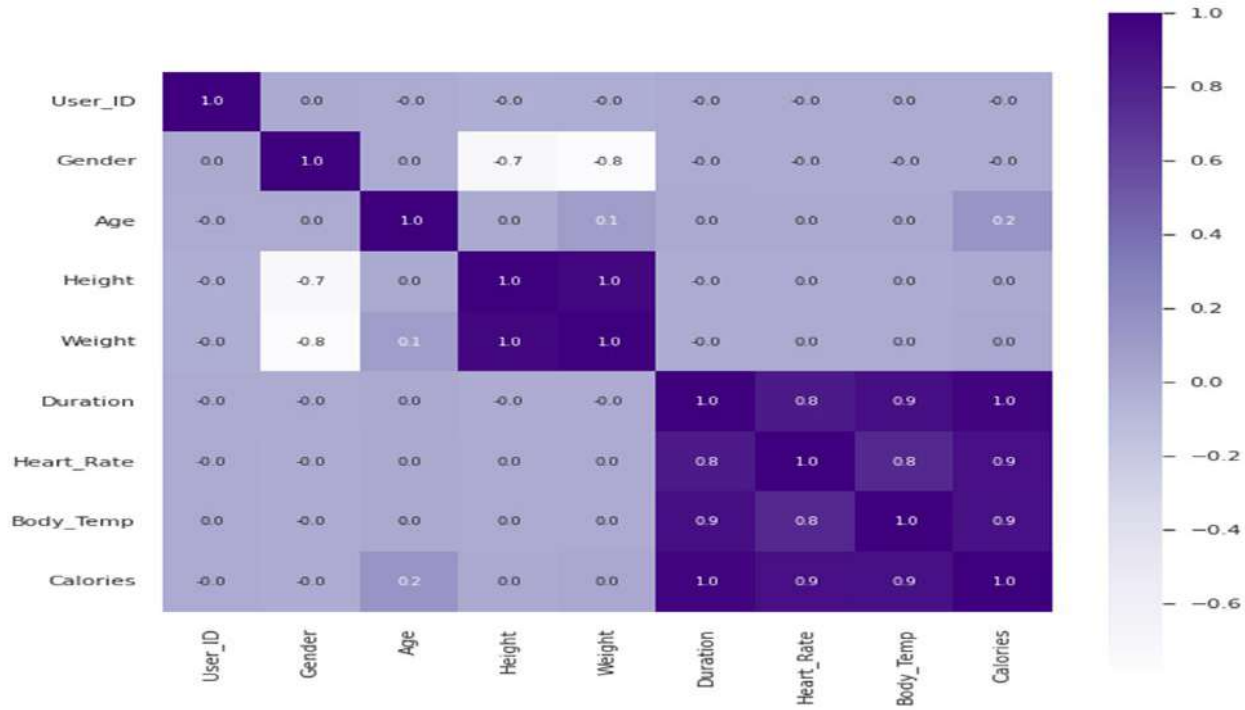


**Interpretation:** From the above plot, we can observe that the data is evenly distributed for both males and females.

**Now we plot heat map for understanding correlation:**

Heat map gives colors based on values, and these values are calculated based on the relationship between the data. Each column would be compared to the other column and if the value is 1 than two columns are positively correlated if the value is 0 than two columns are not correlated, if less than 0 than two columns are negatively correlated. So from graph it can be seen duration, heart rate and body temperature are positively related to calories burnt.

```
plt.figure(fig size=(10,10))
sns.heatmap(correlation, cbar=True, square=True, fmt='.1f', annot=True, annot_kws={'size':8
}, cmap='Purples')
```




**Interpretation**: From above plot we can see that Duration, Heart Rate & Body Temp is highly correlate in target variable (Calories).

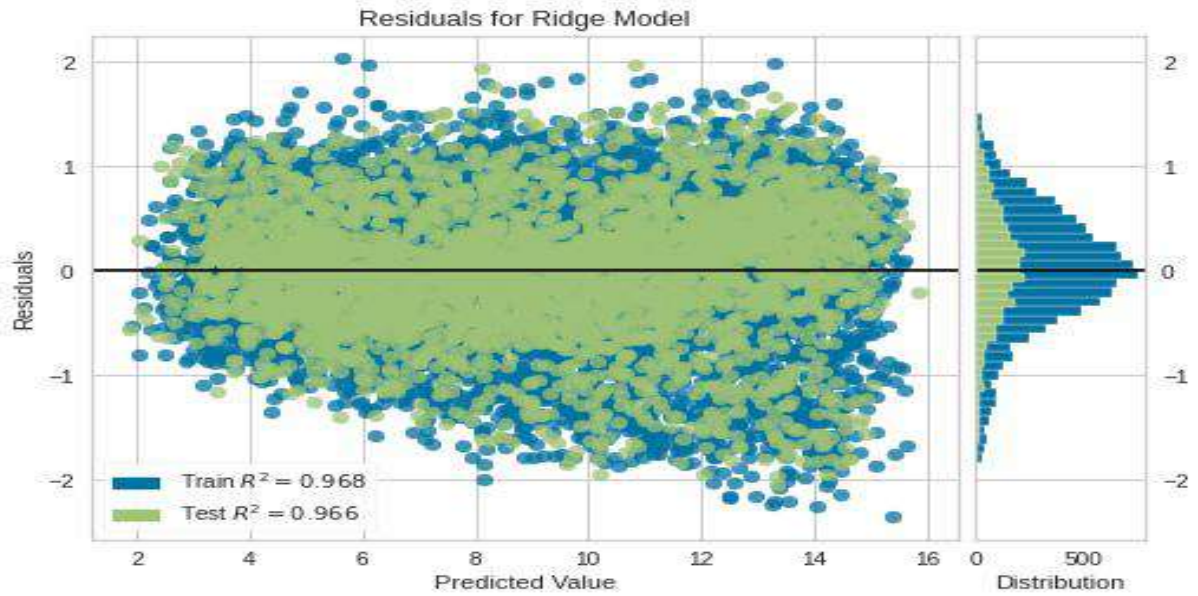
Here, we can see that Height and weight are highly correlated with each other, and Duration, Heart Rate & Body Temp are highly correlated with each other.

Therefore multi co- linearity exists.

From this we get that the height and weight are positively correlated and the duration, heart rate and body temperature are highly positively correlated with calories.

**Residual Plot of Ridge Regression:**

```
from yellowbrick.regressor import ResidualsPlot
visualizer = ResidualsPlot(ridge_model)
visualizer.fit(x_train, y_train)
visualizer.score(x_test, y_test)
 visualizer.show()
```




**Interpretation**: In the case above, we see a fairly random, uniform distribution of the residuals against the target in two dimensions. This seems to indicate that our linear model is performing well. We can also see from the histogram that our error is normally distributed around zero, which also generally indicates a well fitted model.

4. **Result**: The analysis of this dataset was done to predict the calories burned depends on the duration of workout and also based on the gender, height, weight, age body temperature and heart rate at some stage in the exercise. By using these machine learning algorithms we are looking for a machine learning model with greater $R^2$, which gives more accurate results. By comparing the two models, Random forest regression and Ridge regression we get that the Random forest regression gives the more accurate results of the calories burned with $R^2$ 0.9995515067354402 than the Ridge regression.

| **Machine Learning Model** | **$R^2$** | **Adj $R^2$** |
|---|---|---|
| Ridge Regression | 0.9875484041534515 | 0.9875239960966254 |
| Variable selection | 0.9660611777097421 | 0.9660326975791489 |

| **Machine Learning Model** | **Input data** | **Predicted Calorie result** | **Expected Calorie result** |
|---|---|---|---|
| Random Forest Regression | Female{1},21,139.0,43.0,17.0,98.0,40.2 | 9.45825413 | 9.539392 |
| Variable selection | 17.0,98.0,40.2 | 9.52504734 | 9.539392 |
| Ridge Regression | Female{1},21,139.0,43.0,17.0,98.0,40.2 | 8.87111566 | 9.539392 |
| Variable selection | 17.0,98.0,40.2 | 9.28860214 | 9.539392 |

**5. Conclusions:** From the analysis we met with a conclusion that the Random Forest Regression has more accurate results than the Ridge regression model. The $R^2$ value that is getting in Random forest regression is 0.9995 which is very close to 1 than the Ridge regression model $R^2$ value 0.9875. It means that the Random forest model $R^2$ value is larger than the Ridge regression $R^2$ value. Therefore we can conclude that the best model for the calories burn prediction is Random Forest Regression.

**References:**

1. Suvarna Shreyas Ratnakar, Vidya S (2022), "Calorie Burn Prediction using Machine Learning" International Advanced Research Journal in Science, *Engineering and Technology*, Vol. 9, Issue 6, June 2022, pp. 781-787
2. Mitchell, T. M. (1999). "Machine learning and data mining". *Communications of the ACM*
3. MacKay, D.J, & MacKay, D. J. (2003). "Information theory, inference and learning algorithms", *Cambridge University press*

Contents available at www.balwantcollege.edu.in
**Emerging Trends in Applied Mathematics and Statistics**
***(Balwant College, Vita (M.S.) India 415 311)***

# Factors Affecting on the Production of Sugarcane

Swarupa Gargate[1*]
[1]*Assistant Professor, Department of Statistics, Balwant College, Vita*

**Abstract:** Sugarcane is regarded as essential crop worldwide due to its extensive use in day to day lives of people and its industrial intended for dietary and economic sustenance. Study was conducted to analyze the effectiveness of various factors that are responsible for sugarcane production. Study discovers that cost invested for production of yield i.e. Urea, Land Preparation, Pesticides, Herbicides and cost of irrigation were important aspects which influenced on the returns of sugarcane growers.

**Keywords**: Sugarcane production, method of irrigation, Kolmogorov Smirnov Test.

## 1. Introduction:

Sugarcane is oldest crop known to man and it is most important cash crop of India. Sugarcane provides raw material for second largest agro based industries after textile. It is full of antioxidants that are essential to building and maintaining a healthy immune system. Antioxidants help combat free radicals (molecules that cause damage to cells) that can worsen several medical problems like diabetes, malaria, myocardial infarction, and skin cancer. Sugarcane is long duration crop and faces various abiotic traces like shortage of water, high temperature during summer, low temperature during winter, flooding during rainy season, nutritional stress, salinity, alkalinity and biotic stresses like fungal diseases as red dots, smut, wint, locary mawa etc. Nkosingiphile Samuel Zulu et al. (2019) studied factors affecting sugarcane production by small-scale growers in Ndwedwe Local Unicipality, South Africa.
Approximately 80% of the world's sugar is produced from sugar cane in tropical and subtropical climates. The remaining 20% comes from sugar beets, which are grown mostly in the temperate zones of the Northern Hemisphere. A total of over 120 countries produce sugar**.** India is second largest country in sugarcane production in the world. Improving sugarcane production will greatly help in economic prosperity of the farmer's and stakeholder's associated with sugarcane cultivation. In India sugarcane cultivated over an area of 5.228 million hector, with an annual production of 355 million tons and productivity 85.98 ton per hectare. In Tropical zone Maharashtra is the major sugarcane growing state covering about 9.4 lakh ha area with production of 61.32 Million ton.

## 2. Methods and Materials:

This study is done on primary data. The data is taken from farmers in Bhilawadi area. There are 1443 total families in Bhilawadi region .The total data collected from farmers is 303

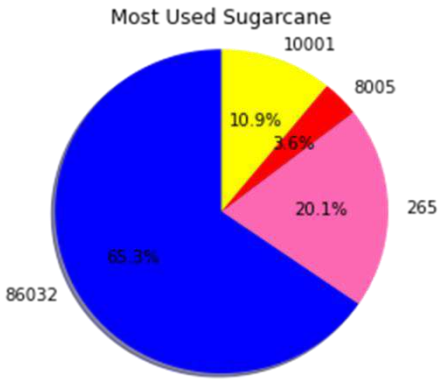
***Corresponding Author:** Miss. Swarupa Gargate, Assistant Professor, Department of Statistics, Balwant College, Vita (M.S.) India 415 311. **E-mail:** swarupagargate@gmail.com*

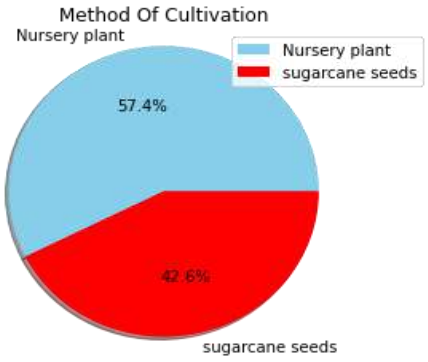
by using sampling method. There are two cultivation method: Nursery Plant and Cane sugar seed's. There are different types of cane sugar as Co-8371(Bhima), Co-85004(Prabha), Co-86032(Nayana), Co-87025(Kalyani), Co-87263(Sarayu) etc. Crop type taken for these project are 86032,10001,8005,265. Type cane sugar.

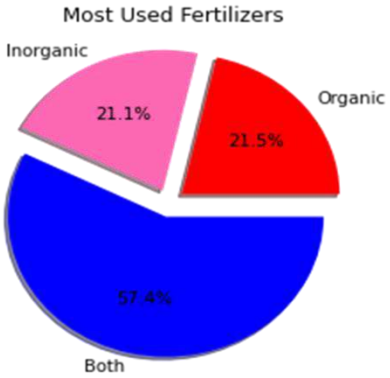
**3. Statistical Analysis**:

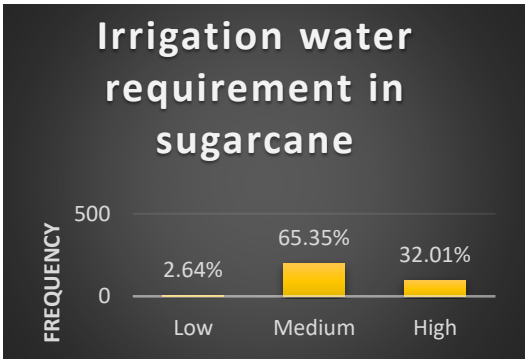
The statistical analysis includes graphical representation and statistical tests as follows.

**3.1. Graphical Representation:**



### 3.2. Statistical Tests:

❖ **Kolmogorov- Smirnov Test**

**Hypothesis:**

$H_0$: There is no significant difference between average yield of rainy and winter season.

$H_1$: There is significant difference between average yield of rainy and winter season.

**Formula:**

Cal D = max|$S_n(x) - S_n(Y)$|

Tab D = $1.36*\sqrt{\frac{n1+n2}{n1*n2}}$

**Observation Table:**

| X (Rainy) | LCF(X) | Sn(X) | Y (Winter) | LCF(Y) | Sn(Y) | $\|(S_n(X)-S_n(Y)\|$ |
|---|---|---|---|---|---|---|
| 21 | 21 | 0.1826 | 57 | 57 | 0.3031 | 0.1205 |
| 53 | 74 | 0.6434 | 83 | 140 | 0.7446 | 0.1012 |
| 29 | 103 | 0.8956 | 33 | 173 | 0.9202 | 0.0245 |
| 12 | 115 | 1 | 15 | 188 | 1 | 0 |

**Calculation:**

Cal D = max|$S_n(x) - S_n(Y)$|

**Cal D = 0.1205**

Tab D = $1.36*\sqrt{\frac{n1+n2}{n1*n2}}$

= $1.36*\sqrt{\frac{115+188}{115*188}}$

= $1.36*\sqrt{\frac{303}{21620}}$

Tab D = 1.36*0.1184

**Tab D = 0.1610**

**Result:** Cal D < Tab D**:** Hence we fail to reject $H_0$ at 5% level of significance.

❖ **Mean Test For Two Sample**

**Hypothesis:**

$H_0$: There is no significant difference between average yield of 86032 and 265 type of sugarcane. i.e. $H_0$: $\mu_1=\mu_2$.

$H_1$: There is significant difference between average yield of of 86032 and 265 type of sugarcane.

i.e. $H_1$: $\mu_1 \neq \mu_2$

Here,

n1=size (86032) = 198

n2=size (265) = 61

μ1=average yield of 86032 type =1.7075

μ2=average yield of 265 type =1.2950

v1=variance of 86032 type =0.4923

v2=variance of 265 type =0.2731

**Calculation:**

Then calculated Z,

$$Z=\frac{(\overline{X_1}-\overline{X_2})}{\sqrt{\frac{\sigma 1^2}{n1}+\frac{\sigma 2^2}{n2}}} \sim N(0,1)$$

Cal $|z|$ = 4.9427, Tab z = 1.96 .

Cal $|z|$>Tab $|z|$

**Result:**

Here, Cal $|z|$is greater than then Tab $|z|$ hence we reject $H_0$ at 5% level of significance.

**4. Conclusions:**

1. Out of 303 famers 65.3% farmer used 86032 type of sugarcane, 20.1% farmers used 265 type of sugarcane, 10.9% used 10001 Type and 3.6% farmers used 8005 Type of sugarcane in Bhilawadi village.
2. Almost 57.4% farmers used Both (Organic&Inorganic) Fertilizers and it also increases Average yield.
3. There are mainly two methods of cultivation used in village, in which Nursery plant is used by 57.4% of farmer &Sugarcane seed is used by 42.6% of farmer which is useful for increasing yield but Seed germination is only 70% then the seeds have to be sown again in non-germinated areas and that is costly so Nursery plants are more used than Sugarcane seeds.
4. Almost 65.35% farmers are of the opinion that sugarcane crop needs medium (moderate)
   Amount of water.
5. There is no significant difference between Average yield of Rainy and winter season. The two attributes type of sugarcane and Average yield are not independent i.e. dependent.
6. There is significant difference between average yield of 86032 and 265 type of sugarcane.

**References:**

1. Nkosingiphile Samuel Zulu, Melusi Sibanda and Bokang Stephen Tlali (2019) "Factors Affecting Sugarcane Production by Small-Scale Growers in Ndwedwe Local Unicipality, South Africa", MDPI, *9*(8), 170.
2. Narayan, P.K. An empirical analysis of sugarcane production in Fiji, 1970–2000. Econ. Anal. Policy 2004, 34, 53–62.
3. http://agropedia.iitk.ac.in/content/varieties-sugarcane
4. https://sugarcane.icar.gov.in/index.php/en/resrch/popular-promising- varieties.

Contents available at www.balwantcollege.edu.in
**Emerging Trends in Applied Mathematics and Statistics**
***(Balwant College, Vita (M.S.) India 415 311)***

# A study of risk factors affecting on Heart Attack

Rupali Suryawanshi[1*], Pallavi Sakhare[2], Abhijeet Jadhav[3], Prasad Gavali[4]
*[1,2,3] PG Student, Department of Statistics, Balwant College, Vita*
*[4]Assistant Professor, Department of Statistics, Balwant College, Vita*

**Abstract:** The goal of present research paper was to find the significant factors affecting on heart attack and to predict the pulse rate of patients by using the independent variables like gender, age, area of living, oxygen level, high blood pressure, diabetes of patients who are suffering for heart attack. The study was carried out among 100 heart attack patients. With the help of multiple linear regression, we found that the variables gender, age, high blood pressure and diabetes are significantly contribute to predict the pulses of heart attack patients. Thus, the most important risk factors that affect the heart attack are gender, age, high blood pressure and diabetes.

**Keywords:** Heart attack, diabetes, high blood pressure, regression.

## 1. Introduction:

Heart attack is very common disease in today's life. It is also called as cardiovascular disease. Heart attack happens when blood flow to the heart is blocked or cut off. If there is not sufficient oxygen rich blood flowing to the heart, it can cause damage to the affected area. As a result, the muscles begins to die. When your heart isn't getting the blood and oxygen it needs to function properly, it can put you at a higher risk of heart failure and other serious complications.

There are so many major risk factors can affect the heart attack or heart beats, such as high blood pressure, Diabetes, Cholesterol level, Obesity, …etc. High blood pressure is very common risk factors can affect the heart disease. When your blood pressure is high, it makes your heart have to work harder. i.e. Your heart works harder to pump blood. High blood pressure can strain your heart, damage blood vessel and increase your risk of heart attack.

Diabetes is most risky factor to damage heart, especially it is not controllable. Nearly 68% of diabetes over the age of 65 die from heart attack. The older you get, the more you are at risk of heart attack. Although you have a heart attack at any age the risk increases significantly after age of 45 for men and 50 for women.

Kannel WB et al. (1971) carried out the comparative study of systolic versus diastolic blood pressure and risk of coronary heart disease. Peter W. F. et al. (1998) predicted the Coronary Heart Disease Using Risk Factor Categories. Present study deals about the prediction of pulse of patients by using the independent variables like gender, age, area of living, oxygen level, high blood pressure, diabetes of patients who are suffering for heart attack and to find the risk factors affecting on heart attack.

***Corresponding Author:** Miss. Rupali Suryawanshi, PG student, Department of Statistics, Balwant College, Vita (M.S.) India 415 311. **E-mail:** rupali2910sury@gmail.com*

## 2. Methods and Materials:

The data of patients who are suffering from heart attack are collected from "Suvidha hospital" Atpadi, Dist: Sangli. The study of heart attack patients are carried out over 100 patients among 71 are males and 29 are females. The collection takes place through the patients register with in Hospital. Multiple linear regression were used to predict the heart pulses. The analysis was carried out with the help of SPSS software.

## 3. Statistical Analysis:

The descriptive statistics such as N, Minimum, Maximum, Mean, Std. deviation for the continuous variable pulse rate, diabetes, high blood pressure (HBP), and Oxygen level are shown in Table-3.1. For the variable pulse rate, the minimum pulse rate is 38, maximum pulse rate is 140. The mean of pulse rate is 80.66 with standard deviation 20.345.

**Table-3.1: Descriptive Statistics**

| | N | Minimum | Maximum | Mean | | Std. Deviation |
|---|---|---|---|---|---|---|
| | Statistic | Statistic | Statistic | Statistic | Std. Error | Statistic |
| Pulse | 100 | 38 | 140 | 80.66 | 2.034 | 20.345 |
| Age | 100 | 24 | 85 | 63.00 | 1.148 | 11.476 |
| Diabetes | 100 | 73 | 534 | 179.10 | 8.835 | 88.348 |
| HBP | 100 | 70 | 220 | 131.99 | 2.740 | 27.398 |
| Oxygen level | 100 | 52 | 100 | 95.81 | .662 | 6.624 |

In this study, the dependent variable is pulse rate of patients who are suffering from heart attack and the predictors are gender, age, area of living, oxygen level, high blood pressure, diabetes of patients who are suffering for heart attack. To do multiple linear regression, the basic assumption is that, the dependent variable should be normally distributed. Therefore, we have to check the normality of dependent variable pulse rate. The null and alternative hypotheses for normality as follows:

$H_0$: Pulse rate of patients follows normal distribution. Vs

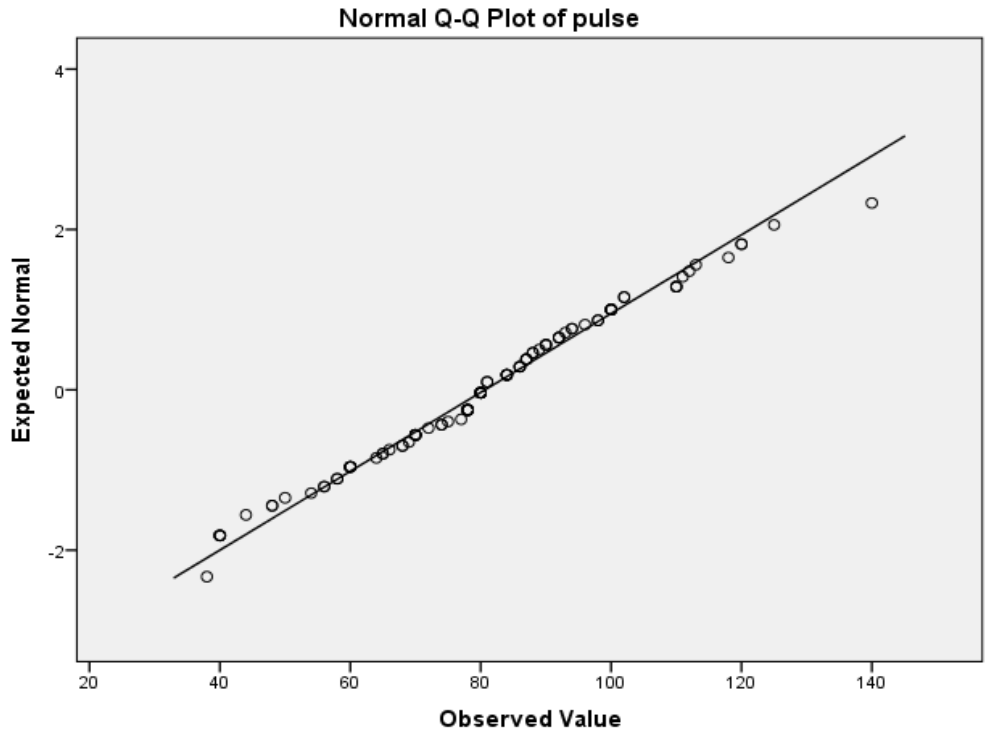
$H_1$: Pulse rate of patients does not follow normal distribution.

The Kolmogorov-Smirnov and Shapiro-Wilk tests of normality are given in Table-3.2. The p-value of Shapiro-Wilk test is greater than 0.05 (level of significance). Hence we fail to reject the null hypothesis and conclude that the pulse rate of patients follows normal distribution.

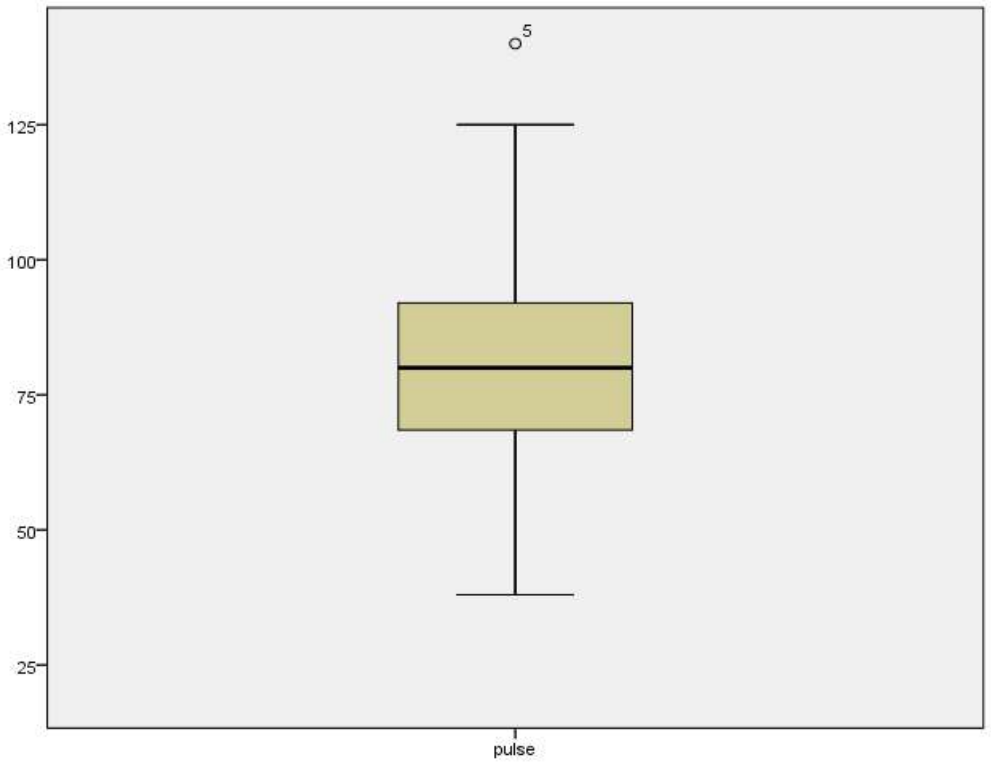
**Table-3.2: Tests of Normality**

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| Pulse rate | .088 | 100 | .054 | .986 | 100 | .359 |
| a. Lilliefors Significance Correction | | | | | | |

The Normal Q-Q plot of pulse shows that all quantile points lie on the theoretical straight line. Hence it indicates that the normality assumption is satisfied.




The Box-Plot shown in below graph shows the five number summary including minimum, first quartile, median, third quartile and maximum. It also displays the outlier if it exists in the data. From the Box-Plot, we can see that the median is in middle of Box and whiskers are about the same on both sides of box. Hence the distribution of pulse rate of patients is Normal (symmetric). It also represent the fifth observation is outlier.



**Multiple Linear Regression:**

To predict the pulse rate of patients by using the predictors like gender, age, area of living, oxygen level, high blood pressure, diabetes of patients who are suffering for heart attack the multiple linear regression is used.

To select final model, the "backward elimination" method is applied. In backward elimination method, we have to set a significance level for deciding when to remove a predictor into the model called as alpha-to remove and it set as $\alpha_R$= 0.10. For this study the backward elimination method builds three models (Table-3.5). In first model, all the predictors are included. The predictor oxygen level has highest p-value (0.181) which is greater than 0.10 is removed from first model, then the new model fitted called as model 2. Now the p-value of predictor area is 0.150 which is greater than 0.10, hence the predictor area is removed from the second model and build new model is called model 3. In model 3, the p-value for all the predictors are less than 0.10. Hence the model 3 is confirmed as final model. The model summary Table -3.3, gives the value of R, R-square, Adjusted R-square, Std. Error of the Estimate for all three models.

**Table-3.3: Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .531[a] | .282 | .236 | 17.787 |
| 2 | .518[b] | .268 | .229 | 17.863 |
| 3 | .502[c] | .252 | .220 | 17.967 |
| a. Predictors: (Constant), diabetes, oxygen level , age, HBP, Area, gender | | | | |
| b. Predictors: (Constant), diabetes, age, HBP, Area, gender | | | | |
| c. Predictors: (Constant), diabetes, age, HBP, gender | | | | |

Now we can interpret the results of model 3 because it is final model for this study. The ANOVA Table-3.4 (model 3) indicates that, the p-value of F test 0.000 which is too small. It is indicating that the model fit better.

**Table-3.4: ANOVA**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 11554.941 | 6 | 1925.824 | 6.087 | .000[b] |
| | Residual | 29421.499 | 93 | 316.360 | | |
| | Total | 40976.440 | 99 | | | |
| 2 | Regression | 10981.489 | 5 | 2196.298 | 6.883 | .000[c] |
| | Residual | 29994.951 | 94 | 319.095 | | |
| | Total | 40976.440 | 99 | | | |
| 3 | Regression | 10310.153 | 4 | 2577.538 | 7.985 | .000[d] |
| | Residual | 30666.287 | 95 | 322.803 | | |
| | Total | 40976.440 | 99 | | | |

| |
|---|
| a. Dependent Variable: pulse rate |
| b. Predictors: (Constant), diabetics, oxygen level , age, HBP, Area, gender |
| c. Predictors: (Constant), diabetics, age, HBP, Area, gender |
| d. Predictors: (Constant), diabetics, age, HBP, gender |

The Coefficient Table-3.5 gives the Unstandardized Coefficients with standard error, value of t test statistic and their respective p-value for each predictor for all three models. The final model for this study is model 3. Therefore, we focus on the results of model 3.

**Table-3.5: Coefficients**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | T | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 105.934 | 30.321 | | 3.494 | .001 |
| | gender | -7.935 | 4.297 | -.178 | -1.847 | .068 |
| | Area | -5.703 | 4.210 | -.124 | -1.355 | .179 |
| | Oxygen level | -.369 | .274 | -.120 | -1.346 | .181 |
| | age | -.441 | .162 | -.249 | -2.719 | .008 |
| | HBP | .295 | .068 | .397 | 4.355 | .000 |
| | diabetes | .034 | .022 | .148 | 1.564 | .121 |
| 2 | (Constant) | 70.766 | 15.463 | | 4.577 | .000 |
| | gender | -7.571 | 4.307 | -.170 | -1.758 | .082 |
| | Area | -6.116 | 4.216 | -.133 | -1.450 | .150 |
| | age | -.465 | .162 | -.262 | -2.866 | .005 |
| | HBP | .299 | .068 | .403 | 4.409 | .000 |
| | diabetes | .037 | .022 | .160 | 1.690 | .094 |
| 3 | (Constant) | 64.445 | 14.922 | | 4.319 | .000 |
| | gender | -7.730 | 4.330 | -.173 | -1.785 | .077 |
| | age | -.434 | .162 | -.245 | -2.686 | .009 |
| | HBP | .315 | .067 | .424 | 4.667 | .000 |
| | diabetes | .042 | .022 | .182 | 1.933 | .056 |
| a. Dependent Variable: pulse rate | | | | | | |

## 4. Result and discussions:

In model -3, the p-values of the predictors gender, age, HBP and diabetes are less than 0.10 ($\alpha_R$= 0.10). Thus the predictor's gender, age, HBP and diabetes are contributing significantly for predicting the pulse rate of patients who are suffering from heart attack. Thus, the final model for prediction is:

$$\widehat{pulse} = 64.445 - 7.730\ gender - 0.434\ age + 0.315\ HBP + 0.042\ diabetes$$

**Illustration:** We want to predict the pulse rate of 55 year old male patients who is suffering from heart attack with high blood pressure and was diabetes 110, 144 respectively.

$$\widehat{pulse} = 64.445 - 7.730(1) - 0.434(55) + 0.315(110) + 0.042(144)$$

$$\widehat{pulse} = 73.543$$

## 5. Conclusions:

The result of present study shows that the age, gender, high blood pressure, and being diabetes are highly significant risk factors for the patients suffering from heart attack. Doctors can use this model for the prediction of pulse rate based on predictor's gender, age, HBP and diabetes. In future, researcher can use the logistic regression to predict the chance of heart attack.

## References:

1. Hajar R. Risk Factors for Coronary Artery Disease: Historical Perspectives. Heart Views. 2017, Jul-Sep;18(3):109-114.
2. Gupta R, Joshi P, Mohan V, Reddy KS, Yusuf S. Epidemiology and causation of coronary heart disease and stroke in India. Heart. 2008;94:16–26
3. Peter W. F. Wilson, Ralph B. D'Agostino, Daniel Levy, Albert M. Belanger, Halit Silbershatz and William B. Kannel. Prediction of Coronary Heart Disease Using Risk Factor Categories, AHA journals, 1998, Vol. 97, No. 18, 1837-1847.
4. Kannel WB, Gordon T, Schwartz MJ. Systolic versus diastolic blood pressure and risk of coronary heart disease. The Framingham study. Am J Cardiol. 1971 Apr; 27(4):335-46.

Contents available at ***www.balwantcollege.edu.in***

**Emerging Trends in Applied Mathematics and Statistics**

***(Balwant College, Vita (M.S.) India 415 311)***

# Profit Prediction of Raisin Production: Special reference to Sangli District

Aishwarya Bhosale[1], Bhagvan Narale[2*]

*[1,2]PG Student, Department of Statistics, Balwant College, Vita*

**Abstract:** The partially dried grapes are called raisins. In state of Maharashtra, especially in Sangli district, the varieties, namely, Anushka, Clone, Manik Chaman, S.S.N., Thompson, Super Sonaka are used to produce raisins. The period required to produce raisins after pruning is between 130 days to 190 days. Tasgaon is an important supplier of raisins in Maharashtra state. Sangli district has favorable climate condition for the crop and soil fertility. Therefore, grapes are grown as a major crop in Sangli. Raisin have great demand at weddings and festival. The sale of good quality raisins ranges from Rs. 160 to Rs. 250 per kg. After selling raisin in the market, the profit to the farmers should be satisfactory. In present research we predicted profit of raisin using predictors such as variety of raisin (Variety), no of days required for raisin (Days) , production area (Acre), production place (Tahsil), selling price (Rate), Sulfur furnace, expenditure, production(kg) of raisin through multiple regression model.

**Key Words:** Raisin, Sulfur furnace, Production, Expenditure, Profit, Regression

## 1. Introduction:

Raisins or currants are semi-dried grapes. Currants are also called raisins in some places. Currants fall under the category of semi-dry fruits. After pruning, the grapes are formed and become bigger. The grapes are harvested after they have matured, then the grapes are kept in sheds to dry and the whole process takes 130 days to 190 days to form raisins. During this period, diseases also occur on the grapes, to protect grapes from such type of diseases, many expensive drugs and chemicals are sprayed on the grapes. Also shed costs a lot to dry those grapes. After the currants are prepared, they are given a Sulfur furnace, which also costs a lot of money to the farmers. After spending all this, currants are produced and then after selling them, the farmers get their compensation.

Tasgaon- Sangli is famous in Maharashtra for delectable raisins. The Maharashtra state ranks first in terms of production of grapes and annually exports more than 1.22 lakh tonnes of grapes. The approximate annual production of raisins in Sangli is 1.25 lakh tonnes of which 30 – 40 percent is exported. A big auction market of raisins was started in 1994 in Tasgaon which now has more than 100 trading centres and 60 – 65 cold storage warehouses.

M.G. Kerutagi et al. (2018) carried out economic analysis for production and marketing of raisins in Vijayapura District. Dr. J. G. Mulani (2018) studied raisins marketing in Sangli agriculture produce market committee and discussed about prices of raisin in Sangli APMC.

**Corresponding Author:** *Mr. Bhagvan Narale, PG Student, Department of Statistics, Balwant College, Vita (M.S.) India 415 311.* ***E-mail:*** *bhagvannarale6015@gmail.com*

M.S. Ladaniya et al. (2005) recorded negative correlation co-efficient in monthly prices and arrivals at national level and major markets in Maharashtra. Mrs. Dhanashri Y Jadhav et al. (2021) discussed problems and prospectus of raisin production in Sangli district of Maharashtra. The main purpose of this study is to fit multiple linear regression on profit of raisin.

## 2. Methods and Materials:

The present study is based on the primary data. The primary data were collected from farmers of five tahasils of Sangli district through GOOGLE form. Out of 216 respondents, 126, 37, 17, 20, 16 from Juth, Tasgaon, Atpadi, Kavathe Mankhal and Miraj are respectively. The Statistical technique called as multiple linear regression used to predict the profit of raisin. The data analysis has been done in SPSS.

## 3. Statistical analysis:

For this study, the response variable is profit and predictors are variety of raisin (Variety), no of days required for raisin (Days), production area (Acre), production place (Tahsil), selling price (Rate), Sulfur furnace, expenditure, production (kg) of raisin. In this predictors, variety of raisin (Variety), production place (Tahsil) and Sulfur furnace are categorical variable. To do multiple linear regression, we need to code this categorical variable. Thus, the variable coded variety of raisin (Variety) as Anushka (1), Clone 2 (2), Manik Chaman (3), SSN (4), Super Sonaka (5) and Thompson (6). The variable Tahsil coded as Juth (1), Tasgaon (2), Atpadi (3), Kavathe Mankhal (4) and Miraj (5). The variable Sulfur furnace coded as, No (0) and Yes (1). The multiple linear regression model that there is linear relationship between one dependent variable and two or more independent variables. The general form of multiple linear regression described as follows:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + - - - + \beta_k x_k + e_i$$

Where, $y_i$ : Value of i$^{th}$ observation for the dependent variable.

$\beta_j$ : Value of the j$^{th}$ coefficient, j = 1, 2, …., k

$x_j$ : j$^{th}$ independent variable, , j = 1, 2, …., k

$e_i$: Error in the observation value for the i$^{th}$ case.

**Hypothesis for overall F test:**

$H_0$: $\beta_1 = \beta_2 =$ --- $= \beta_8 = 0$

$H_1$: At least one $\beta_i^s$ are different from zero.

In this research study, we want to establish the relationship between profit from raisin and variety of raisin (Variety), no of days required for raisin (Days) , production area (Acre), production place (Tahsil), selling price (Rate), Sulfur furnace, expenditure, production(kg) of raisin. Model summary (Table-3.1) provides the values of R, R-square, Adjusted R-square and Std. Error of the Estimate for the fitted model. The value of R-square 0.981, indicates that 98.1% variation in profit of raisin explained by independent variables, namely, Rate, Days, Variety, Acre, Sulfur furnace, Tahsil, Expenditure, production(kg).

**Table-3.1: Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .990[a] | .981 | .980 | 87374.025 |
| a. Predictors: (Constant), Rate, Days, Variety, Acre, Sulfur furnace, Tahsil., Expenditure, production(kg) | | | | |

The ANOVA Table -3.2 contains Sum of Squares (SS), Degrees of Freedom (df), Mean Square (MS), F-test statistic and P-value (Sig.). The value of F statistic is 1332.197. The p-value for F statistic is 0.000 which is less than 0.01 (level of significance). Hence we reject the null hypothesis at 1% level of significance and conclude that least one $\beta_i^s$ are different from zero. It indicates that fit of model is better. (i.e. The model is useful for prediction of profit).

**Table-3.2: ANOVA**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 81362299481300.830 | 8 | 10170287435162.604 | 1332.197 | .000[b] |
| | Residual | 1580283584949.175 | 207 | 7634220217.146 | | |
| | Total | 82942583066250.000 | 215 | | | |
| a. Dependent Variable: Profit | | | | | | |
| b. Predictors: (Constant), Rate, Days, Variety, Acre, Sulfur furnace, Tahsil, Expenditure, production(kg) | | | | | | |

The coefficient Table-3.3, gives the value of regression coefficients for each predictor with their standard error. It also includes t test statistic and p-value (Sig.) for each predictor.

**Table-3.3: Coefficients**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | -979903.343 | 113445.101 | | -8.638 | .000 |
| | Tehsil | 2090.689 | 4806.082 | .004 | .435 | .664 |
| | Variety | -1182.593 | 3836.424 | -.003 | -.308 | .758 |
| | Acre | 2669.047 | 5939.418 | .007 | .449 | .654 |
| | Production(kg) | 168.961 | 2.746 | 1.051 | 61.527 | .000 |
| | Expenditure | -.876 | .084 | -.165 | -10.408 | .000 |
| | Sulfur furnace | 20061.386 | 14817.861 | .013 | 1.354 | .177 |
| | Days | 186.903 | 669.729 | .003 | .279 | .780 |
| | Rate | 5432.903 | 250.387 | .221 | 21.698 | .000 |
| a. Dependent Variable: Profit | | | | | | |

## 4. Result and discussions:

The general form of the equation to predict profit of raisin is:

$\widehat{Profit}$ = -979903.343 + 2090.689Tehsil -1182.593Variety + 2669.047Acre + 168.961 Production (kg) -.876Expenditure + 20061.386Sulfur furnace + 186.903Days + 5432.903Rate

The Unstandardized coefficients indicate how much the dependent variable varies with an independent variable when all other independent variables are held constant.

The unstandardized coefficient of Tahsil is 2090.689. This indicates that for each one unit increase in Tahsil, there is increase in profit of Rs. 2090.689. (e.g. Here we coded the tahsil variable as: Jath (1), Tasgaon (2), Atpadi (3), Kavathe Mahankal (4) and Miraj (5). Thus, the place of production (Tahsil) changed Jath to tasgaon, the profit increases by Rs.2090.689 or the place of production (Tahsil) changed Jath to Atpadi, the profit increases by Rs.4181.378 when other variables held constant).

The unstandardized coefficient of Acre is 2669.047. This indicates that for each one unit increase in Acre, there is increase in profit of Rs. 2669.047 when other variables held constant.

The statistical significance of each predictors is tested by using t-test statistic. The predictor which has p-value of t test statistic is less than or equal to 0.01 are statistically different from zero. However, those predictors p-value of t test statistic is greater than 0.01 are statistically not different from zero.

**Illustration:** A farmer belongs to Jath has 3 acres of manik chaman. He produced 8000 kg of raisin and it cost him Rs. 400000. He was given a sulfur furnace for raisin, costing him 155 days to produce raisin and he got rate Rs.180 for per kg then the predicted profit of raisin is Rs. 1054889.

i.e. $\widehat{Profit}$ = *-979903.343 + 2090.689(1) -1182.593(3) + 2669.047(3) + 168.961(8000) -.876(400000) + 20061.386(1) + 186.903(155) + 5432.903(180)*

$\widehat{Profit}$ = 1054889 Rupees.

## 5. Conclusions:

Farmers spend money on crops without any thought of profit and this can lead to huge losses. To overcome this loss of farmer, we fitted multiple linear regression model to predict the profit of Raisin. So farmers can use this model, to spend less money on crop and earn more profit. In future, the researcher will maximize the raisin profit with the help of Optimization techniques.

## References:

1. M.G. Kerutagi, A.S. Pavithra and Shweta Byahatti (2018). "Production and Marketing of Raisins in Vijayapura District - An Economic Analysis", *International Journal of Current Microbiology and Applied Sciences,* Volume 7 Number 09, 658-671
2. Dr. J. G. Mulani (2018). "A Study Of Raisins Marketing In Sangli Agriculture Produce Market Committee", *Review Of Research,* Valume-7, Issue-6, 1-4.
3. M.S. Ladaniya, Vinod Wanjari and Bipinchandra Mahalle (2005), "Marketing of Grapes and Raisins and Post-Harvest Losses of Fresh Grapes in Maharashtra" *Indian J. Agric. Res.*, 39,167 – 176.
4. Mrs. Dhanashri Y Jadhav and Dr. Yuvraj S Jadhav (2021) "Problems And Prospectus Of Raisin Production In Sangli District Of Maharashtra", *International Journal Of Multidisciplinary Educational Research,* VOLUME:10, ISSUE:1(4), 17-18

Contents available at www.balwantcollege.edu.in
**Emerging Trends in Applied Mathematics and Statistics**
***(Balwant College, Vita (M.S.) India 415 311)***

# Prediction of Chance of Covid-19 after First Vaccination by using Binary Logistic Regression

Pratik Kadam[1*], Ajay Jadhav[2]
[1]*PG student, Department of Statistics, Sardar Patel University, Anand*
[2]*Assistant Professor, Department of Statistics, Balwant College, Vita*

**Abstract:** The goal of present research was to predict the covid-19 infection after first dose of vaccine. The study was carried out among 306 peoples in Khanapur Tehsil. To analyze data, we applied binary logistic regression model. Result indicates that, age of person increases, the chance of having infection of Covid-19 disease increases. The Covaxin is more efficient than Covishield.

**Keywords:** Binary Logistic Regression, model, Covid-19, Covishield, Covaxin, Sputnik-V.

**1. Introduction:** Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-CoV-2 virus. Equitable access to safe and effective vaccines is critical to ending the COVID-19 pandemic, so it is hugely encouraging to see so many vaccines proving and going into development. Vaccination is a simple, safe and effective way of protecting you against harmful diseases, before you come into contact with them. It uses your body's natural defenses to build resistance to specific infections and makes your immune system stronger.

Vaccines train your immune system to create antibodies, just as it does when it's exposed to a disease. However, because vaccines contain only killed or weakened forms of germs like viruses or bacteria, they do not cause the disease or put you at risk of its complications. Covaxin is India's first indigenous covid-19 vaccine, Covishield and Sputnik-V is also approved in India.

Saad E, et al. (2022) predicted the death risk analysis in fully vaccinated people using novel extreme regression-voting classifier. Experton B, et al. (2021) studied the predictive model for severe COVID-19 in the medicare population which is a tool for prioritizing primary and booster covid-19 vaccination.

We conduct study to know possibility of Covid-19 disease after taking the first dose of vaccination. To study the efficacy of vaccine and check which is better among studied vaccines.

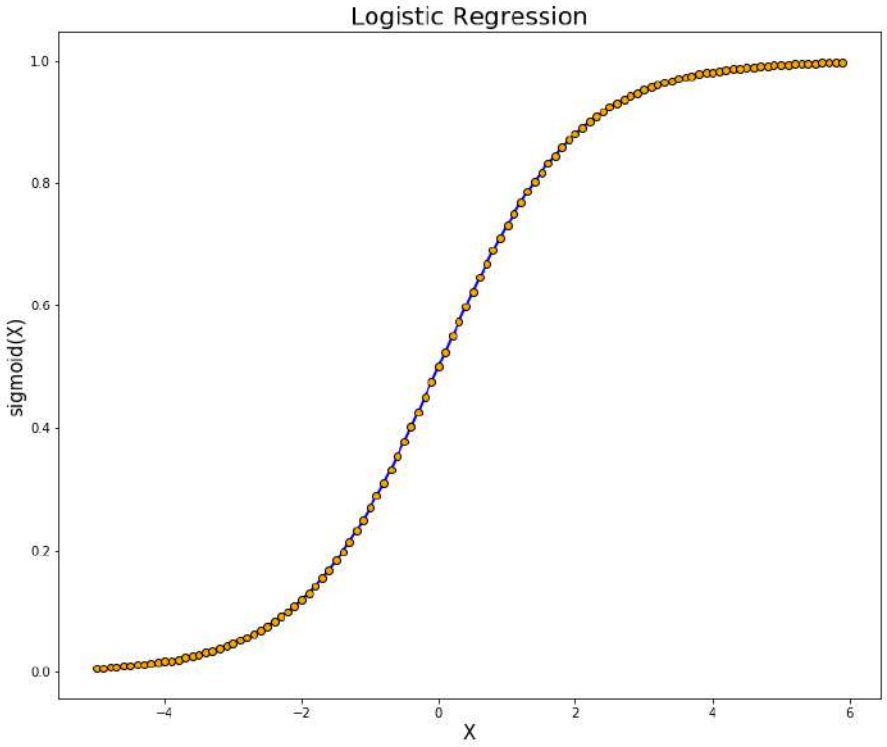
**2. Methods and Material:** The present study was carried out among 306 people lived in Khanapur Tehsil. It is Primary data collected randomly and voluntarily through Google form. The study was conducted in current year March, 2022. The Binary logistic regression was used to predict the chance of covid-19 after first dose. The Statistical software SPSS is used to analyze the data.

***Corresponding Author:** Mr. Pratik Kadam, PG Student, Department of Statistics, Sardar Patel University, Anand, Gujarat. **E-mail:** 2002pratikkadam@gmail.com*

**Logistic Regression:**

Binary logistic regression is useful where the dependent variable is dichotomous (e.g., succeed/ fail, live/die, graduate/dropout, vote for A/B, Yes/No, etc.). We may be interested in predicting the likelihood that a new case will be in one of the two outcome categories. The following is graph of logistic regression:

.




Logistic regression from this model by creating a new dependent variable, the logit(P). If P is the probability of a 1 at any given value of X, the odds of a 1 vs. a 0 at any value for X are P(1-P). The logistic is the natural log of this odds ratio.

Definition: Logit(P) = ln(odds) = ln[P(1-P)]

This looks ugly, but it leads to a beautiful model. In logistic regression, we solve for logit(P) = a+bX, where logit(P) is a linear function of X, very much like ordinary regression solving for Y.

**3. Statistical analysis:** The Binary Logistic regression is used when our outcome variable is binary and we won't to establish the relationship between outcome and predictor variables. Thus, we can write, binary logistic regression model Y on $x_1, x_2, \ldots, x_k$ as,

$$Log\left(\frac{Pi}{1-Pi}\right) = \beta_0 + \beta_1 x1 + \beta_2 x2 + \beta_3 x3 + \ldots\ldots + \beta_k xk \qquad \text{----------(1)}$$

$$\frac{Pi}{1-Pi} = Exp(\beta_0 + \beta_1 x1 + \beta_2 x2 + \beta_3 x3 + \ldots\ldots + \beta_k xk)$$

$$\frac{1-Pi}{Pi} = \frac{1}{Exp(\beta_0 + \beta_1 x1 + \beta_2 x2 + \beta_3 x3 + \ldots\ldots + \beta_k xk)}$$

$$\frac{1}{Pi} - 1 = \frac{1}{Exp(\beta_0 + \beta_1 x1 + \beta_2 x2 + \beta_3 x3 + \ldots\ldots + \beta_k xk)}$$

$$\frac{1}{Pi} = 1 + \frac{1}{Exp(\beta_0 + \beta_1 x1 + \beta_2 x2 + \beta_3 x3 + \ldots\ldots + \beta_k xk)}$$

$$\frac{1}{Pi} = \frac{1 + Exp(\beta_0 + \beta_1 x1 + \beta_2 x2 + \beta_3 x3 + \ldots\ldots + \beta_k xk)}{Exp(\beta_0 + \beta_1 x1 + \beta_2 x2 + \beta_3 x3 + \ldots\ldots + \beta_k xk)}$$

Therefore,

$$Pi = \frac{Exp(\beta_0 + \beta_1 x1 + \beta_2 x2 + \beta_3 x3 + \ldots\ldots + \beta_k xk)}{1 + Exp(\beta_0 + \beta_1 x1 + \beta_2 x2 + \beta_3 x3 + \ldots\ldots + \beta_k xk)} \quad \text{----------(2)}$$

**SPSS output for Binary logistic regression:**

Dependent (outcome) Variable: Binary (CAFV=Yes/No)

Independent (explanatory) Variables: Any (Gender, Area, etc.)

The case processing summary (Table-3.1) gives the total no of cases available for logistic regression analysis and also the missing cases with their percentage.

**Table-3.1: Case Processing Summary**

| Unweighted Cases[a] | | N | Percent |
|---|---|---|---|
| Selected Cases | Included in Analysis | 306 | 100.0 |
| | Missing Cases | 0 | .0 |
| | Total | 306 | 100.0 |
| Unselected Cases | | 0 | .0 |
| Total | | 306 | 100.0 |
| a. If weight is in effect, see classification table for the total number of cases. | | | |

In table-3.2, the coding for the criterion variable is given first – the largest coded group (Yes=1) is identified as the "target".

**Table-3.2: Dependent Variable Encoding**

| Original Value | Internal Value |
|---|---|
| No | 0 |
| Yes | 1 |

The categorical Coding for all categorical predictors is specified in table-3.3: For example the variable, Vaccine type: Sputnik-V will be the comparison group, 1st parameter will be compared the Covishield to sputnik-V and 2nd parameter will be compared the Covaxin to sputnik-V.

**Table -3.3: Categorical Variables Coding's**

| | Frequency | Parameter coding | |
|---|---|---|---|
| | | (1) | (2) |

| | | | | |
|---|---|---|---|---|
| Vaccine Type | Covishield | 244 | 1.000 | .000 |
| | Covaxin | 60 | .000 | 1.000 |
| | Sputnik V | 2 | .000 | .000 |
| Area | Urban | 166 | 1.000 | |
| | Rural | 140 | .000 | |
| POV | Private | 8 | 1.000 | |
| | Government | 298 | .000 | |
| Diabetes | No | 265 | 1.000 | |
| | Yes | 41 | .000 | |
| Blood Pressure | No | 244 | 1.000 | |
| | Yes | 62 | .000 | |
| Gender | Female | 165 | 1.000 | |
| | Male | 141 | .000 | |

The full model is tested using a Likelihood Ratio (LR) test to see if it is a significant improvement (p-value < 0.05) on the null model in a 'Model' row of the 'Omnibus Tests of Model Coefficients' table. The model was statistically significant when compared to the null model, $X^2(8) = 99.059$, p-value (Sign) < 0.001

**Table -3.4: Omnibus Tests of Model Coefficients**

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 99.059 | 8 | .000 |
| | Block | 99.059 | 8 | .000 |
| | Model | 99.059 | 8 | .000 |

The variable in equation table-3.4 gives the value of coefficient (B), S.E., Wald test statistic with p-value and Exp(B).

**Table-3.4: Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | Age | .005 | .015 | .088 | 1 | .766 | 1.005 |
| | Gender (1) | -.489 | .328 | 2.224 | 1 | .136 | .613 |
| | Area (1) | 1.646 | .341 | 23.301 | 1 | .000 | 5.186 |
| | Vaccine Type | | | 2.283 | 2 | .319 | |
| | Vaccine Type (1) | -1.529 | 1.608 | .905 | 1 | .341 | .217 |
| | Vaccine Type (2) | -2.004 | 1.635 | 1.503 | 1 | .220 | .135 |
| | Diabetes (1) | -1.226 | .561 | 4.775 | 1 | .029 | .293 |
| | Blood Pressure (1) | -1.835 | .463 | 15.738 | 1 | .000 | .160 |
| | POV (1) | -1.429 | 1.326 | 1.162 | 1 | .281 | .240 |
| | Constant | 2.279 | 1.917 | 1.414 | 1 | .234 | 9.767 |

a. Variable(s) entered on step 1: Age, Gender, Area, Vaccine Type, Diabetes, Blood Pressure, POV.

## 4. Results and discussion:

A logistic regression was carried out to assess the effect of age, gender, area, vaccination type, Diabetes and BP of certain person on the likelihood of infection of Covid-19 Virus on that person after taking first dose. The binary logistic regression model to predict the chance of infection of covid-19 after first vaccination from variable in equation (table-3.4) given as:

$$\text{Log}\,(Pi/1\text{-}Pi) = \beta_0 + \beta_1\,\text{Age} + \beta_2\,\text{Gender} + \beta_3\,\text{Area} + \beta_4\,\text{V}\,(1) + \beta_5\,\text{V}\,(2) + \beta_6\,\text{Dia.} + \beta_7\,\text{BP} + \beta_8\,\text{POV}$$

Where, TOV – Type of Vaccine, Dia. – Diabetes, BP – Blood Pressure, POV – Place of Vaccination

Therefore, we can write [from equation (2)]

$$P_i = \frac{\text{Exp}(2.279+0.005\text{Age}-0.489\text{Gender}+1.646\text{Area}-1.529\text{V}(1)-2.004\text{V}(2)-1.226\text{Dia.}-1.835\text{BP}-1.420\text{POV})}{1+\text{Exp}(2.279+0.005\text{Age}-0.489\text{Gender}+1.646\text{Area}-1.529\text{V}(1)-2.004\text{V}(2)-1.226\text{Dia.}-1.835\text{BP}-1.420\text{POV})}$$

**Illustration 1:** The chance of Covid-19 after first vaccination for 45 year old male belongs to rural area, who had taken Covishield vaccine on government vaccination center and he is patient of diabetes is:

$$P_i = \frac{\text{Exp}(2.279+0.005(45)-0.489(0)+1.646(0)-1.529(1)-2.004(0)-1.226(0)-1.420(0))}{1+\text{Exp}(2.279+0.005(45)-0.489(0)+1.646(0)-1.529(1)-2.004(0)-1.226(0)-1.420(0))}$$

$P_i$ = **0.297339**

**Illustration 2:** The chance of Covid-19 after first vaccination for 45 year old male belongs to rural area, who had taken Covaxin vaccine on government vaccination center and he is patient of diabetes is:

$$P_i = \frac{\text{Exp}(2.279+0.005(45)-0.489(0)+1.646(0)-1.529(0)-2.004(1)-1.226(0)-1.420(0))}{1+\text{Exp}(2.279+0.005(45)-0.489(0)+1.646(0)-1.529(0)-2.004(1)-1.226(0)-1.420(0))}$$

$P_i$ = **0.208334**

**5. Conclusions:** The results indicate that, as age of person increases, the chance of having infection of Covid-19 disease increases. If a person has some diseases like diabetes, BP, etc. then the chances are also increases and vice-versa. The illustration of two examples reflects that the Covaxin is more efficient than Covishield. With the help of this study we can stop sudden surge of covid cases in the world by promoting people to take second dose and buster doses.

## References:

1. Saad E, Sadiq S, Jamil R, Rustam F, Mehmood A, Choi GS, Ashraf I. Predicting death risk analysis in fully vaccinated people using novel extreme regression-voting classifier. Digit Health. 2022 Jul 21;8
2. Experton B, Tetteh HA, Lurie N, Walker P, Elena A, Hein CS, Schwendiman B, Vincent JL, Burrow CR. A Predictive Model for Severe COVID-19 in the Medicare Population: A Tool for Prioritizing Primary and Booster COVID-19 Vaccination. Biology (Basel). 2021 Nov 15;10(11):1185.
3. Binary Logistic Regression by Mark Tranmer and Mark Elliot (Cathie Marsh Centre for Census and Survey Research).
4. Introduction to Binary Logistic Regression and Propensity Score Analysis by Dale E. Berger from Claremont Graduate University.

Contents available at www.balwantcollege.edu.in
**Emerging Trends in Applied Mathematics and Statistics**
*(Balwant College, Vita (M.S.) India 415 311)*

# View of Businessmen towards G.S.T.: A special reference to Tasgaon city

Sandhya Ghail[1*]
[1]*Assistant Professor, Department of Statistics, Balwant College, Vita*

**Abstract:** The GST has been launched by 1st July 2017. Atal Bihari Vajpayee is known as father of GST. It is the only indirect tax that directly affects all sectors and sections of our economy. This study was conducted to check awareness of GST. After GST, taxes in Agri, Cloth, Gold, is maximum than previous taxes. Also to analyze proportion of mode of GST paying and types of area. In GST there are three types, namely, CGST, SGST and IGST. This study discovers that CGST & SGST type is mostly used in business. Assam is the first state of India to pass GST Bill. This survey is limited to main Tasgaon area.

**Key Words:** GST, Tax, Businessman, Paying.

**1. Introduction:** In India, there are different kinds of tax that are leaved in the form of indirect taxes. GST is the simple tax system. GST is being launched in India to reduce the complication of taxes & improve the system of indirect taxes & as well the problem faced like double taxation. GST would combine several central & state taxes into one tax. The GST has been launched by 1St July 2017. In GST there are three types, i) CGST ii) SGST iii) IGST We need think about GST because pervious taxes had given headache to many businessman. We want to analyze view of Businessman towards GST.


**2. Method & Materials:** As the main aim of our project is "View of Businessmen towards GST''. Considering different age groups, different business tax, so we decide to collect all primary data from the TASGAON which will provide us the necessary and sufficient data within less time and more accuracy.

We made 120 observations to collect the specific information of the given person so that we may access more sample in short span. The related data and information collected is true to

***Corresponding Author:** Miss. Sandhya Ghail, Assistant Professor, Department of Statistics, Balwant College, Vita (M.S.) India 415 311. **E-mail:** sandhyaghail26824@gmail.com*

this project which has been collected within two days of sampling study. The Analysis have been done in Excel

## 3. Statistical Analysis:

### a) Graphical Representation:

To represent our data graphically, we use the multiple bar diagram, pie chart method …..etc. To construct the appropriable divided bar diagram of given data. We use the Microsoft excel.

**1) Bar Chart:** A bar chart or bar graph is a chart that presents grouped data with rectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally .A vertical bar chart is sometimes called a column bar chart.

**2) Pie chart** –A pie chart displays data, information, and statistics in an easy –to-read 'pie –slice' format with varying slice sizes telling you how much of one data element exists. The bigger the slice, the more of that particular data was gathered.

### b) In case of a 2*2 contingency table

The hypothesis, that one factor is independent of the other or not, can be tested by the chi-square test. The hypothesis,

$H_0$: Two attributes are independent of each other.

V/S

$H_1$: Two attributes are dependent of each other.

The hypothesis can be tested by the statistic

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)} = x_1^2$$

In case of 2*2 contingency table, if any one of the theoretical cell frequencies less than 5, then the use of pooling method for chi-square test results in chi-square with zero d.f. which meaningless .This consist in adding 0.5 to cell frequency which is less than 5 and then adjusting for the remaining cell frequencies, so that the row sum, column sum and grand sum remains the same. For 2*2 contingency table, suppose a<5 then we rewrite the table as follows:

| | B1 | B2 | Total |
|---|---|---|---|
| A1 | a+0.5 | b-0.5 | a+b |
| A2 | c-0.5 | d+0.5 | c+d |
| Total | a+c | b+d | N |

According to Yates correction, corrected value of chi-square is given by,

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)} \quad = \chi^2 1(\text{d.f})$$

**Decision Rule:** If cal., $\chi^2$ >Tab. $\chi^2$ with 1 d.f. We reject $H_0$.This means two attributes are not independent.

**c) Test for testing Equality of Two Population Proportion:**

Let P1 and P2 be the proportion of units of possessing the given attribute in two populations. Let a random sample of n1 units be a drawn from a population with population proportion P1. In which the number of units possessing the given attribute is X1 and a random sample of n2 units be a drawn from population with population proportion P2 in which the number of units possessing the given attribute is X2.The corresponding standard normal variate is,

$$Z= \frac{(p1-p2)-(P1-P2)}{\sqrt{\frac{P1Q1}{n1}+\frac{P2Q2}{n2}}} \sim N(0, 1)$$

Here, we want to test the null hypothesis

$H_0$:P1=P2 (i.e. There is no significant difference between population proportion)

V/S

$H_1$:P1≠P2 (i.e. There is significant difference between population proportion)

Under the null hypothesis,

$$Z= \frac{P1-P2}{\sqrt{PQ(\frac{1}{n1}+\frac{1}{n2})}}$$

Let α be the level of significance and $Z_\alpha$ being a critical value.

**Decision Rule:**

1. Suppose $H_1$:P1≠P2 if computed lZl > tab. $Z_{\alpha/2}$ then we reject $H_0$ at α % l.o.s and conclude that there is significant difference between population proportions.
2. Suppose $H_0$: P1 >P2. If computed Z>tab. $Z_\alpha$. We reject $H_0$ at α % l.o.s. O.W. accept $H_0$ at α % l.o.s.
3. Suppose $H_1$: P1 < P2. If computed Z <tab.-$Z_\alpha$.We reject $H_0$ at α % l.o.s. O.W. accept $H_0$ at α % l.o.s.

**Chi-square test for Independence:**

**1. To check dependency between mode of payment and location**

$H_0$: Mode of payment & Location are Independent.

V/S

$H_1$: Mode of payment & Location are Dependent.

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$$

**Data:**

| Area | Online | Offline | Total |
|---|---|---|---|
| Rural | 38 | 7 | 45 |
| Urban | 61 | 14 | 75 |
| Total | 99 | 21 | 120 |

**Calculation:**

$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)} =$ 0.37553311

The table value of $\chi^2$ at 5% level of significance,

Tab $\chi^2 = 3.841$

**Result:** Cal $\chi^2$ < Tab $\chi 2$ , Hence we accept $H_0$ at 5% level of significance.

**2. To check dependency between and location:**

H0: Opinion about tax method is independent on location.

H1: Opinion about tax method is dependent on location.

$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$

Data:

| Area | Old method | GST | Total |
|---|---|---|---|
| Rural | 21 | 24 | 45 |
| Urban | 40 | 35 | 75 |
| Total | 61 | 59 | 120 |

Calculation: $\chi^2 = \frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$

=0.9961

The table value of $\chi^2$ at 5% level of significance,

Tab $\chi^2 = 3.841$

**Result:** Cal $\chi^2$ < Tab $\chi^2$ ,Hence we accept the H0 at 5% l. o. s.

**Proportion test (Z- test):**

$H_0$: There is no significance difference between proportion of online paying method of GST in rural & urban area. V/S

$H_1$: There is significance difference between proportion of online paying method of GST in rural & urban area.

**Calculation:**

$P = \frac{p1n1+p2n2}{n1+n2} = 0.8249$

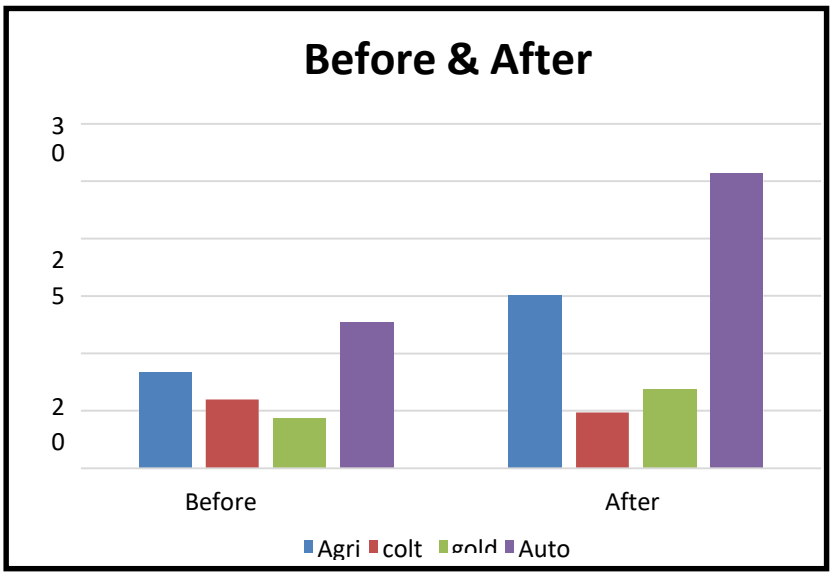
$$Z = \frac{p1-p2}{\sqrt{PQ}\left(\frac{1}{n1}+\frac{1}{n2}\right)} = 0.002383$$

The table value of Z at 5% level of significance,

Tab Z = 1.96

**Result:** Z (cal.) < Z (tab.). , Hence we accept the $H_0$ at 5% l. o. s.

**Graph-1: Difference between Previous tax & GST Tax.**




**Graph -2: Types of GST**

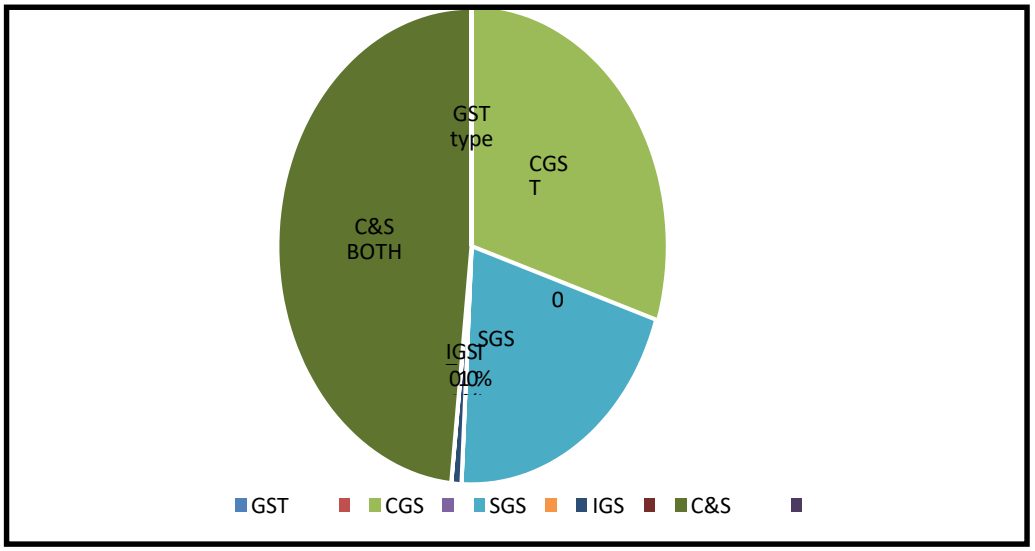
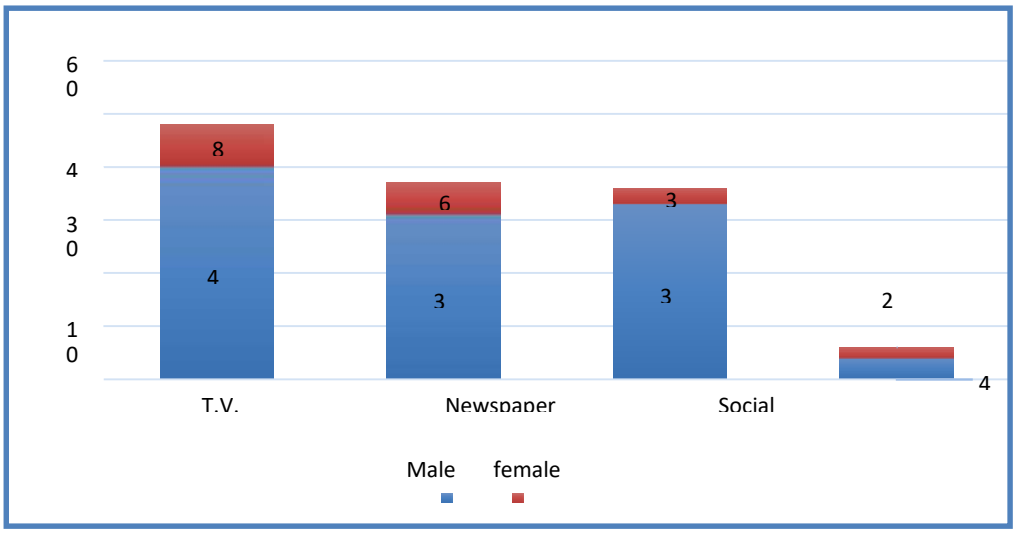


**Table-3: Gender wise receiving Information about GST fromvarious sources**




## 4. Result and Discussion:

1) There is no significance difference between proportions of agree for old &GST method.
2) There is no significance difference between proportion of online paying method of GST in rural & urban area.

3)) we observe that GST taxes Are maximum than Previous tax.
4) Male & female got the moreInformation about the GST through the TV.

## 5. Conclusion:

1. The type of area does not affected on paying method of GST.
2. After GST, taxes in Agri, Cloth, Gold, Automobile is maximum than previous taxes.
3. The central and state GST type is mostly used in business.
4. Removal of previous taxes for uniform GST across the country with make India a common market.

## References:

**1.** R. Lavanya Kumari (2017), "Impact of Goods and Service Tax (GST) on Indian MSMEs", International Journal of Research in Economics and Social Sciences (IJRESS), Vol. 7 (7). July 2017 pp.334-348
**2.** Ankita Verma and Others (2018), "Impact of GST on the regulation of small business", IOSR Journal of Business and Management, Vol. 20(7) July 2018 pp. 81-83
**3.** Gopal K.Kanji. 100 Statistical Tests,
**4.** Kapoor & Gupta. Fundamental of Mathematical Statistics,

Contents available at www.balwantcollege.edu.in
**Emerging Trends in Applied Mathematics and Statistics**
***(Balwant College, Vita (M.S.) India 415 311)***

# Comparative study of IQ Score among College Students

Dhanshree Mali[1*], Pritee Kumbhar[2], Shrutika Kirdat[3], Aditya Pawar[4]
*[1]PG Student, Department of Statistics, Balwant College, Vita*
*[2]PG Student, Department of Statistics, SGM, Karad*
*[3]PG Student, Department of Statistics, YCIS, Satara*
*[4]PG Student, Department of Statistics, MIT, Pune*

**Abstract:** The present study explores the general intelligence score of college students. It also compares the IQ of students in terms of the college type (govt./ private), area of student living (urban/rural), the stream in which students study (arts/ commerce/ Science) and the gender (male/ female). Data was collected from a representative sample of 509 students from different colleges of Vita. Test of Intelligence Quotients (IQ) for College Students was employed to conduct the study. Result showed that, there is significant difference in IQ Score between the Arts and Science stream, Commerce & Science stream However there is no significance differences between the Arts and Commerce stream.

**Keywords:** independent t-test, ANOVA, IQ score, Students, Stream

**1. Introduction:** The intelligence is the ability to learn about, learn from, understand, and interact with one's environment. IQ is defined intelligence as a sort of mental energy, in the form of mental or cognitive abilities, available with an individual which enable him to handle his environment in terms of adaptation to face novel situations as effectively as possible.

Intelligence is normally distributed in nature. It is a product of both heredity and environment. The factors like gender, stream of students, area of student living, type of college and class of family, etc. are may influence the degree of intelligence.

Dr. Sandip Kumar Mandal (2020) carried out a study on rural and urban children in respect of intelligence. Lehmann, I. J. (2014) studied the differences in intelligence of students belong to rural-urban area. The goal of present study is to compare the students IQ score among arts, commerce and science stream.

**Objectives:**

1. To compare the IQ score of male and female students studying in colleges.
2. To compare IQ score of college students studying in govt. and private colleges.
3. To compare the IQ score of college students living among urban and rural areas.
4. To compare IQ score of college students studying in Arts, Commerce and Science streams.

***Corresponding Author:** Miss. Dhanshree Mali, PG Student, Department of Statistics, Balwant College, Vita (M.S.) India 415 311. **E-mail:** dhanshreemali946@gmail.com*

**2. Materials and Method:** In this study, we have started survey in March 2022 and it took around one month for collection the data. According to my aim for the survey, we have randomly selected students of different college of Vita through Google form. The Statistical techniques, namely, independent two sample t-test and One Way ANOVA was used. The data analysis has been done in SPSS software.

**3. Statistical analysis:** In this study score of IQ level of college students is dependent variable. We have given thirty questions to college students to calculate IQ level. Each Question of about IQ is coded Low (1), Average (2), Superior (3), Genius (4). Then summarized score of this thirty questions about IQ to form dependent variable score of IQ level which is coded as: Low (1), Average (2), Superior (3), Genius (4). Here range for Low is 0 to 6, for Average 7 to 14, for Superior 15 to 22 and for Genius 23 to 30. Descriptive Statistics Table -3.1 gives the frequency distribution of each categorical variable is with IQ level. For categorical variables gender, out of 509 people 211 Male 298 Female. Out of 211 male respondents, 22 male has Low IQ level, 140 male has Average IQ level, 42 male has Superior IQ level, 7 male has Genius IQ level. Out of 298 Female respondents, 17 Female has Low IQ level, 183 Female has Average IQ level, 93 Female has Superior IQ level, 5 Female has Genius IQ level.

**Table-3.1: Descriptive statistics**

| **Variable** | | IQ Level | | | |
|---|---|---|---|---|---|
| | | Low | Average | Superior | Genius |
| Gender | Male | 22 | 140 | 42 | 7 |
| | Female | 17 | 183 | 93 | 5 |
| Type of College | Government | 38 | 309 | 130 | 12 |
| | Private | 1 | 14 | 5 | 0 |
| Area of Student Living | Rural | 31 | 281 | 118 | 11 |
| | Urban | 8 | 42 | 17 | 1 |
| Stream of Student | Arts | 8 | 48 | 14 | 1 |
| | Commerce | 16 | 151 | 53 | 3 |
| | Science | 15 | 124 | 68 | 8 |

**3.1 Independent two sample t-test:**

The independent two sample *t*-test is used to test whether there is significant average difference between two groups of population.

**Hypothesis for independent t-test:**

*$H_0$: There is no significant difference between IQ Score of Male and Female students.*

*$H_1$: There is a significant difference between IQ Score of Male and Female students.*

The independent two sample t-test (Table-3.1.1) displays the number of students in category of dependent variable, mean IQ score, t-test statistics and p-value.

**Table-3.1.1: Independent two sample t-test**

| Independent variables | Categories | N | Mean | t-test statistic | P-value |
|---|---|---|---|---|---|
| Gender | Male | 211 | 11.63 | -2.406 | 0.017 |
| | Female | 298 | 12.57 | | |
| Area of Student Living | Rural | 441 | 12.28 | 1.294 | 0.196 |
| | Urban | 68 | 11.54 | | |
| Type Of College | Government | 489 | 12.18 | 0.030 | 0.976 |
| | Private | 20 | 12.15 | | |

For the variable gender, the P-value of independent two sample t-test is 0.017, which is less than 0.05 (level of significance). Hence we reject the null hypothesis at 5% level of significance and conclude that there is a significant difference between IQ score of Male and Female students. The variables which has p-value of t test statistic is greater than 0.05 (level of significance) are not significantly different in IQ score.

### 3.2 One way ANOVA:

One Way Analysis of variance is collection of statistical models used in order to analyze the differences between group means and their associated procedure (such as" variation " among and between groups), developed by R.A. Fisher. ANOVA is useful in comparing testing three or more means (groups of variables) for statistical significance.

**Hypothesis for ANOVA:**

*$H_0$ = Mean IQ Score of three streams of students is same.*
*$H_1$ = At least two streams of students differ significantly in mean IQ.*

The ANOVA Table-3.2.1 gives the sum of squares, degrees of freedom, mean square, F test statistic and p-value (Sig.) to compare the IQ score of three streams.

**Table-3.2.1: ANOVA**

| | Sum of Squares | DF | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 213.512 | 2 | 106.756 | 5.751 | .003 |
| Within Groups | 9393.219 | 506 | 18.564 | | |
| Total | 9606.731 | 508 | | | |

**4. Result and discussion:** The above output of ANOVA Table-3.2.1 shows that, the P-value of F test is 0.003. This p-value is less than 0.05 (level of significance). Hence we reject the null hypothesis at 5% level of significance and conclude that there is significant difference between mean IQ score of three streams of students. But we do not known which of the specific streams of students differs significantly. Hence we have to go for the multiple comparison test which is known as Tukey test shown in Table-4.1.

**Table-4.1: Multiple Comparisons**

| Dependent Variable: IQ Score Tukey HSD | | | | | | |
|---|---|---|---|---|---|---|
| (I) Stream of Student | (J) Stream of Student | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
| | | | | | Lower Bound | Upper Bound |
| Arts | Commerce | -.453 | .587 | .721 | -1.83 | .93 |
| | Science | -1.621* | .590 | .017 | -3.01 | -.23 |
| Commerce | Arts | .453 | .587 | .721 | -.93 | 1.83 |
| | Science | -1.167* | .412 | .013 | -2.14 | -.20 |
| Science | Arts | 1.621* | .590 | .017 | .23 | 3.01 |
| | Commerce | 1.167* | .412 | .013 | .20 | 2.14 |
| *. The mean difference is significant at the 0.05 level. | | | | | | |

We can see that from multiple comparisons table, there is significant difference in IQ Score between the Arts and Science stream (P-value = 0.017), Commerce & Science stream (P-value = 0.013). However there is no significance differences between the Arts and Commerce stream (P-value = 0.721).

**5. Conclusions:** The independent two sample t test suggests that, there is sufficient evidence to conclude that the IQ score between male and female students is different. The result indicate that, there is sufficient evidence to conclude that the average IQ score of science students differ significantly from arts and commerce students.

**References:**

1. Dr. Sandip Kumar Mandal (2020). A Study on Rural And Urban Children In Respect Of Intelligence. *Int. J. Adv. Res. 8(08), 1106-1109.*
2. Cofnas, N. (2019). Research on group differences in intelligence defense of free inquiry. Jn. of philosophical psychology, 33(1):125-147
3. Lehmann, I. J. (2014).Rural-urban differences in intelligence. The journal of educational research, 53(2): 62-68.
4. Emmett, W.G. (2011).The intelligence of urban and rural children. A journal of demography, 7(3): 207-221.

Contents available at www.balwantcollege.edu.in

**Emerging Trends in Applied Mathematics and Statistics**

***(Balwant College, Vita (M.S.) India 415 311)***

# Analysis of Physical activities and Fitness Pattern of GYM: Case Study of Vita Fitness Club

Ashwini Mahadik[1*], Shivani Divate[2]

*[1,2]PG students, Department of Statistics, Balwant College, Vita*

**Abstract:** Fitness is the first and foremost thing that makes both mind and body fit & relax. Everyone needs to maintain health due to which the body gets more flexible and energetic. Countless fitness canters have been set in the market for the betterment of individual's wellbeing. Just like other industries, the competition amongst the fitness and gym canters are also on the rise. Due to the current increasing stress, workload, as well as increasing competition in the young generation health and physical is put as secondary priority among them .the cause of ignoring this things they are facing some health problems .in this research we have studied young men and women who take care of their health and other physical activities.it is also looks at reason why it is important to physically active. Here we studied their opinion about health and physical fitness. Physical activity is one of the most basic human function and needs, which has benefits across the lifespan and has sufficient evidence that the "Vita Fitness Club". The main purpose of this research is to study the level of physical activity, general attitude, perceived impact of physical activity on health. The goal of present research study is to check the gym effect.

**Keyword:** Physical activity, Fitness, Active, Weight, t-test, Descriptive Statistics.

**1. Introduction:** In the present era, many people are suffering from the problem of obesity. Through this research, we tried to find out whether there is any difference in the measurements of body weight, height, wrist, chest, thies, etc. after joining the gym. Many people are joining gyms to get rid of this problem. Through this research, we verified whether there is a difference in your weight after joining the gym. Also see if it improves your health too. To do this research, we took the help of Vita Fitness Club, a gym. For this we collected secondary data of Feb -2022 and Dec -2022.From the total data we got from this information randomly no. chose. In this we selected randomly sampled of 25 girls and 25 boys. Also we got some information from the questionnaire created by Google form.

Descriptive Statistics were used to analyse data where the mean, median, mode and standard deviation was calculated, and a number of correlations were made using the same. Moreover, when a comparison was made between the male and female on their levels of physical activity and there measurement. Also we used here statistical method of paired test to see the result after joining the gym. For this we compare the resulting p value with the level of significance (0.05). The BMI was likewise determined, so as to discover the class of weights that individual's fall into.

***Corresponding Author:** Miss. Ashwini Mahadik, PG Student, Department of Statistics, Balwant College, Vita (M.S.) India 415 311. **E-mail:** ashwinimahadik518@gmail.com*

**Objectives:**

1. To test if there is any difference after joining gym on weight and health.
2. To study the level of physical activity.
3. To study which physical activity useful for men and women.
4. To study the body mass index (BMI) of male and female. Classify it as underweight, overweight, normal and obesity.

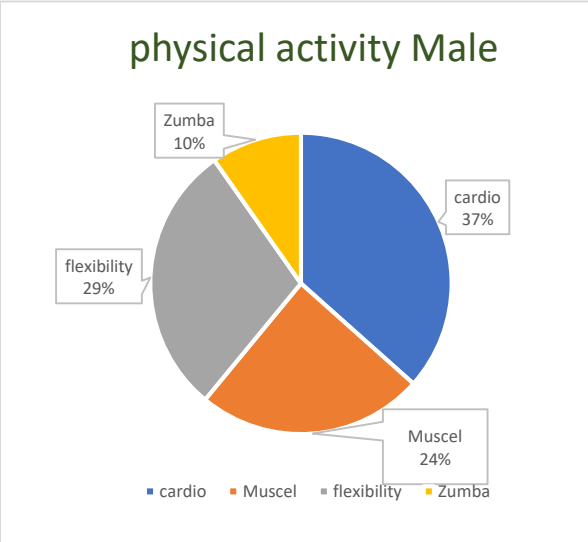
**2. Materials and Method:** The data required for this study was collected between Feb-2022 to Nov-2022. According to aim of this research we randomly selected persons of "Vita Fitness Club". The statistical techniques called as paired t-test" were used to analysis data. The analysis has been done in 'MS-EXCEL'.

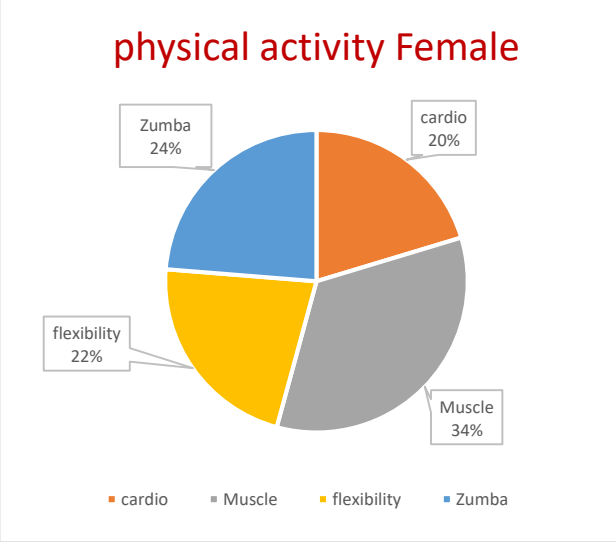
**3. Statistical Analysis:** The frequency table -3.1 shows the number of male and female in each category of physical activity. It indicates that 15 male and 12 female are doing physical activity Cardio.

**Table-3.1: Frequency table**

| Physical activity | Male | Female |
|---|---|---|
| Cardio | 15 | 12 |
| Muscle | 10 | 20 |
| Flexibility | 12 | 13 |
| Zumba | 4 | 14 |

The Pie chart of the above frequency (Table-3.1) shows that, the 37% of male respondent carried out Physical activity called as cardio. However, 10% of male respondents carried out Zumba physical activity. The 34% of female respondent carried out muscle physical activity and 20% of female respondents carried out Cardio physical activity.




**Hypothesis:**

$H_0$: There is no effect of gym on body measurement. Vs

$H_1$: There is effect of gym on body measurements.

The paired t-test (table-3.2 & 3.3) gives the mean of body measurements (such as weight, Neck, Chest etc.) before and after joining the GYM for male and female respectively. It also provides the value of paired t test statistic with respective degrees of freedoms, p-value of test statistic and decision based on p- value & level of significance (α).

**Table-3.2: Paired t –test (for Male)**

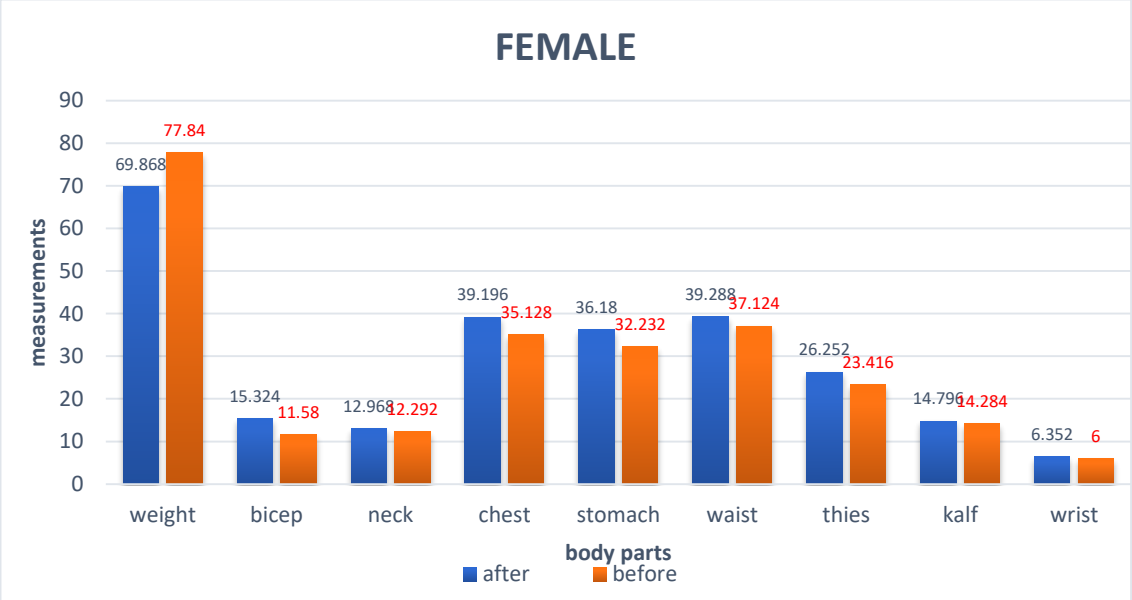
| | | | MEAN | t- Test Statistic | D.F. | P-Value | Decision |
|---|---|---|---|---|---|---|---|
| Male | Weight | After | 72.628 | 2.6256 | 24 | 0.0148 | Reject $H_0$ |
| | | Before | 79.58 | | | | |
| | Neck | After | 13.824 | 2.6174 | 24 | 0.0075 | Rejected $H_0$ |
| | | Before | 14.856 | | | | |
| | Bicep | After | 12.756 | 1.8433 | 24 | 0.0388 | Rejected $H_0$ |
| | | Before | 14.94 | | | | |
| | Chest | After | 37.736 | 1.0265 | 24 | 0.1574 | Fail to rejected $H_0$ |
| | | Before | 54.24 | | | | |
| | Stom | After | 33.868 | 3.0450 | 24 | 0.0028 | Rejected $H_0$ |
| | | Before | 35.336 | | | | |
| | Waist | After | 34.244 | 5.6753 | 24 | 0.0000 | Rejected $H_0$ |
| | | Before | 36 | | | | |
| | Thies | After | 21.272 | 3.867 | 24 | 0.001 | Rejected $H_0$ |
| | | Before | 22.188 | | | | |
| | Kalf | After | 13.76 | 5.5693 | 24 | 0.0000 | Rejected $H_0$ |
| | | Before | 14.296 | | | | |

**Table-3.3: Paired t –test (for Female)**

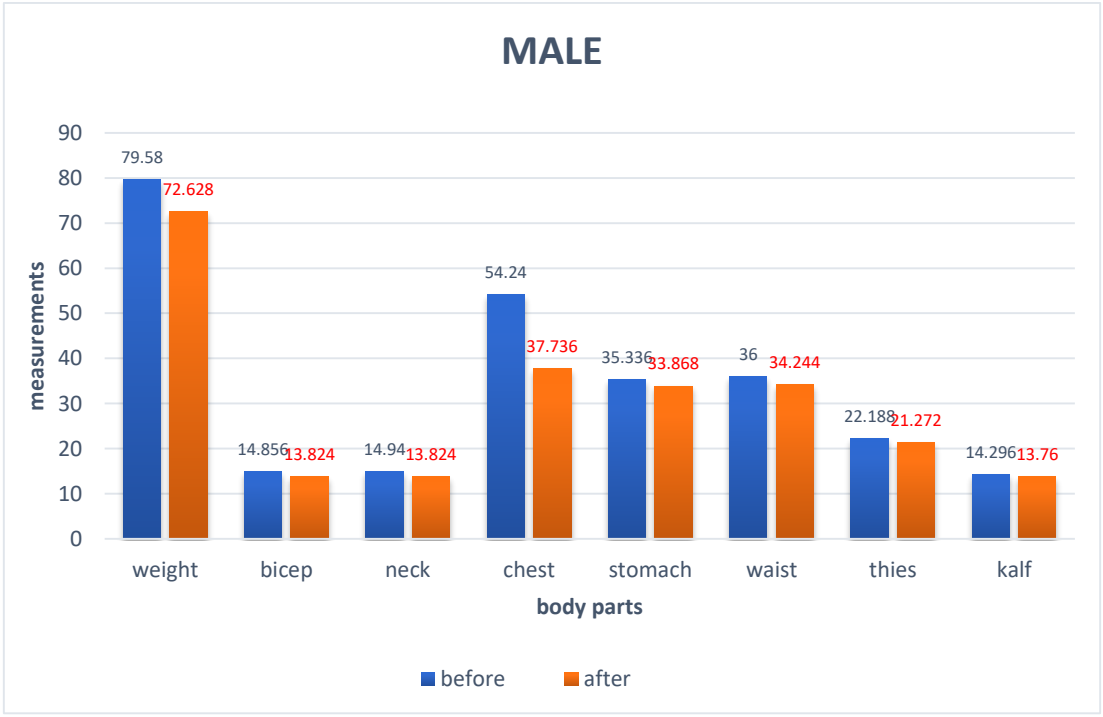
| | | | Mean | t-Test Statistic | D.F. | P-Value | Decision |
|---|---|---|---|---|---|---|---|
| Female | Weight | After | 69.868 | -15.3663 | 24 | 0.0000 | Reject $H_0$ |
| | | Before | 77.84 | | | | |
| | Neck | After | 12.292 | 10.3783 | 24 | 0.0000 | Reject $H_0$ |
| | | Before | 12.968 | | | | |
| | Bicep | After | 11.58 | 22.549 | 24 | 0.0000 | Reject $H_0$ |
| | | Before | 15.324 | | | | |
| | Chest | After | 35.128 | 14.435 | 24 | 0.0000 | Reject $H_0$ |
| | | Before | 39.196 | | | | |
| | Stom | After | 32.232 | 11.9252 | 24 | 0.000 | Reject $H_0$ |
| | | Before | 36.18 | | | | |
| | Waist | After | 37.124 | 9.3678 | 24 | 0.000 | Reject $H_0$ |
| | | Before | 39.288 | | | | |
| | Thies | After | 23.416 | 4.9996 | 24 | 0.0000 | Reject $H_0$ |
| | | Before | 26.252 | | | | |
| | Kalf | After | 14.284 | 7.7094 | 24 | 0.0000 | Reject $H_0$ |
| | | Before | 14.796 | | | | |
| | Wrist | After | 6 | 6.06296 | 24 | 0.0000 | Reject $H_0$ |
| | | Before | 6.352 | | | | |

## 4. Result and discussion:

The p-value of paired t test for each variable for male and female (except Chest of male) is less than 0.05 (Level of Significance). Hence we reject the null hypothesis at 5% level of significance and conclude that there is significance difference between the body measurements before and after joining the GYM for male and female. i.e. The workout in gym are affected on body measurements and thus, it is helpful for loss the weight.




From the above graph, we can see the difference between body measurement before and after joining for women. More importantly, we observe that the gym workout is mostly affected on the weight of female.

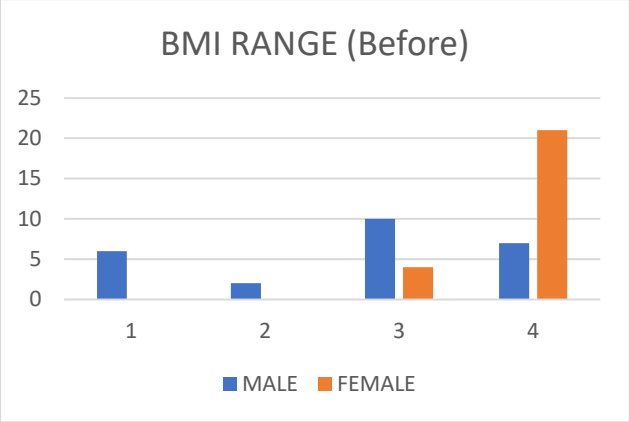


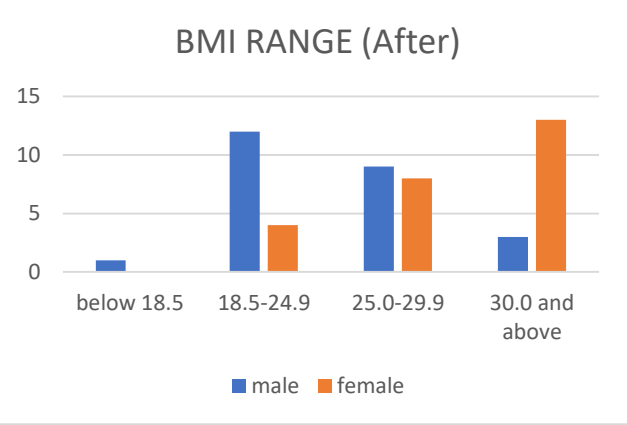
From the above graph, we can see the difference between body measurement before and after joining the GYM for Men. It is observed that the most significant difference weight and chest of mean before and after joining the GYM.

The BMI classification Table-4.1 gives the change in BMI before and after joining the GYM of male and female. It shows that the number of male and female increased after joining the GYM as compared to the before joining the GYM.

**Table-4.1: BMI Classification**

| BMI range | Before | | After | |
|---|---|---|---|---|
| | Male | Female | Male | Female |
| below 18.5 (Under Weight) | 6 | 0 | 1 | 0 |
| 18.5-24.9 (Normal Weight) | 2 | 0 | 12 | 4 |
| 25.0-29.9 (Over Weight) | 10 | 4 | 9 | 8 |
| 30.0 & above (Obesity) | 7 | 21 | 3 | 13 |




**5. Conclusions:** The result of this research study indicate that, GYM has significant effect on body measurements of male and female respondents. This study helps to the male and female who are suffering from underweight or obesity problems. With the help of this, the GYM managers can decide the appropriate physical activity to the male and female regarding their underweight or obesity problems.

**References:**

1. Kljajevic V, Stankovic M, Đorđevic D, Trkulja-Petkovic D, Jovanovic R, Plazibat K, Oršolic M, Curic M, Sporis G. Physical Activity and Physical Fitness among University Students-A Systematic Review. Int J Environ Res Public Health. 2021 Dec 24;19(1):158.
2. Darden S. Weight changes in African American college students: a review of literature. ABNF J. 2014 Winter;25(1):10-2
3. Teixeira PJ, Carraça EV, Markland D, Silva MN, Ryan RM. Exercise, physical activity, and self-determination theory: a systematic review. Int J Behav Nutr Phys Act. 2012 Jun 22;9:78.
4. Ajibade PB. Physical activity patterns by campus housing status among African American female college students. J Black Stud. 2011; 42(4): 548-60.

Contents available at www.balwantcollege.edu.in
**Emerging Trends in Applied Mathematics and Statistics**
***(Balwant College, Vita (M.S.) India 415 311)***

# Face Anti-Spoofing Security System by Using CNN

Rani Mali[1*]
[1]*Assistant Professor, Department of Statistics, Balwant College, Vita*

**Abstract:** To effectively protect the privacy of a person, it is very critical to build a face authentication and anti-spoofing system. This project introduces a novel and appealing face spoof detection technique, which is primarily based on the study of real and spoofed photos. Though several face anti-spoofing or liveness detection methods (which determine at the time of capture whether a face is live or spoof) have been proposed, the issue is still unsolved due to difficulty computationally inexpensive features and methods for spoof attacks. In addition, existing techniques use whole face image or complete video for liveness detection. However, often certain face regions (video frames) are redundant or correspond to the clutter in the image (video), thus leading generally to low performances. Therefore, we propose CNN classifier then gives the output is real or fake face.

**Keyword:** Face Anti-Spoofing Security System, CNN, Real, Spoofing

**1. Introduction:** Nowadays, the security of face recognition systems is challenged. A face image on a mobile can spoof the systems and then the access is granted to attackers. A practical face recognition system demands not only high recognition performance, but also the capability of anti-spoofing to differentiate faces from real persons (genuine face) and those from attackers (fake face). Recently, many papers on face anti-spoofing have been published. We know that machines are gives only numbers so images are converted into numbers. Face anti-spoofing, as a security measure for face recognition system, are drawing increasing attentions in both academic and industrial fields. However, due to the diversity of spoofing types, including print-attacks, replay-attacks, mask attacks, etc., it is still a difficult work to distinguish various fake faces. Some randomly sampled real and fake face images are shown to evaluate the anti-spoofing ability of our eyes. Among all the face images are 3400 and 1700 real and fake face images are 1700. Admittedly, no obvious visual cues are available for us to pick the real face images from the gallery. To train face anti-spoofing models for these subjects, we propose a subject domain adaptation method to transfer the information provided by the subjects which have both genuine samples and fake samples (source subjects) to the subjects having no fake samples (target subjects) to synthesize fake samples Recently, researchers are devoted to come up with more generalized and discriminative features for face anti-spoofing.

In general, these features are all called hand-crafted features because they are designed manually. In this project, however, we exploit deep convolutional neural network (CNN) for face anti-spoofing. To the best of our knowledge, this is the first attempt. Compared with above

**Corresponding Author:* *Miss. Rani Mali, Assistant Professor, Department of Statistics, Balwant College, Vita (M.S.) India 415 311.* ***E-mail:*** *ranimali3597@gmail.com*

hand-crafted features, the features learned from CNN are able to catch more discriminative cues in a data-driven manner. More importantly

**Objectives:** To build a system with high accuracy which takes detection of spoofing i.e Real and fake faces.

## 2. Methods and Materials:

As we make the dataset of real and fake images of faces for M.Sc.-IIyear class, we take 100 images of each student in different angle. Statistics s there are 17 students thus we have total 3400 images of all students.

**Research Methodology:**

I also referred some research paper then some research papers are included different methods such that deep convolutional neural networks and also use this security system. Overall process has been done in three steps namely,

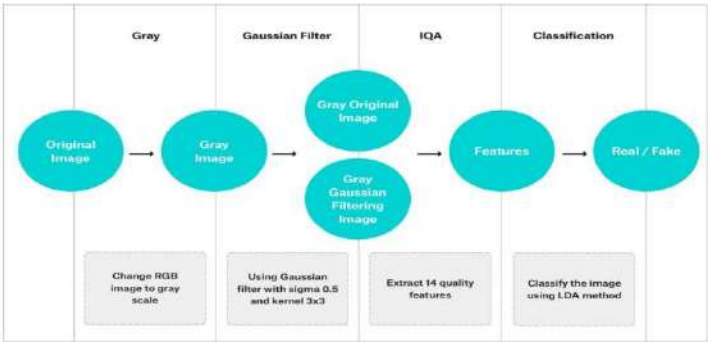
➢ Pre-processing on captured image

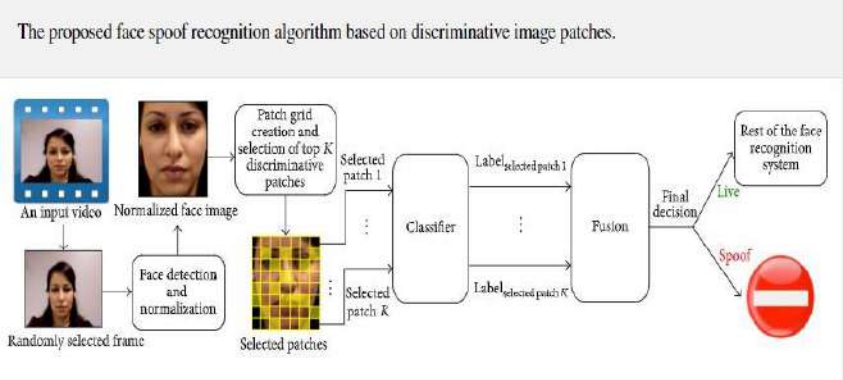
➢ Face Detection

➢ Real and Fake Face Recognition

Face Recognition The system consists of a camera that captures the images of person and then pre-processing is done on that image. After Pre-processing first task is to detect the faces from captured input person face image that was done by open cv with haarcascade features and MTCNN detector. The next step is face recognition which is done Deep deep convolutional neural network (CNN) for face anti-spoofing.

Algorithm:






## 3. Techniques:

### 3.1 Histogram equalization:

To see the students standing on the main door of hotel clearly, we use this technique so that we can be easily recognized them. Histogram equalization is goodtechnique for contrast enhancement in the spatial domain. The method is useful in images with backgrounds and foregrounds that are both bright or both dark. Histogram is a graphical representation of the intensity distribution of an image. Insimple terms, it represents the shows the histogram equalized image of the input image. number of pixels for each intensity value considered.Fig.3.1.1 and 3.1.2

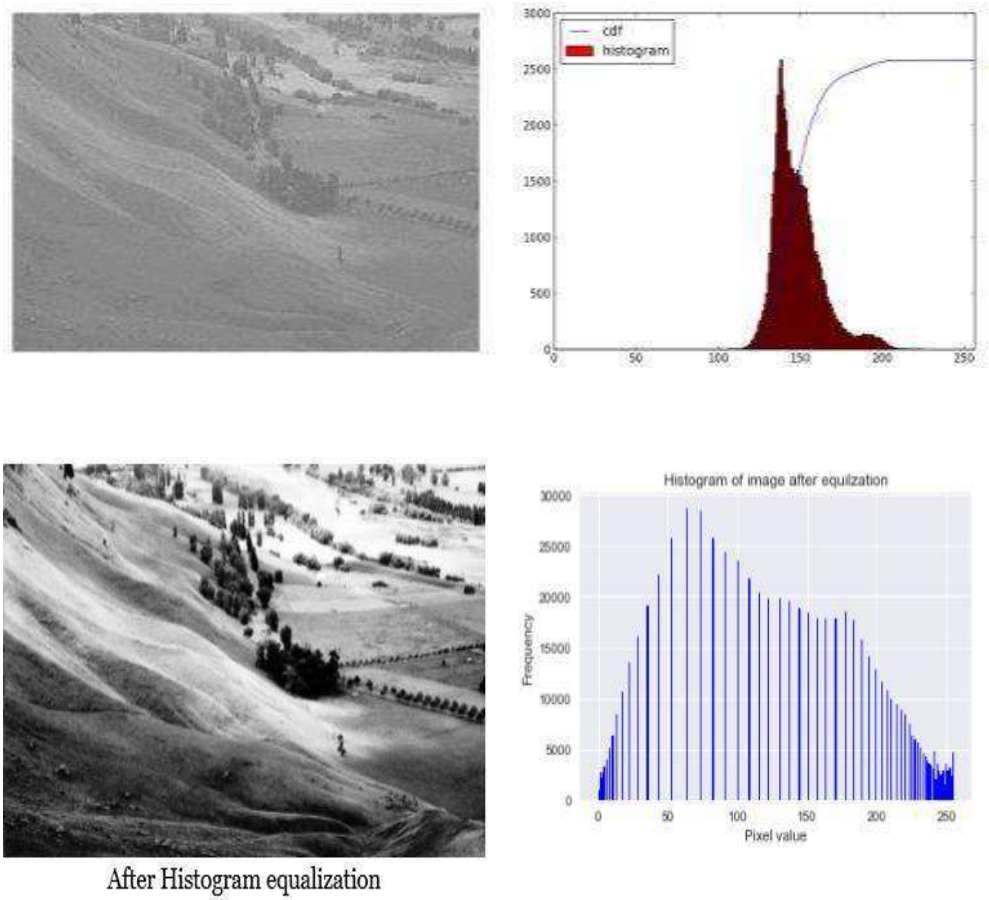



**Fig 3.1.1**

**After Histogram equalization:**

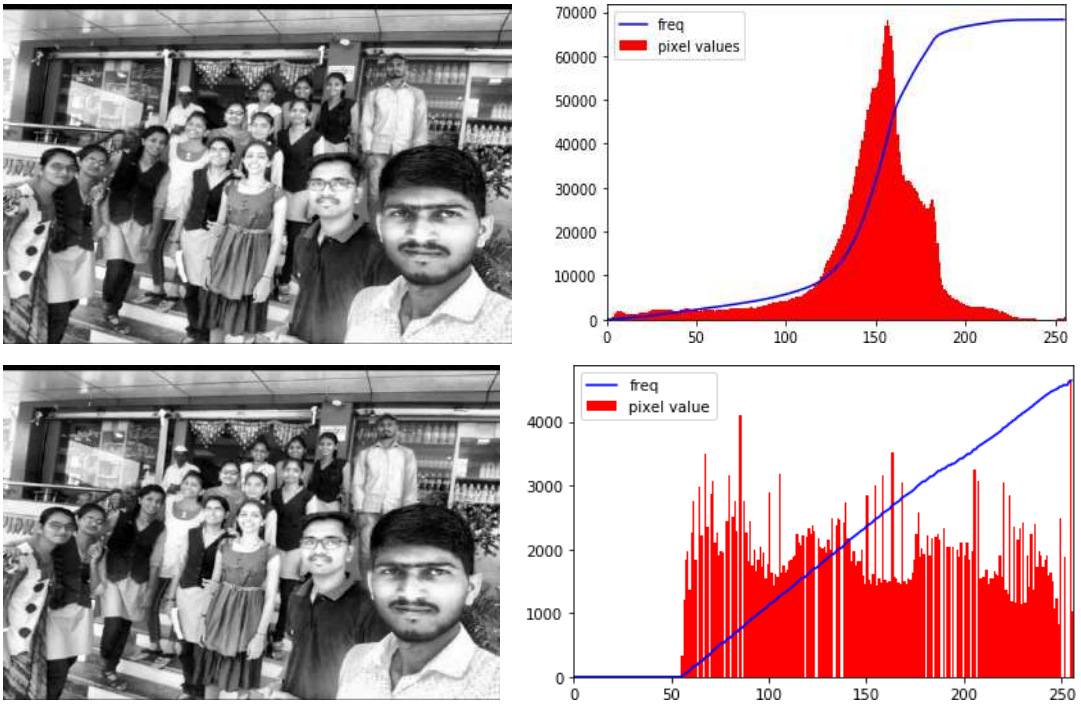



**Fig 3.1.2**

In the above figure, X-axis represents the tonal scale (black at the left andwhite at the right), and Y-axis represents the number of pixels in an image. In the histogram shows the number of pixels for each brightness level (from black towhite), and when there are more pixels, the peak at the certain brightness level is higher.

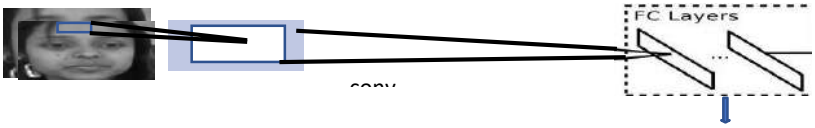
### 3.2 Convolutional Neural Network:

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learningalgorithm similar to Artificial Neural Network (ANN) which can take in an inputimage, assign importance (learnable weights and biases) to various aspects or features in the image and be able to differentiate one image from the other. A convolutional neural network consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of convolutional layers, Relu layer i.e. activation function, pooling layers, fully connected layers and normalization layers. The general architecture of CNN is shown in Fig 3.2.1

Design:

We need three basic components to define a basic convolutional network.

1. The convolutional layer
2. The Pooling layer [optional]
3. The output layer




Input image Real or Fake

Fig 3.2.1

I used deep CNN in this project with 6 convlulation layers then I got the bestaccuracy is 0.9980. It is used to in that security system.

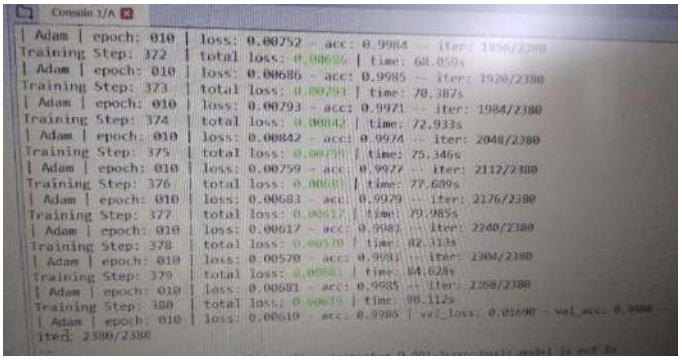



Fig 3.2.2

**4. Model Building:** To build a model for face recognition first we apply open cv by using haarcascade features and mtcnn to detect face in image and crop only the faces as shown in Fig 3.1 The cropped faces are then resized to 200x200 in Gray scale and passed toCNN. As an output we get 200x1 embedding array of each images such that thearray of same person images closes to each other in the feature space.

closes to each other in the feature space.

## 5. Implementation:

5.1 Capturing Image: The camera is placed in such a way that every student is visible in the image captured. We take the image with a high definition camera so that even the back benchers are visible.

5.2 Input Image

5.3 Preprocessing: First the image is converted into grayscale image then Histogram equalizationis used to see the students sitting on the back rows clearly. The Histogram equalized image is shown in below fig.

4. Face detection: In the system presented here, we use CNN classifier. Result seems to be a list that included the coordinates of the bounding box and facial landmarks, as well as thenetwork's confidence in classifying that area as a face. We use the pixel coordinates ofthe bounding box to draw a rectangle around each face as shown in Fig 5.1 and crop each face from the image. The cropped faces are shown in Fig. 5.2

Fig.5.1

Fig .5.2


Fig.5.3

**6. Conclusion:** This project presents a CNN model to address face anti-spoofing in a mutual way. In order to learn more generalizable printed Attack representations for face anti-spoofing. We propose preventing the printed attack representations from overfitting to dataset-specific spoof patterns. The fast CNN achieves not only superior performance for face anti-spoofing on cross-tests, but also high accuracyis 0.9980.

**Limitation:** This project is limited for printed attacks i.e. some once photo only and itmay not respond to video attacks.

**Scope:** It is helpful to prevent such printed attacks at least on a defined scale butdefinitely. And if modified further may be more useful.

**References:**

1. Z. Boulkenafet, J. Komulainen, and A. Hadid. Face antispoofing based on color texture analysis. In Image Processing (ICIP), 2015 IEEE International Conference on, pages 2636–2640. IEEE, 2015.
2. G. Chetty. Biometric liveness checking using multimodal fuzzy fusion. In Fuzzy Systems (FUZZ), 2010 IEEE International Conference on, pages 1–8. IEEE, 2010.
3. L. Feng, L.-M. Po, Y. Li, X. Xu, F. Yuan, T. C.-H. Cheung, and K.-W. Cheung. Integration of image quality and motion cues for face anti-spoofing: A neural network approach. Journal of Visual Communication and Image Representation, 38:451–460, 2016.
4. Medium Blog.

Contents available at www.balwantcollege.edu.in
**Emerging Trends in Applied Mathematics and Statistics**
*(Balwant College, Vita (M.S.) India 415 311)*

# Comparison of Survival Distributions of Lumpy Skin Disease Infected Cattles Using Kaplan-Meier Method

Komal Ghodake[1*], Pranali Kashid[2], Annasaheb Suryawanshi[3]
*[1,2]PG Student, Department of Statistics, Balwant College, Vita*
*[3]Assistant Professor, Department of Statistics, Balwant College, Vita*

**Abstract:** Lumpy skin disease is a viral disease that affects cattle. Most of the cattle suffer from severe emaciation and it may loss of milk production for several months. The severity of the lumpy disease varies between breeds and strains of cattle. Lumpy disease may leads to death of cattle. The main purpose of present study is to compare the survival distribution between species of cattle's who are suffering from the Lumpy Skin Disease. Result showed that, the survival distributions are not same in different levels of species of cattle.

**Key Words**: Lumpy, Cattle, Survival, Kaplan-Meir, Cox regression

**1. Introduction:** LSD (Lumpy skin disease) is an Infectious disease in cattle caused by virus of the family poxviridae. Which is the current issue across all over India in states are Uttarakhand , Himachal Pradesh, Gujrat ,Rajasthan, Jammu& Kashmir ,Madhya Pradesh ,Andaman Nicobar, Panjab. Symptoms observed in cattle are high fevar, reduced milk production ,skin nodules, loss of appetite, increased nasal discharge & water eyes salivation , discharge of water from eyes & nose .This severly affects milk production .When the animal is dying its entire body system is affected. Davies FG (1991) studied about Lumpy skin disease, an African capripox virus disease of cattle.

India is the world's largest milk producer had a cattle population of 192.5 million in 2019 but due to Lumpy and their was 97,000 cattle were resulted in death in three months (July and 23 September). It is major problem facing India. The disease has shown its ability to establish and spread in a wide range of environmental and production systems around the world.

In affected villages, cattle herds should be kept separate from other herd by avoiding communal grazing. Cattle should be treated regularly with insect repellents to minimize the risk of vector transmission of the disease. This measure cannot fully prevent transmission but may reduce the risk. The Indigenous vaccine Lumpy-ProVac$^{IND}$ has been developed to protect livestock from Lumpy skin disease. Vaccine has been developed by national Equine Research center, Hisar (Harayana). LSD created havoc in India. Virus mostly affects cows, buffaloes, and deer. This disease directly effect on Indian Economy, due to reduce in milk production.

The Kaplan-Meier estimator proposed by Kaplan, E. L. and Meier, P. (1958). It is a non-parametric method which is very popular to analyzing time –to-event data. It is also called as

***Corresponding Author:** Miss. Komal Ghodake, PG Student, Department of Statistics, Balwant College, Vita (M.S.) India 415 311. **E-mail:** komalghodake0@gmail.com*

Product limit estimator. It is used to estimate the survival functions of time –to-event data.

Alkhamis M.A. et al. (2016) studied Spatial and Temporal Epidemiology of Lumpy Skin Disease in the Middle East. Their study represents the novel approach to modeling the spatial and temporal dynamics of LSDV in Middle East countries. Roxstrom A et al. (2003) carried out survival analysis of longevity in dairy cattle on a lactation basis. In this research study, we want to compare the Survival distributions for the different levels of species of Cattle and different level of age group of Cattle using Kaplan-Meier method.

## 2. Method and Materials:

We collected the data of 74 Lumpy inspected cattle from Bhendawade village of Sangli district through personal interview with farmer and veterinary doctors. The data consists of 40 species of HF cattle, 10 species of OX cattle, 21 species of Khilar cattle and 3 species of HF cattle. The survival analysis technique, named as Kaplan-Meier is used to analyze data through SPSS software.

## 3. Statistical Analysis:

### 3.1. **Comparison of survival distributions for the different levels of species of Cattle**:

The null and alternative hypothesis for Log-Rank test:

$H_0$: Survival distributions for the different levels of Species of Cattle *are same.*

$H_1$: Survival distributions for the different levels of Species of Cattle *are different.*

The 3.1.1 shows the case processing summary of the each category of the Species of Cattle. This shows that HF Species has 40 cattle, OX Species has 10 cattle, Khilar Species has 21 cattle and Calf Species has 3 cattle. It also gives the number of events and censoring cases in each category of Species of Cattle.

**Table-3.1.1: Case Processing Summary**

| Species of Cattle | Total N | N of Events | Censored | |
|---|---|---|---|---|
| | | | N | Percent |
| HF | 40 | 7 | 33 | 82.5% |
| OX | 10 | 6 | 4 | 40.0% |
| Khilar | 21 | 7 | 14 | 66.7% |
| Calf | 3 | 1 | 2 | 66.7% |
| Overall | 74 | 21 | 53 | 71.6% |

Means for Survival Time (Table -3.1.2) gives the estimate of mean survival time with their standard error and 95% confidence interval each category of Species of Cattle.

**Table-3.1.2: Means for Survival Time**

| Species of Cattles | Estimate | Std. Error | 95% Confidence Interval | |
|---|---|---|---|---|
| | | | Lower Bound | Upper Bound |
| HF | 81.023 | 4.058 | 73.070 | 88.975 |
| OX | 54.780 | 7.716 | 39.656 | 69.904 |

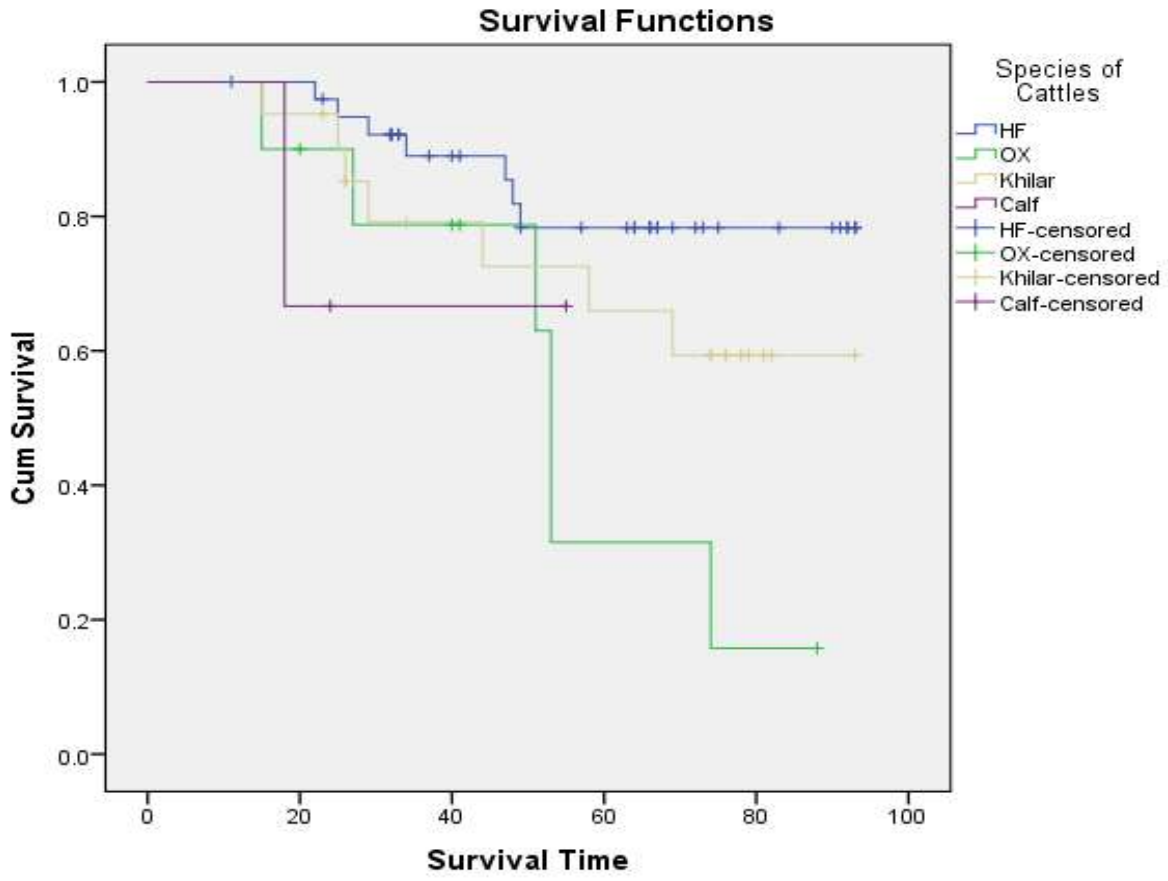
| | | | | |
|---|---|---|---|---|
| Khilar | 71.502 | 6.741 | 58.289 | 84.714 |
| Calf | 42.667 | 10.070 | 22.929 | 62.404 |
| Overall | 73.868 | 3.465 | 67.076 | 80.660 |

The log rank test in overall comparison (Table-3.1.3) is used to test the null hypothesis of no difference in survival between two or more independent groups. The log rank test compares the entire survival distribution between groups and test of whether the survival curves are identical or not.

**Table-3.1.3: Overall Comparisons**

| | Chi-Square | df | Sig. |
|---|---|---|---|
| Log Rank (Mantel-Cox) | 7.894 | 3 | .048 |
| Test of equality of survival distributions for the different levels of Species of Cattle. | | | |

The plot of the cumulative survival proportion against survival time with different levels of Species of Cattle is shown below. The horizontal axis shows the survival time of Lumpy Skin Disease Infected Cattles. In this plot, drops in the survival curve occur whenever the cattle get died due to Lumpy Skin Disease. The vertical axis shows the probability of survival.



The plot above will help to understand how the survival distributions compare the different levels of Species of Cattle .When considering the Species of Cattle, there is difference between survival curves for four levels of Species of Cattle. Thus the survival time is different for four levels of Species of Cattle.

**3.2. Comparison of survival distributions for the different levels of Age group of Cattle:**

The null and alternative hypothesis for Log-Rank test:

$H_0$: Survival distributions for the different levels of Age group of Cattle *are same.*

$H_1$: Survival distributions for the different levels of Age group of Cattle *are different.*

The Table 3.2.1 shows the case processing summary of the each category of four levels of age groups of lumpy infected Cattle. This shows that, 1-5 years age group has 25 cattle, 6-10 years age group has 45 cattle, 11-15 years age group has 03 cattle and 16-20 years age group has 01 cattle. It also gives the number of events and censoring cases in each category of age groups of lumpy infected Cattle.

**Table-3.2.1: Case Processing Summary**

| Age group of Cattle | Total N | N of Events | Censored | |
|---|---|---|---|---|
| | | | N | Percent |
| 1-5 Yrs | 25 | 10 | 15 | 60.0% |
| 6-10 Yrs | 45 | 8 | 37 | 82.2% |
| 11- 15 Yrs | 3 | 2 | 1 | 33.3% |
| 16-20 Yrs | 1 | 1 | 0 | 0.0% |
| Overall | 74 | 21 | 53 | 71.6% |

Means for Survival Time (Table 3.2.2) gives the estimate of mean survival time with their standard error and 95% confidence interval for four levels of age groups of lumpy infected Cattle.

**Table-3.2.2: Means for Survival Time**

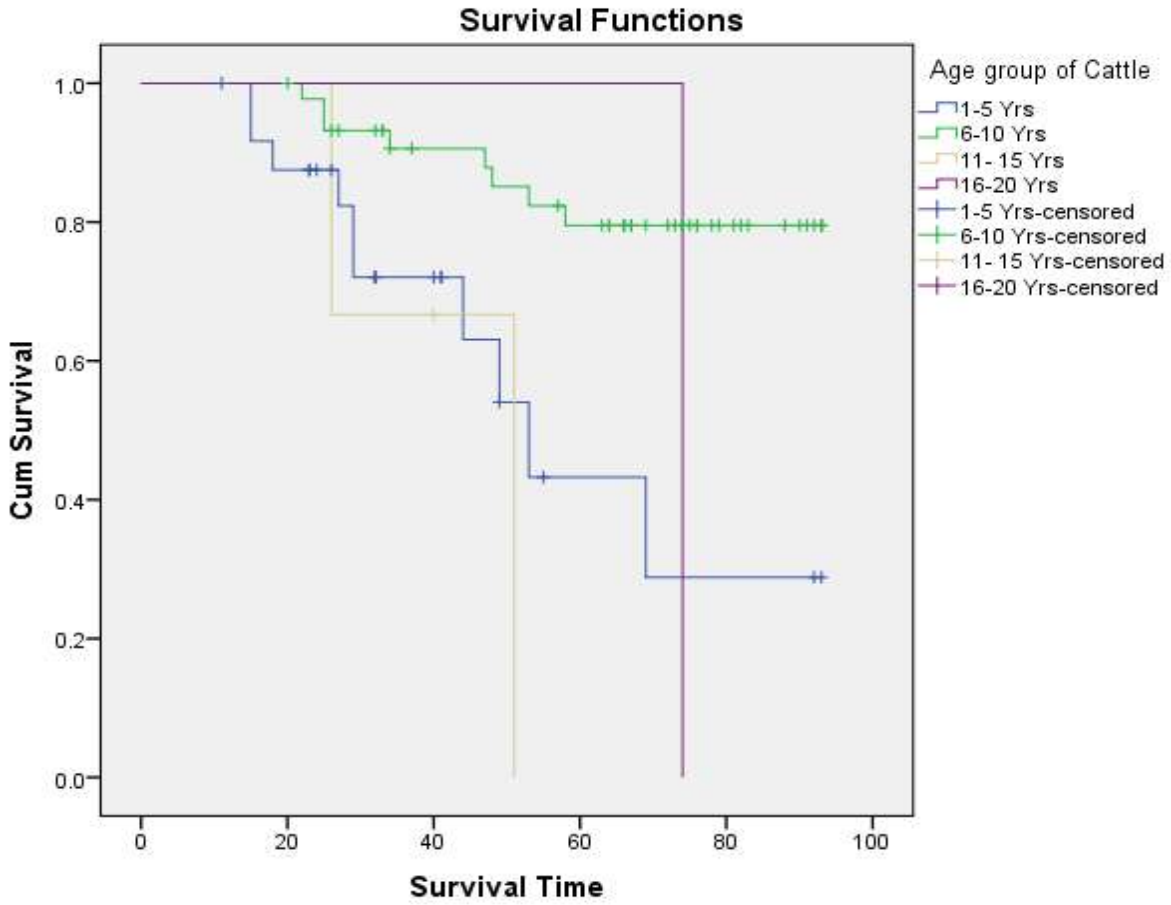
| Age group of Cattle | Estimate | Std. Error | 95% Confidence Interval | |
|---|---|---|---|---|
| | | | Lower Bound | Upper Bound |
| 1-5 Yrs | 57.231 | 7.211 | 43.096 | 71.365 |
| 6-10 Yrs | 82.178 | 3.493 | 75.332 | 89.024 |
| 11- 15 Yrs | 42.667 | 9.623 | 23.807 | 61.527 |
| 16-20 Yrs | 74.000 | .000 | 74.000 | 74.000 |
| Overall | 73.868 | 3.465 | 67.076 | 80.660 |

The log rank test in overall comparison (Table 3.2.3) is used to test the null hypothesis of no difference in survival between two or more independent groups. The log rank test compares the entire survival distribution between groups and test of whether the survival curves are identical or not.

**Table-3.2.3: Overall Comparisons**

| | Chi-Square | df | Sig. |
|---|---|---|---|
| Log Rank (Mantel-Cox) | 14.789 | 3 | .002 |
| Test of equality of survival distributions for the different levels of Age group of Cattle. | | | |

The plot above will help to understand how the survival distributions compare the different levels of age group of Cattle. When considering the age groups of Cattle, there is difference between survival curves for four levels of age groups of lumpy infected Cattle. Thus the survival time is different for four levels of age groups of lumpy infected Cattle.




**4. Result and Discussion:**

i. The log rank test calculates a Chi-statistic in overall Comparisons (Table-3.1.3). The value of Chi-statistic is 7.894 with p-value 0.048. This p-value is less than 0.05 (level of significance). Hence we reject the null hypothesis at 5% level of significance and conclude that Survival distributions for the different levels of Species of Cattle are different.

ii. The log rank test calculates a Chi-statistic in overall Comparisons (Table-3.2.3). The value of Chi-statistic is 14.789 with p-value 0.002. This p-value is less than 0.05 (level of significance). Hence we reject the null hypothesis at 5% level of significance

and conclude that survival distributions for the different levels of Age group of Cattle are different.

**5. Conclusions:** Farmers are afraid of diseases like Lumpy that come suddenly. This research will be useful to reduce the fear of the farmer about the lumpy disease of Cattle. Because in this study, the distribution of survival according to age and species of cattle has been studied which shows that the survival distributions for the different levels of species of Cattle are different. It also indicates that, the survival time is different for each age group of cattle who inspected from Lumpy skin diseases. However the cattle whose ages between six to ten years have more survival chance as compared to other age group of lumpy infected cattle**.** Hence this findings will helps to the veterinary doctors as well as farmer.

**References:**

1. Alkhamis M.A. & VanderWaal K. (2016). “Spatial and Temporal Epidemiology of Lumpy Skin Disease in the Middle East”, 2012–2015. Front. Vet. Sci., 3, 19.
2. Roxstrom A, Ducrocq V, Strandberg E (2003). “Survival analysis of longevity in dairy cattle on a lactation basis”. Genet SelEvol. May-Jun;35(3):305-18.
3. Davies FG (1951). “Lumpy skin disease, an African capripox virus disease of cattle”. Br Vet J.,Nov-Dec;147(6):489-503.
4. Kaplan, E. L.; Meier, P. (1958). "Nonparametric estimation from incomplete observations". J. Amer. Statist. Assoc. 53 (282): 457–481.

Contents available at www.balwantcollege.edu.in
**Emerging Trends in Applied Mathematics and Statistics**
*(Balwant College, Vita (M.S.) India 415 311)*

# Statistical Analysis of Financial Literacy with reference to Devikhindi Village in Khanapur Tehsil

Nikita Dharane[1*], Jagruti Tambave[2], Priti Kashid[3], Prashant Ganjave[4]
*[1,2,3]UG students, Department of Statistics, Balwant College, Vita*
*[4]Assistant Professor, Department of Economics, Balwant college vita*

**Abstract:** Financial literacy is the ability of an individual to take informed decision on basic financial practices. Financial literacy is essential not only for the individual but also for development of markets of a country .this study investigates the relation between financial literacy level and gender of individual. The goal of present study is to determine whether gender affects the level of financial literacy. In order to test this claim, the ransom sample of 117 people living in Devikhindi village was selected. The t -test was used to analysis the data. The study found out which gender has more sufficient financial knowledge and determine the gender of individuals who needed to receive finance education. As the result of analysis, it was concluded that gender has no significant effect on financial literacy and negligible financial literacy.

**Keywords:** Internet Banking, Financial education, Debit card, t-test.

**1. Introduction:** Financial literacy is an essential life skill that impact on individual, family well-being on the wider economy. The rise of this concept is also directly related to concept of behavioural finance. The concept of behavioural finance based on financial freedom and self – decision of individuals, it has become essential for individual to take correct financial decisions.

The best decision for that will be to increase financial literacy level of individual. Financial literacy is very important for all societies and it has macroeconomic effect as well. The goal of the study is to measure effect of gender on financial literacy and also to know how many peoples are aware about financial literacy.

Huston et al. (2010) measured the financial literacy. They studied about the cause and effect relationship between financial education and financial literacy. Adam et al. (2018) studied about the gender disparity in financial literacy still persist after retirement. In this research we want to compare the financial literacy level between male and female.

**2. Methods and Materials:** For this study we select Devikhindi village having population 2012. With the help of questionnaire we collected primary data about financial literacy of person individually. The data collection took place on 11th December 2022.

**Population of the study:** The population of study was all individuals living in Devikhindi Village with potential of having financial literacy.

**Corresponding Author: Miss. Nikita Dharane, UG student, Department of Statistics, Balwant College, Vita (M.S.) India 415 311. **E-mail:** nikitadharane20@gmail.com*

**Sample of the study:** The sample of study was randomly selected no certain region limit was applied when selecting sample.

**Data analysis:** The data where analysed through the SPSS and excel. The Independent two sample t-test where applied to compare the financial literacy level between male and female

## 3. Data Analysis and Interpretation:

**$H_0$:** There is no significant difference between financial literacy level of male and female. Vs

**$H_1$:** There is significant difference between financial literacy level of male and female.

The Group statistics Table-3.1 indicates that the male and female in terms of financial literacy level. This table gives, the number of male & female, mean, standard deviation and standard error of mean.

**Table-3.1: Group Statistics**

| | Gender | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| Difference | Male | 65 | .3538 | .1118 | .0139 |
| | Female | 52 | .3856 | .1262 | .0175 |

The independent sample test (table-3.2) gives the results of Levene's Test for Equality of Variances and t-test for Equality of Means. As shown in the table 3.2, the t-test results are reported twice. In first result the equal variances assumed and in second result equal variances not assumed. The assumptions of equal variances is evaluated using Levene's Test for equality of variances. As a rule of thumb if P-value $\leq 0.05$, we reject null hypothesis of equal variances and use the second result of t-test. Reversely if its P-value > 0.05 then accept null hypothesis of equal variances and use first result of t-test.

**Table-3.2: Independent Samples Test**

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| | | | | | | | | | Lower | Upper |
| Difference | Equal variances assumed | 1.249 | .266 | -1.442 | 115 | .152 | -.0318 | .0220 | -.0754 | .0119 |
| | Equal variances not assumed | | | -1.423 | 102.842 | .158 | -.0318 | .0223 | -.0760 | .0125 |

## 4. Result and Discussions:

The p-value of Levene's test for equality of variances is 0.266 which is greater than 0.05, hence we fail to reject the null hypothesis of equal variances and use the first result of t-test. In the first result of t-test, the value of t test statistic is -1.442 with corresponding p-value is 0.152.

The p-value of independent two sample t-test is greater than 0.05 (level of significance). Hence we fail to reject the null hypothesis at 5% level of significance. Thus, there is no sufficient evidence to conclude that financial literacy level of male and female are different.

**5. Conclusions:** The result of this study indicate that the financial literacy level is nearly equal in males and females. Obtaining financial literacy is most important task for human beings. If people are financially literate, then their ability to make financial decisions will be high and therefore they will be able to manage money effectively. Hence they reaches to their financial goal.

**References:**

1. Adam, A.M., Boadu, M.O. and Frimpong, S. (2018), "Does gender disparity in financial literacy still persist after retirement? Evidence from Ghana", *International Journal of Social Economics*, Vol. 45 No. 1, pp. 18-28.
2. Hathaway, Ian, and Sameer Khatiwada. 2008. "Do Financial Education Programs Work?" Federal Reserve Bank of Cleveland, Working Paper No. 08-03
3. Huston, Sandra J. (2010), "Measuring financial literacy". The journal of consumer affairs, 44(2), 296-316.
4. Bodie Zvi, (2006), "A note on Economic principles and financial literacy", NFI Policy Briefs 2006-PB-07, Indiana State University policy brief.

Contents available at www.balwantcollege.edu.in
**Emerging Trends in Applied Mathematics and Statistics**
***(Balwant College, Vita (M.S.) India 415 311)***

# Ordinary Differential Equations (ODE) using SCILAB Programming and Euler's Technique

Amruta Hasabe[1*]
[1]*Assistant Professor, Department of Mathematics, Balwant College, Vita*

***Abstract*:** For engineers and scientists, mathematics is crucial, but understanding it can be challenging without the right equipment and measurements. A mathematical approach is one of the mathematical formulas used in computer programming. In this study, Scilab is utilized to handle issues with mathematical models like matrices, ODE operations, and integration problems. It is nevertheless true that analytical methods cannot be used to solve the vast majority of first order initial value problems. It is crucial to have alternative strategies for solving the issue. There are many techniques used nowadays to approximate the solutions of differential equations numerically. Here, we provide the earliest and most straightforward of these approaches, developed by Euler in 1768. It is also known as the Euler method or the tangent line approach. The approximate solution is generated using a set step size of h. The goal of this study is to demonstrate the implementation of Euler's technique in detail, compare the modified Euler's solution to the actual value obtained from integration, and solve ODEs using Scilab's built-in functions.

***Key-Words*:** - Euler methods, Euler Modified methods Numerical Methods, Scilab programming, Ordinary Differential Equation.

## 1 Introduction

Numerical methods are methods for formulating mathematical issues so that arithmetic operations can be used to solve them. Although there are many different kinds of numerical approaches, they all share the same flaw: they all require a lot of time-consuming and tedious mathematical calculations. It should come as no surprise that, in recent years, the use of numerical approaches in engineering problem solving has grown significantly, thanks to the advent of quick, effective digital computers. These complex computations can now be done using computers and numerical techniques. It can approach these computations without resorting to simplifying assumptions or time-consuming procedures by using computing power to achieve solutions directly.Numerical approaches are alternatives that considerably expand your ability to confront and solve difficulties, even though analytical solutions are still very helpful for both problem solving and offering insight. As a result, you have more time to exercise your creative abilities. Thus, the phrasing of the problem, the interpretation of the solution, and the inclusion of entire system, or "holistically" awareness can all be given more attention. Numerical technique use and development have virtually skyrocketed since the late 1940s as a result of the widespread use of digital computers [1]. Initially, the cost of

***Corresponding Author:** *Mrs. Amruta Hasabe, Assistant Professor, Department of Mathematics, Balwant College, Vita (M.S.) India 415 311.* ***E-mail:*** *amruta.hasabe15@gmail.com*

access to huge mainframe computers served as a barrier to this growth, and as a result, many engineers continued to use straightforward analytical methods for a sizable percentage of their work.

## 2 Backgrounds

### 2.1 Benefit of Numerical Methods

Numerical techniques are incredibly effective instruments for solving issues. Large systems of equations, nonlinearities, and complex geometries that are frequently impossible to solve analytically and are typical in engineering practice can all be handled by them. As a result, they significantly improve your ability to solve problems. When working as an engineer, you may frequently need to employ pre-made, commercially available computer programs that apply numerical methods. Understanding the fundamental theory that underpins the methodologies is frequently required for the wise usage of these programs. Canned software cannot be used to solve many issues. It is commonly recognized that writing computer programs is an efficient approach to learn programming Mathematical Models and Methods in Modern Science. Numerical approaches are perfect for this because they are often created for computer implementation. Additionally, they are particularly suitable for demonstrating the potential and constraints of computers. You can increase your comprehension of mathematics by using numerical approaches. The "nuts and bolts" of some otherwise esoteric issues can be reached using numerical methods because one of its functions is to reduce higher mathematics to elementary arithmetic operations. This new viewpoint can result in improved comprehension and insight.

### 2.2 Introduction to Scilab Programming

Scilab is a free and open source software package for high-performance numerical computations and visualization for engineering and scientific applications. It provides an interactive computation and visualization environment along with a large number of easy to use built-in functions for technical and scientific computations, powerful 2-D and 3-D graphics and animations. Further, it is also a high level programming language with advanced data structures to implement complex computations, to develop new features and to modify/extend the existing features Scilab stands for Science laboratory. It is available as an open source software package, freely downloadable, under the CeCILL license (GPL compatible) for GNU/Linux, Mac OS and Windows for both 32 bits and 64 bits platforms Scilab is a very powerful programming language with a very rich library of scientific and engineering functions. From the software point of view, it is an interpreted programming language. It provides an Integrated Development Environment (IDE) for programming and supports features and programming constructs that are commonly available in general purpose programming languages like java, C and C++. It also has easy to use built-in tools for creating Graphical User Interfaces (GUI) for Scilab applications [2] From the software point of view Scilab is an interpreted language. Scilab language allows dynamically compiling and linking other languages such as FOTRAN, MATLAB, C++ if they were a part of built-in features [3].Here is a brief summary of Scilab's capabilities: matrices, polynomials, linear equations, signal processing, ordinary differential equations, statistics, and linear equations. Fig. 1 displays the Scilab console for the Windows mode. The Scilab also

provided an editor to edit script easily. The editor can be accessed from the menu of the console, under the Application> Editor menu or from the console [4]. The Fig. 2 shows the Scilab Text Editor. This editor allows managing several files at the same time. The following example, the function display is used in interactive mode to print the string 'Hello World!'.

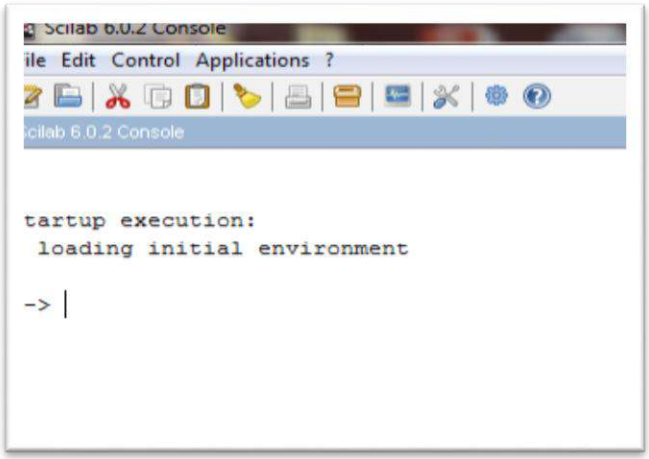



Figure 1: Overview of Scilab console

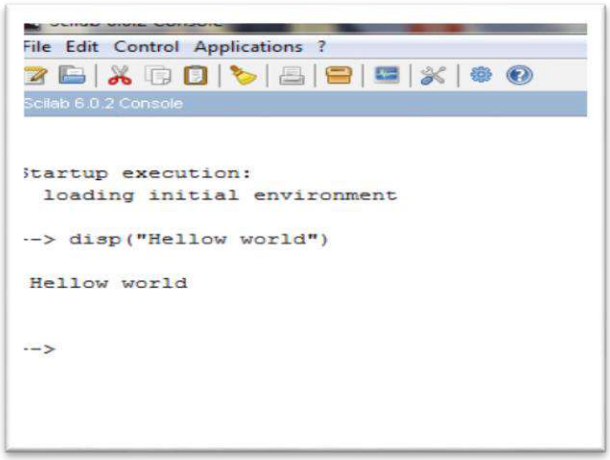



Figure 2: Overview Scilab Text Editor (SciNote)

### 2.3 Introduction to iterative ODE's

Solving linear equations, ordinary differential equations, and numerical integration are all topics that are connected to producing matrices in ODE's study of numerical methods.Numerous postsecondary mathematics and engineering courses place a significant emphasis on numerical methods that involve mathematical modeling [5].When it is difficult or even impossible to find precise answers to mathematical issues posed in science and engineering, numerical approaches are frequently applied. Differential equations can only be solved analytically for a select few.an ordinary differential equation (ODE).It has recently gotten increased attention and recognition in other disciplines like bioengineering, economics, biology, epidemiology, and the medical sciences [6].In the past, deterministic models based on ideas from algebra, vector calculus, regression, differential equations, etc. have been the only approaches and skills that have been applied. The numerical application in Scilab is actually more than capable of solving ODEs [7].

In this paper, we go through several typical engineering problem-solving issues as well as some fundamental Scilab programming concepts. Mathematical Models and Methods in Modern Science an equation involving one or more derivatives of the unknown functioncalled an ODE, or ordinary differential equation. The largest derivative's order establishes the equation's order. For instance, the equation is referred to be a first-order ODE if the first derivative is the only derivative. The equation is referred to as a second-order ODE if the largest derivative is second order. Depending on how the conditions at the domain's endpoints are described, the issues associated with solving an ODE can be divided into boundary value problems (BVP) and initial-value problems (IVP) [8].In particular for Taylor's series, the Gauss-Legendre-Quadrate formula can be used for ODE integration.

### 2.4 Euler's Methods

The Euler methods are straightforward techniques for resolving first-order Ordinary Differential Equations yet, despite their simplicity and low accuracy, they are particularly

well suited for fast programming. Euler's can be derived from the expansion of the Taylor's series.

$f(x_{i+1}) = f(x_i) + f'(x_i)h + \frac{f''(x_i)}{2}h^2 + \cdots$ Then rewrite only the first order forward difference

will give them $f'(x_i) \approx \frac{f(x_{i+1})-f(x_i)}{h}$…….(1)

and use h as a step size $f(x_{i+1}) = f(x_i) + f'(x_i)h$

Let us consider $f'(x_i)= f(x_i) + hf(x_i, y_i)$ therefore equation 1 can be written as;

$f(x_{i+1}) = f(x_i) + hf(x_i, y_i)$ ……..(2)

Equation (2) is Called Euler's method. Then consider i as a integer numbers where i=0,1,2,3…….n now equation (2) can be expanded into;

$f(x_1) = f(x_0) + hf'(x_0)$

$f(x_{i+1}) = f(x_i) + hf(x_i, y_i)$

$f(x_2) = f(x_1) + hf(x_1, y_1)$

$f(x_3) = f(x_2) + hf(x_2, y_2)$

.

.

.

$f(x_n) = f(x_{n-1}) + hf(x_n, y_n)$

The final answer of Euler's can be more accurately and more stable if modify it by using the slope of the function at the estimated midpoints of $(x_i, y_i)$ and $(x_{i+1}, y_{i+1})$[8] to approximate $y_{i+1}$ .Again we use the Euler's Equation

$f(x_{i+1}) = f(x_i) + hf(x_i, y_i)$

$$y_{i+1} = y_i + f\left(\frac{x_i + x_{i+1}}{2}, \frac{y_i + y_{i+1}}{2}\right)h$$

$$=y_i + f(x_i + \tfrac{h}{2}, y_i + \tfrac{\Delta y}{2})\text{……..(3)}$$

Where $\Delta y$ is the estimated incremental value of y from $y_i$ and can be obtained using Euler's formula as ;$\Delta y = hf(x_i, y_i)$ then equation (3) can be written as

$$y_{i+1} = y_i + h\, f\left(x_i + \tfrac{h}{2}, y_i + \tfrac{h}{2}\, f(x_i, y_i)\right)\text{……(4)}$$

Now Equation (4) is called as the Modified Euler's Method.

## 3 Comparison Results Euler's and Scilab

### 3.1 Iterative Euler's methods

An iterative numeric answer for Euler's was compared the answer by using the below an example. The answer should be comparatively. Solve the ODE's over the interval from x =1 to 2 using step size of 0.25 where y(1) = 2. Consider theinitial equation$\frac{dy}{dx} = \frac{2y}{x}$Here we wish to approximate the value y(2) using the Euler's methods with step sizes h = 0.25 and modify Euler's. when $x_0 = 1, y_0 = 2$.

#### 3.1.1 Euler's Method

$f(x_{i+1}) = f(x_i) + hf(x_i, y_i)$

$f(1.25) = 2 + (0.25)(4) = 3.0000$

$f(1.5) = 3.0 + (0.25)\left[\frac{2(3.0000)}{1.25}\right] = 4.2000$

$f(1.75) = 4.2 + (0.25)\left[\frac{2(4.2000)}{1.5}\right] = 5.6000$

$f(2.0) = 5.6000 + (0.25)\left[\frac{2(5.6000)}{1.75}\right]$ =7.2000

#### 3.1.2 Modify Euler's Methods

$y_{i+1} = y_i + h\, f\left(x_i + \frac{h}{2}, y_i + \frac{h}{2}\, f(x_i, y_i)\right)$

$f(1.25) = 2.0 + (0.25)f(1 + 0.125,2 + 0.125f(1,2))$

= 2.0+0.25 f(1.125,2.5)

= 3.1100

$f(1.5) = 3.11 + (0.25)f(1.25 + 0.125,3.11 + 0.125f(1.25,3.11))$

=3.11+.25f (1.375, 3.732)

=4.4700

$f(1.75) = 4.4700 + (0.25)f(1.5 + 0.125,4.4700 + 0.125f(1.5,4.4700))$

$=4.4700 + (0.25)f(1.625,5.125)$

= 6.0746

$f(1.75) = 6.0746 + (0.25)f(1.75 + 0.125,6.0746 + 0.125f(1.75,6.0746))$

$=5.6289 + (0.25)f(1.875,6.9424)$

= 7.4802

The exact value of y(x) and comparison between Euler's and modified are tabulated at Table 1 below:

Table 1: Comparison value of Euler's, Modify Euler's and Exact value.

| x | y(x) | | | | |
|---|---|---|---|---|---|
| | Euler's methods | Modify Euler's | Exact value | Error Euler's | Error Modify Euler's |
| 1.00 | 2.0000 | 2.0000 | 2.0000 | - | - |
| 1.25 | 3.0000 | 3.1100 | 3.1250 | 4.00% | 0.4480% |
| 1.50 | 4.2000 | 4.4700 | 4.5000 | 6.67% | 0.6667% |
| 1.75 | 5.6000 | 6.0746 | 6.1250 | 8.57% | 0.8229% |
| 2.00 | 7.2000 | 7.4802 | 8.0000 | 10.00% | 6.4975% |

### 3.2 ODE's using Scilab programming

The exact value was compared with Euler's and modified Euler's in graphical presentation by using Scilab show at Fig.3. Form the graph showed thatthe exact value is nearest to modified Euler's value. Errorinvolve for this case can be calculated as below;

$Error=\frac{|Ex-Ev|}{Ex} \times 100\%$, Ex=Exact_ value, Ev=Euler's _Value

Error calculation from the Table 1, showed that modified Euler's give the smaller error compared to the Euler's value. While x's value is increased with the no. of error also increased and this was shown for x = 2.00, both error for Euler value andmodify Euler's give tremendous error more than 5%. Then for x's value from 1.0 to 1.75 with the error is (< 1%) actually give the good result and it was agreed with David Houcque [8]. The details scilab note and code for this ODE's also presented and show at Fig. 2.

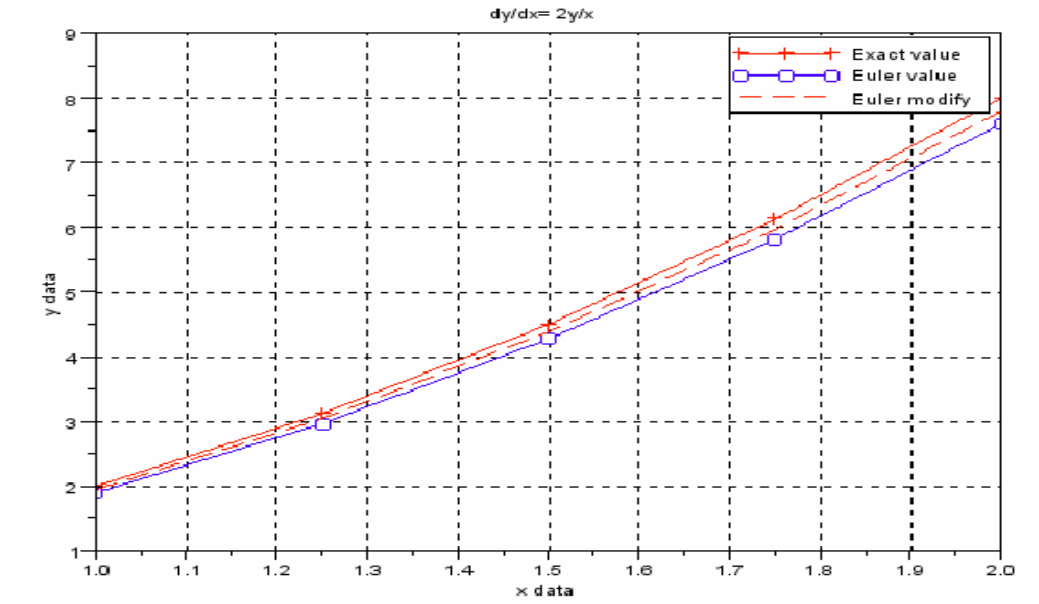



**Figure 1:** Comparison of Exact value with Euler's and Modify Euler's value.

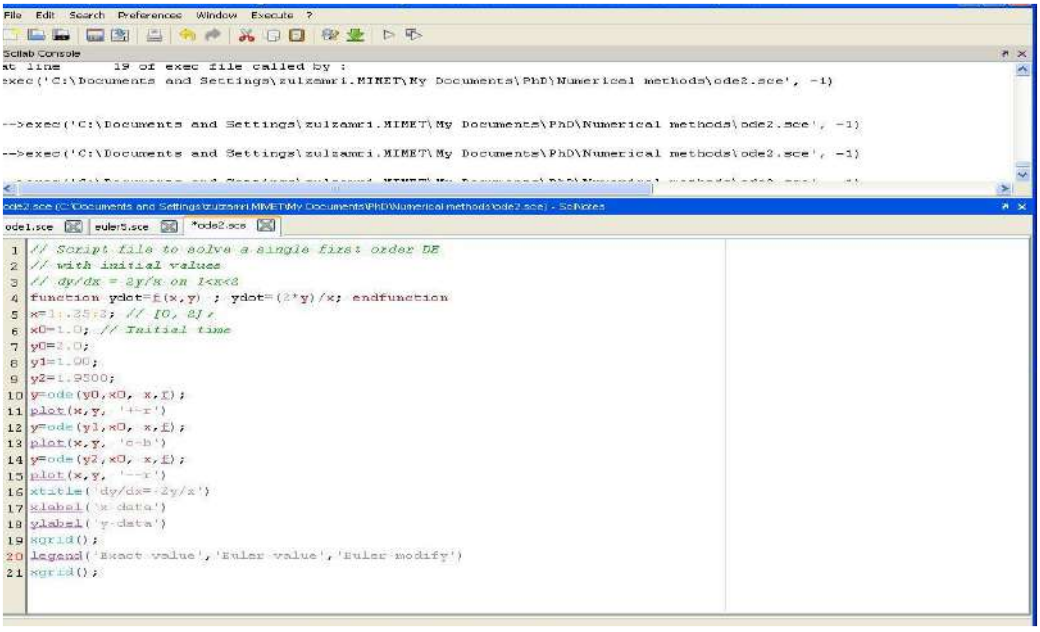



**Figure 2:** Scilab note and code

## 4. Conclusions:

The above graphical plots at Fig. 1 show the results obtained from different algorithms. Consequently, we can see better that modified Euler's algorithm is even more accurate with same step size (h = 0.25) than Euler algorithms. It can easily adapt these algorithm using Scilab codes as needed for a different type of problem. In numerical procedure, such as Euler's method, one must always keep in mind the question of whether the results are accurateenough to be useful. This was done whencompared between exact value and Euler's andmodification value. All these result willfinalized with their individual error's. Formodify Euler's, the error is small (< 1%) then led the accuracy of the numerical results andcould be ascertained directly by a comparisonwith the solution obtained analytically handcalculation of exact value.

## References:


1. Steven C. Chapra, Raymond P. Canale,Numerical Methods for Engineers, Mc Graw Hill, International Edition (6th Edition),2010.
2. Scilab_rachna_arvind_Sample_chapters_1_2and_2.pdf
3. Salleh. Z, Fundamental of Numerical Methods for Scientists and Engineers, Basic Application,Lambert Academic Publishing, 2011.
4. INRIA-ENPC, Scilab (Version 3.1.1). URL http://www.scilab.org/,2005.
5. S.Rao, Applied Numerical Methods for Engineers and Scientist, Pearson Prentice Hall Education, 2002.
6. K.C. Ang, A simple stochastic model for an epidemic numerical experiments with matlab, The Electronic Journal of Mathematics and Technology,Vol.1, No.2, 2009, pp 117-128.
7. Gilbeto.E.Urroz, Ordinary Differential Equations with SCILAB, In for Clearinghouse.com, 2001, pp 3-4.
8. D. Houcque, Matlab Application Ordinary Differential Equation, MathWorks Inc. 2010.
9. M.Podisuk, International Journal of Mathematical Models and Methods in Applied *Science*, Vol.5, No.2, 2011, pp 387- 394.
10. M.Kurulay, M.Bayram, European Journal of and Applied Mathematics,Vol.2, No.2,2009, pp 268- 277.
11. Zulzamri Salleh,Ordinary Differential Equations (ODE) using Euler's Technique and SCILAB Programming.

Contents available at www.balwantcollege.edu.in

**Emerging Trends in Applied Mathematics and Statistics**

***(Balwant College, Vita (M.S.) India 415 311)***

# Application of Vedic Mathematics in Remedial Primary School Mathematics Classes

Mahesh Pawar[1*], Nitin Jagtap[2], Rakeyshh Byakuday[3]

*[1,2]Assistant Professor, Department of Mathematics, Balwant College, Vita*

*[3]DKTE Society's Yashwantrao Chavan Polytechnic Ichalkaranji*

**Abstract:** Many Indian Secondary School pupils consider Mathematics a very difficult subject. Some pupils encounter difficulty with basic arithmetical operations. Some pupils arises difficult to manipulate symbols and balance equations. In other words, abstract and logical reasoning is their hurdle. Many such difficulties in learning Mathematics enter into a long list if prepared by an experienced teacher of Mathematics. Learning Mathematics is an unpleasant experience to some pupils mainly because it involves mental exercise. Vedic Mathematics is based on 16 formulas with the purpose of simplification of lengthy and cumbersome mathematics. The whole approach of Vedic math is suitable for slow learners as it is so simple and easy to use. Thus we have observed that these methods can be extended to long calculations. This gives enormous confidence to the student who lost their fear of numbers and go on to tackle harder math in a more open manner. Thus Vedic mathematics offers methods that are simpler, more efficient and more readily acquired than conventional methods.

**Key words:** Vedic philosophy, mathematics education, ancient Indian mathematics, mathematical knowledge, symbolic algebra, Vedic literature.

**1. Introduction:** Vedic mathematics - a gift given to this world by the ancient sages of India. Vedic Mathematics introduced by Jagadguru Swami Sri Bharati Krsna Tirthaji Maharaja (Sankaracharya of Govardhana Matha, Puri, Orissa, India). He was born in March 1884 to highly learned and religious parents. In 1957, when he decided finally to undertake a tour of the USA he rewrote from his memory the present volume of Vedic Mathematics giving an introductory account of the sixteen formulae reconstructed by him.

The Vedic Mathematics simplicity, means that calculations can be solved mentally though the methods can also be written. There are plenty advantages in using a flexible, mental system. Pupils can develop their own methods; they are not limited to proposed one. This follows to more creative, interested and intelligent pupils. Vedic Mathematics context to the technique of Calculations based on a set of 16 Sutras, or aphorisms, as algorithms and their upa-sutras or corollaries derived from these Sutras. Any mathematical problems (algebra, arithmetic, geometry or trigonometry) solved mentally with these sutras. Vedic Mathematics is more coherent than modern mathematics.

**Corresponding Author: Mr. Mahesh Pawar, Assistant Professor, Department of Mathematics, Balwant College, Vita (M.S.) India 415 311.* ***E-mail:*** *mahesh.msctech@gmail.com*

Vedic Mathematics offers a recent and more efficient approach to mathematics covering a wide range - begin with elementary multiplication and determination with a relatively advanced topic, the solution of non-linear partial differential equations. But the Vedic method is not simply a collection of quick methods; it is a system, a unified approach. Vedic Mathematics extensively fitted the properties of numbers in every practical application. Now we proceed on to give the 16 sutras and their corollaries

**2. The sixteen sutras and their corollaries are as follows:**

1. Ekadhikina Purvena -**By one more than the previous one** (Cor: Anurupyena)
2. Nikhilam Navatashcaramam Dashatah -**All from 9 and the last from 10** (Cor: Sisyate Sesasamjnah)
3. Urdhva-Tiryagbyham- **vertically and crosswise** (Cor: Adyamadyenantyamantyena)
4. Paraavartya Yojayet-**Transpose and adjust** (Cor: Kevalaih Saptakam Gunyat)
5. Shunyam Saamyasamuccaye-**When the sum is the same, that sum is zero**. (Cor: Vestanam)
6. Anurupye) Shunyamanyat-**If one is in ratio, the other is zero** (Cor: Yavadunam Tavadunam)
7. Sankalana-vyavakalanabhyam- **By addition and by subtraction** (Cor:Yavadunam Tavadunikritya Varga Yojayet)
8. Puranapuranabyham-**By the completion or non-completion** (Cor: Antyayordashak
9. Chalana-Kalanabyham- **Differences and Similarities** (Cor: Antyayoreva)
10. Yaavadunam-**Whatever the extent of its deficiency** (Cor: Samuccayagunitah)
11. Vyashtisamanstih-**Part and Whole** (Cor: Lopanasthapanabhyam)
12. Shesanyankena Charamena-**The remainders by the last digit** (Cor: Vilokanam)
13. Sopaantyadvayamantyam-**The ultimate and twice the penultimate** (Cor: Gunitasamuccayah Samuccayagunitah)
14. Ekanyunena Purvena-**By one less than the previous one** (Cor: Dhvajanka)
15. Gunitasamuchyah-**The product of the sum is equal to the sum of the product** (Cor: Dwandwa Yoga)
16. Gunakasamuchyah-**The factors of the sum is equal to the sum of the factors**

Low level of education has implications for the general quality of life and life conditions. Starting from the reality that today's schools have become meeting places of different cultures, religions, languages and points of view, the development of teachers' attitude towards culturally different pupils based on intercultural competence becomes crucial when dealing with Roma pupils. Student is the subject of the process of education, a teacher's partner in the common work, and the most important reason for the existence of the entire school system and the education system

When the education of the Indian children is concerned, the specific qualities of such a relationship must be taken into consideration, and that applies to teaching mathematics in the classroom as well. A Indian child approaches the educational process not only by bringing to it his or her individuality and competence, but also as a member of a culture and a language that is, in a way, different from the culture from which most of the children in our schools approach the teaching process. For this reason, the approach to teaching, including

teaching mathematics, poses a challenge not only for the Indian student, but for the teachers who are not familiar with the specific characteristics of the Indian culture. That is why it is necessary, when working with the Indian pupils, to find a methodological approach to teaching mathematics in which a teacher will try to adjust the learning process on some new level, with the aim of better and easier acquisition of mathematical contents.

In an attempt to find a better approach to learning mathematics, as an alternative option it is possible to apply calculation by using Vedic mathematics. Vedic mathematics is a calculation system based on 16 sutras (Sanskrit formulas) the basic feature of which is a simplicity of calculation without the written computation, which can stimulate the pupils' interest in mathematical thinking and creativity in finding solutions to mathematical problems. In contemporary researches, scientists are trying to find alternative usages and importance of Vedic mathematics apart from its computational aspect. In that process, they are discovering a correlation with various aspects of a daily life, with other sciences, as well as the width of observing through the significance of sutras for everyday work and activities; their philosophy and meaning are applicable not only in the mathematical sense, but also to the events and order of social achievements, moral development, interest in individual and social progress The 16 sutras, through formulas which are expressed in words and are easily understandable, memorable, and applicable, represent ways of solving mathematical problems in the areas of arithmetic, algebra, geometry and calculus

Sutras which are the focus of this work are related to mathematical contents and ways of calculation that could be applied in extracurricular work with the Indian pupils, particularly in remedial mathematics classes. Precisely because of all the above mentioned characteristics, sutras might be a useful alternative or a supplement to a traditional method of teaching, in cases where such a method of teaching does not achieve the expected results or educational progress.

Some Ancient Vedic Methods:

The use of Vedic mathematics lies in the fact that it reduces the typical calculations in conventional mathematics to very simple ones. This is so because the Vedic formulae are claimed to be based on the natural principles on which the human mind works. Vedic Mathematics is a methodology of arithmetic rules that allow more efficient speed implementation. This is a very interesting field and presents some effective algorithms which can be applied to various branches of engineering such as computing.

## 3. Elementary Examples of How Vedic Sutras Provide in terms of their Algorithms

Addition using the Shuddhah Sutra,

Example: 93 + 28 + 65

Algorithm: Shuddhah addition follows this procedure.

1) When the sum reaches 10 or more, purify 10 out of the sum (take away 10) and note it as a dot.
2) Addition continues with the remaining amount.
3) The number of dots in a column is the amount to be carried.

```
  9● 3
  2  8●
+ 6  5
 --------
1  8  6
```

Beginning in the ones column, 3 + 8 = 11. Take away 10 from the sum and note the transaction as a dot. Continue adding with 1: 1 + 5 = 6. Record the 6. There is 1 dot in the ones column so carry 1. In the tens column, 1 + 9 = 10. Record the 10 as a dot and continue adding: 2 + 6 = 8. There is 1 dot in the tens column which is carried to the hundreds place of the answer: 186.

Subtraction with the Shuddhah Sutra,

Example: 5125 – 3608

Algorithm: The procedure for Shuddhah subtraction as follows.
1) If lower digit is less than or equal to the upper, subtract.
2) If lower digit is larger than the upper digit, then:
A) Write a dot beside the next digit in the lower portion of the problem. The dot is a reminder to increase that digit by 1 and thereby purify 10, or a power of 10, out of the upper portion of the problem (minuend).

B) Add the complement of the lower digit to the upper digit and record the sum.

```
  5 1  2  4
- 3●6 0 ●8
 -------------
  1  5 1  6
```

Beginning in the ones column, 8 is larger than 4 so write a dot near the 0. Add the complement of 8 to the four: 2 + 4 = 6. In the tens column, the dot beside the 0 reminds the student to add 1 to the 0. One is less than 2 so subtract normally:
2 - 1 = 1. In the hundreds column, 6 is greater than 1 so write a dot near the 3. Add the complement of 6 to the 1: 4 + 1 = 5. In the thousands column, 4 is less than 5 so subtract normally: 5 - 4 = 1.

Multiplication with the Urdhva Sutra, meaning vertically and crosswise

Example: 27 x 34

Algorithm: The following procedure is for problems of two digits multiplied by two digits.

```
  2 7
x 3 4
-------
9/1/8
  3   2
```

1) Multiply vertically in the ones column:
(4x7) = 28. Record the 8 and carry the 2.
2) Multiply crosswise and add the two products:
(4x2)+(3x7) = 8+21 = 29. Add the carry, 29+2=31. Record the 1 and carry the 3.
3) Multiply vertically in the tens column:
(3x2) = 6. Add the carry, 6+3=9.
The answer is 918.

Special case multiplication when both factors are near a base of 10 using the Nikhilam Sutra, meaning all from nine and last from ten:

Example: 92 x 97

Algorithm: When applying Nikhilam Sutra, all digits are subtracted from 9 except the last one (the ones place digit in this example) which is subtracted from 10.
1) Apply Nikhilam Sutra to the factors:
For 92, 9-9=0 and 10-2=8 so the deviation from the base of 100 is -08.
For 97, 9-9=0 and 10-7=3 so the deviation from the base is -03.
Now rewrite the problem as shown:
2) The right-hand side of the answer is found by multiplying the deviations: -03 x -08 = 24.
3) The left-hand side of the answer is found by adding diagonally in either direction:
92 + (-03)=89 or 97+(-08)=89. The answer is 8,924.

```
  92  -08
x 97  -03
----------
89 / 24 = 8,924
```

**4. Conclusions:**

The whole approach of Vedic math is suitable for slow learners as it is so simple and easy to use. Thus we have observed that these methods can be extended to long calculations. This gives enormous confidence to the student who lost their fear of numbers and go on to tackle harder math in a more open manner. When children learn about algebra and Pythagorean Theorem then they are easily use square root and square of numbers. A few observations are stated hereunder

- Mathematics, derived from the Veda, provides one line, mental and superior methods along with fast cross checking systems.
- Vedic Mathematics converts a tedious subject into a playful and blissful one which pupils learn with smiles.
- Vedic Mathematics offers a new and entirely different approach to the study of Mathematics based on pattern recognition. It allows for constant expression of a student's creativity, and is found to be easier to learn.
- In Vedic mathematics for any problem, there is always one general technique applicable to all cases and also a number of special pattern problems. The element of choice and flexibility at each stage keeps the mind lively and alert to develop clarity

of thought and intuition, and thereby a holistic development of the human brain automatically takes place.

- Vedic Mathematics with its special features has the inbuilt potential to solve the psychological problem of Mathematics.

Thus Vedic mathematics offers methods that are simpler, more efficient and more readily acquired than conventional methods. It brings out the beauty and pattern in numbers and the world around us.

**References:**

[1] Brooks, T. & Brooks, K. (1988, Spring). Vedic mathematics: Taking the anxiety out of math. In Maharishi International University, Maharishi International University News Maharishi International University Press.

[2] Cajori, F. (1897). A history of elementary mathematics with hints of methods and teaching. London: MacMillan and Company.

[3] Glover, James T. Vedic Mathematics for Schools Book 1, Motilal Banarsidass Pub. Pvt. Ltd., Delhi 2004.

[4] Jagadguru Swami Sri Bharati Krisna Tirthaji Maharaja,"Vedic mathematics", Motilal Banarsidass Publishers Pvt Ltd,Delhi, 2009

[5] Jan Hogendijk, Vedic Mathematics and the Calculations of Guru Tirthaji, Secrets of Ancient Mathematics, 24-27, (2004).

[6] Jones, C.H. (1989). The impact of Maharishi Vedic Science Based Education in higher education: The example of Maharishi International University. Modern Science and Vedic Science, 3.

[7] Kandasamy, W. B. V.; Smarandache, F. (2006). Vedic mathematics-'Vedic' or 'mathematics': a fuzzy & neutrosophic analysis.

[8] Maharishi Mahesh Yogi. (1986). His Holiness Maharishi Mahesh Yogi: Thirty years around the world-Dawn of the age of enlightenment, Vol. 1, Netherlands: Maharishi Vedic University Press.

[9] Muehlman, J.M. (1994). Maharishi's Vedic Mathematics at the elementary level: Improving achievement, affect, and mental mathematics through Vedic sutra based computation.

[10] Nicholas, A.P., Williams, K.R., & Pickles, J. (1984). Applications of the Vedic mathematical Sutra Ûrdhva-tiryagbhy. vertically and crosswise. Vedic Mathematics Research Group, Roorkee University, Roorkee, India.

[11] Wallace, R.K., Fagan, J.B., & Pasco, D.S. (1988). Vedic physiology. Modern Science and Vedic Science

[12] Williams, K.R. (1984). Discover Vedic mathematics: A practical system based on sixteen simple formulae. Available through Vedic Mathematics Research Group, Roorkee University, Roorkee India.

[13] http://en.wikipedia.org/wiki/Vedic_mathematics

[14] http://www.vedicmaths.org

[15] http://www.cs.uml.edu/~asaxena/vedic-maths

Contents available at www.balwantcollege.edu.in
**Emerging Trends in Applied Mathematics and Statistics**
***(Balwant College, Vita (M.S.) India 415 311)***

# Solution of quadratic equation by using C++

Tushar Suryawanshi[1*]
[1]*Assistant Professor, Department of Mathematics, Balwant College, Vita*

**Abstract:** Utilizing C++ programme application, this research is used to design the software applications. The C++ language was used to create this package. using a Dev-C++ compiler is used to run, build, and discover programme flaws. Usually, quadratic equations are used to solve major mathematical algorithmic faults. We can sketch out a solution to this quadratic to use this application. In consideration of this, the study also uncovered the students' outstanding performance. In view of the implications of the findings, it was advised that computer educational software be designed and developed for use in classroom instruction. However, to be aware of these, one must be familiar with numerical mathematics.

***Key Words*:** Application Program, Quadratic equation, C++.

**1. Introduction:** Technology advancements are used to enhance pedagogy, student learning, and mathematical education, as well as to build more effective organisational structures. Modern math technology frequently takes the shape of computer programmes. These technological tools benefit students' cognitive processes by reducing memory burdens and increasing their knowledge of the problem-solving process. These aid logical reasoning by making hypothesis testing straightforward. These recordings of the problem-solving process—the fits, starts, and different paths pupils take—can be played again to gain insight into their minds. Classroom studies usually come into play for everyone.

### 1.1 Quadratic Equation

An equation is created when a polynomial is set equal to a value, such as an integer or another polynomial. Quadratic Equation is the name given to an equation that has the form $ax^2 + bx + c = 0$. By following the laws of algebra, employing factoring strategies where appropriate, and the principle of zero products, you may solve a quadratic equation**.**

### 1.2 The Principle of Zero Products

According to the Principle of Zero Products, if the sum of two numbers is zero, at least one of the elements must also be zero. (This is actually nothing new.)

If ab is equal to 0, then either an or b, or both a and b, are zero.

Although this characteristic may seem rather clear, it has significant implications for the solution of quadratic equations. A factored polynomial that is equal to zero indicates that either all or both of the factors are equal to zero. This approach can be used to resolve quadratic equations.

***Corresponding Author:** Mr.* Tushar Suryawanshi, *Assistant Professor, Department of Mathematics, Balwant College, Vita (M.S.) India 415 311.* ***E-mail:*** *suryawanshit15@gmail.com*

## 2. Factoring the Quadratic

Often, factoring the quadratic, setting each factor equal to zero, and then solving each factor is the quickest way to find the answer to "$ax^2 + bx + c = 0$" given the value of x. However, there are situations when the quadratic is too complicated, it does not factor at all, or you're just not in the mood to factor. The Quadratic Formula always can locate the answer, even while factoring might not. The "a", "b", and "c" in "$ax^2 + bx + c$," which are just numbers and are the "numerical coefficients" of the quadratic equation they have given you to solve, are used in the quadratic formula. The quadratic formula, which is officially written as follows, is obtained from the square-filling procedure.

The values of x that are the solutions of the equation for $ax^2 + bx + c = 0$ are given by:

$$x = -b + \sqrt{(b^2 - 4ac)}/2a$$

$$x = -b - \sqrt{(b^2 - 4ac)}/2a$$

Your equation must be set up in the form "(quadratic) = 0" in order for the quadratic formula to function. Additionally, not just the square root, but also the "2a" appears beneath everything above in the denominator of the Formula. And underneath, instead of just a "2," there is a "2a." I can almost guarantee that if you forget to "insert them back in" on the test, you'll mess up because you dropped the square root or the "plus/minus" in the middle of your calculations. Always keep in mind that "$b^2$" refers to "the square of ALL of b, including its sign," thus even if b is negative, the square of a negative is positive.
Here are a few illustrations of how the quadratic formula functions.
$x^2 + 3x - 4 = 0$
Interestingly, this quadratic factors:
$x^2 + 3x - 4 = (x + 4)(x - 1) = 0$

**Steps to solve the given equation**

1. Rearrange the equation to the standard quadratic form $ax^2 + bx + c = 0$:
   $x^2 + 3x - 4 = 0$
2. Find the values of a, b and c:
   a = 1, b = 3, c = -4
3. Find the discriminant ($b^2$ - 4ac):
   $b^2$ - 4ac = 9 - 4(1)(-4) = 25
4. Use the discriminant to determine the number of solutions:
   The discriminant is greater than 0, so there are two solutions.
5. Solve the equation using the Quadratic Formula:
   x = (-b +/- sqrt($b^2$ - 4ac)) / 2a
   x = (-3 +/- sqrt(25)) / 2
   x = (-3 +/- 5) / 2
6. Find the solutions:
   x = -3/2 + 5/2 = 1
   x = -3/2 - 5/2 = -4

### 2.1 Methods to Solve

There are several ways to solve quadratic equations as follows:

1. Manual factoring
2. Completing the square

### 2.2 Completing the Square

The algebraic identity, which stands for a clear technique that can be utilised to solve any quadratic problem, is used in the process of completing the square. $ax^2 + bx + c = 0$ is a quadratic equation in standard form.

1. Divide a, the squared term's coefficient, from each side.
2. Subtract both sides from the constant term c/a.
3. Increase both sides by the square of the coefficient of x, half of b/a. By doing so, the left side becomes a perfect square and "completes the square."
4. If necessary, simplify the right side and write the left side as a square.
5. Create two linear equations by adding the positive and negative square roots of the right side to the square root of the left side.
6. Conquer the two linear problems.

### 2.3 Program

```
#include <iostream>
#include<cmath>
using namespace std;
int main ()
{
float a, b, c, OneSolution, Solution1, Solution2,
determinant;
cout<<"Quadratic Equation Form ax^2 + bx + c \n";
cout<<"Enter Integer for a: ";
cin>>a;
cout<<"Enter Integer for b: ";
cin>>b;
cout<<"Enter Integer for c: ";
cin>>c;
if (a == 0)
   if (b == 0)
      if (c == 0)
         cout<<"All Solutions"<<endl;
      else
         cout<<"NO Solutions" <<endl;
else
{
OneSolution = -c/b;
cout<<"Solution for "<<a<<"x^2 + " <<b <<"x +" <<c<<"\nx = "<<One Solution <<endl;
}
```

```
else
{
determinant = b * b - 4 * a * c;
if (determinant < 0)
cout<<"Solution for "<<a<<"x^2 + " <<b <<"x +" <<c
<<"\nNo Solution" <<endl;
else if (determinant == 0)
{
cout<<"Solution for "<<a<<"x^2 + " <<b <<"x +" <<c
<<"\nDuplicate Solution" <<endl;
OneSolution = (-b/2) * a;
cout<<OneSolution<<endl;
}
else
{
cout<<"two solutions" <<endl;
Solution1 = (-b + sqrt(determinant)) / 2 * a;
Solution2 = (-b - sqrt(determinant)) / 2 * a;
cout<<"Solution for "<<a<<"x^2 + " <<b <<"x +" <<c
<<"\nx1 = " << Solution1 <<endl<< "x2 = "<<Solution2 <<endl;
}
}
system("pause");
}
```

**Output:**

```
Quadratic Equation Form ax^2 + bx + c
Enter Integer for a: 1
Enter Integer for b: 2
Enter Integer for c: 1
Solution for 1x^2 + 2x +1
Duplicate Solution
-1
sh: 1: pause: not found
```

**Examples Results Works**

0x^2 + 0x + 0 = 0 Answer = "All Solutions"
0x^2 + 0x + 1 = 0 Answer = "No Solutions"
0x^2 + 2x + 1 = 0 Answer = -1/2
1x^2 + 2x + 1 = 0 Answer = "Duplicate Solution" = -1
1x^2 + 4x + 1 = 0 Answer = "two solutions" = d = 3.464 1st root = -0.267 | 2nd root = -3.73

### 2.4 Advantages using Application software

i. Quick feedback, as the software enables users or students to determine whether or not the problem is accurate right away.
ii. Perseverance in practising the calculations
iii. Encourages kids to work in private spaces.

## 3. Conclusions:

1. Despite the fact that technology is evolving quickly these days, classroom instruction must still use it.
2. The use of an application to solve a mathematical problem increases the independence of the next generation.
3. Effectively use technology to expand learning opportunities and enhance student potential.
4. This quadratic equation is one of the most widely utilised in applied mathematics, making its real application one of the most well-understood computer algorithms.
5. We are pushing the boundaries of what is currently known about computers, even with the simplest challenge.
6. The robustness of these formulas for the quadratic equations must be the focus of attention.

## References:

1. W.H. Press et al., Numerical Recipes in C, Cambridge Univ. Press, 1988, p. 184.
2. Carter, M. B. (2004).Dissertation Abstracts International, 65, 1288.
3. Connell, M. (1998).Technology in Constructivist Mathematics Classrooms. Journal of Computers in Mathematics and Science Teaching, 17(4) ,311-338.
4. Cradler, J., McNabb, M., Freeman, M., & Burchett, R. (2002). How does technology influence student learning? Learning & Leading with Technology, 29(8), 47-56.
5. Dede C.(2000).Rethinking how to Invest in Technology. In D. Roy (Ed.), Technology and Learning. San Francisco
6. Jossey -Bass Doering, A &Veletsianos G.(2009). Teaching with Instructional Software in M.D Roblyer & A. Doering (Eds)

## ABOUT BALWANT COLLEGE, VITA

Balwant College, Vita is one of the biggest and well known multi-faculty college in the jurisdiction of Shivaji University, Kolhapur, Maharashtra. It was established in 1963 at the behest of Late Shri. Yashwantrao Chavan Saheb, supported generously by local community. It is run by Rayat Shikshan Sanstha to provide quality education to deprived strata of the society. Today, college offers 40 different academic programs through 14 UG and 11 PG Departments. Students are benefitted with the large number of support services and inter-disciplinary activities. The college is accredited with 'A' grade by NAAC. The college strives continuously to achieve excellence in all the spheres of academic activities. The college has also well-developed competitive guidance center, rich central library with internet facility and eco-friendly Wi-Fi campus.

## ABOUT DEPARTMENT OF MATHEMATICS

The Department of Mathematics is quite efficient our Department in Balwant College, Vita was started in academic year 1975 and conduct undergraduate courses of Shivaji University, Kolhapur and post graduate courses of YCMOU, Nashik. This department is actively engaged in all-round activities with the success of all its past students. The faculty members are recognized as leaders in their fields well and qualified. The objective of the Department of Mathematics has been to bestow quality education to students. The atmosphere of the department is student centric and healthy. The department is committed to up skill the students through multifarious curricular, co-curricular and extra-curricular activities throughout the year. The department having well equipped laboratory with LCD, Wi-Fi facility, rich in study materials.

## ABOUT DEPARTMENT OF STATISTICS

Statistics plays a key role in the development of modern sciences, management and many other important applied areas. With this aim, the Department of Statistics was established in 1977 to attain excellence in teaching and research. The Department was introduced B.Sc.-I, B.Sc.-II, B.Sc.-III UG courses in June 1977, June 1978 and June 2016 respectively under Shivaji University, Kolhapur. In June 2021, the Department has introduced a M.Sc. (Statistics) course for post graduate students. Apart from teaching, the department has actively participated in research activity conducted by various research institute and Universities all over India. Many alumni of department are currently working in India and abroad. Department is trying to contribute to the development of this subject by innovative teaching - learning with various research activities.

## ABOUT THE BOOK

The proceedings book entitled "Emerging Trends in Applied Mathematics and Statistics" jointly composed by Department of Mathematics and Statistics, Balwant College, Vita aimed to provide a collaborative platform for students, young researchers, academicians to publish their research works in the field of applied Mathematics and Statistics. It will enlighten the participants about emerging trends in applied Mathematics and Statistics.

**ISBN: 978-81-956739-5-7**

Prarup Publication, Kolhapur