

---

# Applied Statistics for Agriculture, Veterinary, Fishery, Dairy and Allied Fields

---

Pradip Kumar Sahu

# Applied Statistics for Agriculture, Veterinary, Fishery, Dairy and Allied Fields

 Springer

Pradip Kumar Sahu  
Department of Agricultural Statistics  
Bidhan Chandra Krishi Viswavidyalaya  
Mohanpur, WB, India

ISBN 978-81-322-2829-5      ISBN 978-81-322-2831-8 (eBook)  
DOI 10.1007/978-81-322-2831-8

Library of Congress Control Number: 2016958114

© Springer India 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature  
The registered company is Springer (India) Pvt. Ltd.

*To  
Kenchu, Mechu, Venpu and Khnako*

---

## Preface

Statistics is now recognized and universally accepted a discipline of science. With the advent of computer technologies, the use of statistics has increased manifold. One can hardly find any area where there is no use of statistics. In the field of Biological Sciences, the use of statistics is also keeping pace with other disciplines. In fact development of many statistical theories has their roots in biological sciences, in particular agricultural sciences. This has led to ever increasing areas of its application in diversified fields. Newer and varieties of problems are being tackled by the subject. Like other branches of science, statistics is being extensively used in agricultural/animal/fishery/dairy and other fields in explaining various basic as well as applied problems. Availability of wide range of statistical techniques suited for various problems has made it possible for its wider application. Everyday type of problem is getting increased and more and more tools or techniques need to be developed to solve various specific problems. Development and/or selection of appropriate statistical technique for a given problem is mostly warranted for getting meaningful explanation of the problems under consideration.

Students/teachers/researchers/practitioners from agriculture and allied fields are to deal with various factors like living flora and fauna, soil, air, water, nutrients, etc. along with socio-economic and behavioral aspects of plant and animal beings for successful research and development. Understanding of the theory and essence of both the agricultural science and the theory of statistics is a must for getting and explaining the problem under consideration in a meaningful way. It is felt increasingly that a user in any field should have well understanding of the logic behind any experimentation as well as the specific statistical tools (during planning, designing, executing, collecting information/data, analytical methods and drawing inference from the results) to be used to draw meaningful conclusion from the experiment.

Statistics is a mathematical science in association with uncertainty. There is a large section of students/teachers/researchers/practitioner who do not have enough mathematical orientation and as such are scares of using statistics, in spite of its wider acceptability. To reach to these huge users remains a challenging task to the statisticians, particularly the biostatisticians. Statistics must reach to the users particularly to these types of user in their terms/manners and language. Biological sciences have moved

on from mostly simple qualitative description to concepts founded on numerical measurements and counts. In order to have proper understanding of phenomena, correct and efficient handling of these measurements is needed and actually done by statistics. Understanding of basic statistics is essential for planning measurement programs and for analyzing and interpreting data but frequently it has been observed that many users lack in good comprehension of statistics, moreover do not feel comfortable while making simple statistics based decisions. A number of books are available, which deal with various aspects of statistics. The need for the present book has been crept in to the mind of the author during his teaching experience. In India only, there are more than hundred colleges where agriculture, veterinary, fishery, dairy and home science are taught at graduation and post-graduation levels as per the syllabi of the Indian Council of Agricultural Research. Outside India, millions of students are there in these wings. A textbook to cater the need of these types of students with a guide to handle their data using easily available statistical software is mostly needed. An attempt has been made in this book to present the theories of statistics in such a way that the students and researchers from biological/agricultural/animal/fishery/dairy and allied field find it easy to handle and use in addressing many real life problems of their respective fields.

This book starts with an introduction to the subject which does not require any previous knowledge about the subject. The ultimate aim of the book is to make it self-instructional textbook, which can be helpful to the users in solving their problems using statistical tools also with the help of simple and easily available computer software like MSEXCEL. It is expected that thousands of students of biological/agricultural/animal/fishery/dairy and allied fields would be benefitted from this book. In each chapter, theories have been discussed with the help of example(s) from real life situations, followed by worked out examples. Simple easily available packages like MSEXCEL, SPSS, etc. have been used to demonstrate the steps of calculation for various statistical problems. Statistical packages used for demonstration of analytical techniques are gratefully acknowledged. Attempts have been made to familiarize the problems with examples on each topic in lucid manner. Each chapter is followed by a number of solved problems (more than 165) which will help the students in gaining confidence on solving those problems. Due care has been taken on solving varied problems of biological/agricultural/animal/fishery/dairy and allied fields and the examination need of the students. It has got 13 chapters. The first chapter is to address and explain the subject statistics, its usefulness and application with particular reference to biological/agricultural/animal/fishery/dairy and allied fields. A brief narration on statistics, highlighting its use, scope, steps in statistical procedure and limitations along with example, has been provided in Chap. 1. Main ingredient of statistics is the varied range of information or data; in second chapter, attempts have been made to explain different types of information/data from relevant fields. In this chapter, discussion has been made on collection, scrutinisation and presentation of data in different forms so as to have first-hand idea about the data. The third chapter deals with measures of central tendency and measures of dispersion along with

skewness and kurtosis. Different measures of central tendencies and dispersion along with their uses, merits and demerits have been discussed. Measures of skewness and kurtosis have also been discussed. The theory of probability has been dealt in Chap. 4. Utmost care has been taken to present the theory of probability in its simplest form, starting from the set theory to the application of different laws of probability. Quite a good number of examples on probability theory and random variable are the special features of this chapter. A few discrete and continuous probability distributions like Binomial, Poisson, Normal,  $\chi^2$ , t and F have been discussed in brief. Introductory ideas about population, types of population, sample, sampling techniques used under different situations, comparison of sample survey techniques and census have been discussed in Chap. 5. Statistical inference has been discussed in Chap. 6. Starting with the introduction of statistical inference, both statistical estimation and testing of hypothesis have been discussed in this chapter. Tests based on distributions mentioned in Chap. 4 have been discussed. Discussions on different non-parametric tests included in this chapter hope to find their applications in various agriculture and allied fields. These tests have been designed with an objective to cater the need of the students of agriculture/animal science/dairy/fishery and allied fields as per the syllabi of the Indian Council of Agricultural Research. Chapter 7 is devoted to the study of correlation. Starting with the idea of bivariate data, bivariate frequency distribution and covariance, this chapter has described the idea of simple correlation and its properties, significance and rank correlation. The idea of regression, need, estimation of parameters of both simple and multiple regression, meaning and interpretations of parameters, test of significance of the parameters, matrix approach of estimation of parameters, partitioning of total variance, coefficient of determination, game of maximization of  $R^2$ , adjusted  $R^2$ , significance test for  $R^2$ , problem of multicollinearity, regression vs. causality, part and partial correlation are discussed in Chap. 8. Discussion on properties and examples are the special features of the correlation and regression chapters. Starting with general idea, the analysis of variance technique has been discussed in Chap. 9. Extensive discussion has been made on assumptions, one-way analysis of variance (with equal and unequal observations), two-way analysis of variance (with one or more than one observations per cell), violation of the assumptions of ANOVA vis-a-vis transformation of data, effect of change in origin and scale on analysis of variance with worked-out examples. Chapter 10 is devoted to basics of experimental design and basic experimental designs. This chapter discusses on experiment, types of experiments, treatment, experimental unit, experimental reliability, precision, efficiency, principles of design of field experiments – replication, randomization, local control, lay out, uniformity trial and steps in designing field experiments. In this chapter, elaborate discussion has been made on completely randomized design, randomized block design and latin square design along with missing plot techniques.

Efforts have been made to explain the basic principles and procedures of factorial experiments in Chap. 11. Factorial experiments, their merits and demerits, types of factorial experiments, two factor factorial (symmetrical and asymmetrical) CRD, two factor factorial (symmetrical and

asymmetrical) RBD, three factor factorial (symmetrical and asymmetrical) CRD, three factor factorial (symmetrical and asymmetrical) CRD, split plot and strip plot designs are discussed in this chapter. Some special types of experimental designs which are useful to the students, teachers, researchers and other users in agriculture and allied fields have been discussed in Chap. 12. In this chapter, attempt has been made to discuss on augmented CRD and RBD, augmented designs with single control treatment in factorial set up, analysis of combined experiments, analysis of data recoded over times and experiments at farmers fields. Computer has come in a great way to help the experimenter not only in analysis of experimental data but also in different ways. But there has been a tendency of using computer software without providing due consideration to 'what for', 'where to use', 'which tool is to use' and so on. In last chapter of this book, an attempt has been made, by taking example, to show how computer technology can be misused without having knowledge of appropriate statistical tools.

A great number of books and articles in different national and international journals have been consulted during preparation of this book which provided in reference section. An inquisitive reader will find more material from these references. The need of the students/teachers/researchers/practitioners in biological/agricultural/animal/fishery/dairy and allied fields remained the prime consideration during the preparation of this book.

I express my sincere gratitude to everyone who has helped during the preparation of the manuscripts for the book. The anonymous international reviewers who have critically examined the book proposal and put forwarded their valuable suggestions for improvement of the book need to be acknowledged from the core of my heart. My PhD research students, especially Mr Vishawajith K P, Ms Dhekale Bhagyasree, Md Noman, L Narsimaiah and others, who helped a lot during analysis of the examples based on real life data and need to be acknowledged. Taking the help of MSEXCELL, SPSS and SAS softwares various problems have been solved as examples in this book; the author gratefully acknowledges the same. My departmental colleagues and our teachers at BCKV always remained inspiration to such book projects, thanks to them. My sincere thanks to the team of Springer India in taking responsibility of publishing this book and continued monitoring during the publication process. Most importantly my family members, who have always remained constructive and inspirational for such projects need to be thanked; without their help and co-operation it would have not been possible to write such a book. All these will have a better success if this book is well accepted by the students, teachers, researchers and other users for whom this book is meant for. I have the strong conviction that like other books written by the author, this book will also be received by the readers and will be helpful to everyone. Sincere effort are there to make the book error free, however any omissions/mistake pointed out, along with constructive suggestions for improvement will be highly appreciated and acknowledged.



---

# Contents

<b>1</b>	<b>Introduction to Statistics and Biostatistics</b> . . . . .	1
1.1	Introduction . . . . .	1
1.2	Use and Scope of Statistics . . . . .	1
1.3	Subject Matter of Statistics . . . . .	2
1.4	Steps in Statistical Procedure . . . . .	2
1.5	Limitation of Statistics . . . . .	7
<b>2</b>	<b>Data–Information and Its Presentation</b> . . . . .	9
2.1	Data . . . . .	9
2.2	Character . . . . .	12
2.3	Variable and Constant . . . . .	12
2.4	Processing of Data . . . . .	13
2.5	Classification/Grouping . . . . .	16
2.5.1	Method of Classification . . . . .	18
2.5.2	Cumulative Frequency . . . . .	18
2.5.3	Relative Frequency . . . . .	18
2.5.4	Frequency Density . . . . .	18
2.6	Presentation of Data . . . . .	20
2.6.1	Textual Form . . . . .	21
2.6.2	Tabular Form . . . . .	22
2.6.3	Diagrammatic Form . . . . .	24
<b>3</b>	<b>Summary Statistics</b> . . . . .	35
3.1	Measures of Central Tendency . . . . .	36
3.1.1	Arithmetic Mean . . . . .	37
3.1.2	Geometric Mean . . . . .	39
3.1.3	Harmonic Mean . . . . .	42
3.1.4	Use of Different Types of Means . . . . .	43
3.1.5	Median . . . . .	44
3.1.6	Partition Values (Percentiles, Deciles, and Quartiles) . . . . .	46
3.1.7	Mode . . . . .	48
3.1.8	Midpoint Range . . . . .	49
3.1.9	Selection of Proper Measure of Central Tendency . . . . .	50

3.2	Dispersion and Its Measures . . . . .	51
3.2.1	Absolute Measures of Dispersion . . . . .	51
3.2.2	Moments . . . . .	62
3.2.3	Relative Measures of Dispersion . . . . .	69
3.3	Skewness and Kurtosis . . . . .	70
3.3.1	Skewness . . . . .	71
3.3.2	Kurtosis . . . . .	73
<b>4</b>	<b>Probability Theory and Its Application . . . . .</b>	<b>77</b>
4.1	Introduction . . . . .	77
4.2	Types of Set . . . . .	78
4.3	Properties of Sets . . . . .	79
4.4	Experiment . . . . .	80
4.5	Probability Defined . . . . .	81
4.5.1	Important Results in Probability . . . . .	82
4.6	Random Variables and Their Probability Distributions . . . . .	85
4.7	Mean, Variance, and Moments of Random Variable . . . . .	86
4.8	Moment-Generating Function . . . . .	89
4.9	Theoretical Probability Distributions . . . . .	91
4.9.1	Binomial Distribution . . . . .	91
4.9.2	Poisson Distribution . . . . .	96
4.9.3	Normal Distribution . . . . .	100
4.10	Central Limit Theorem . . . . .	106
4.11	Sampling Distribution . . . . .	107
4.11.1	$\chi^2$ -Distribution . . . . .	107
4.11.2	$t$ -Distribution . . . . .	108
4.11.3	$F$ Distribution . . . . .	109
4.11.4	Sampling Distribution of Sample Mean and Sample Mean Square . . . . .	110
4.11.5	Fisher's $t$ -Distribution and Student's $t$ -Distribution . . . . .	111
<b>5</b>	<b>Population and Sample . . . . .</b>	<b>113</b>
5.1	Population . . . . .	113
5.2	Sample . . . . .	113
5.3	Parameter and Statistic . . . . .	115
5.4	Estimator . . . . .	115
5.5	Subject Matter of Sampling . . . . .	115
5.6	Errors in Sample Survey . . . . .	116
5.7	Sample Size . . . . .	116
5.8	Selection of Sample (Sampling Technique) . . . . .	118
5.9	Different Sampling Techniques . . . . .	119
5.9.1	Probability Sampling . . . . .	119
5.9.2	Non-probability Sampling . . . . .	130

<b>6</b>	<b>Statistical Inference</b> . . . . .	133
6.1	Introduction . . . . .	133
6.1.1	Estimation . . . . .	134
6.1.2	Testing of Hypothesis . . . . .	139
6.2	Testing of Hypothesis . . . . .	140
6.2.1	Parametric Tests . . . . .	141
6.3	Nonparametric Method . . . . .	176
6.3.1	One Sample Test . . . . .	176
6.3.2	Two Sample Test . . . . .	182
<b>7</b>	<b>Correlation Analysis</b> . . . . .	195
7.1	Introduction . . . . .	195
7.2	Correlation Coefficient . . . . .	196
7.3	Properties . . . . .	197
7.4	Significance of Correlation Coefficients . . . . .	203
7.5	Correlation Coefficient of Bivariate Frequency Distribution . . . . .	204
7.6	Limitations . . . . .	206
7.7	Rank Correlation . . . . .	207
7.8	Correlation Ratio . . . . .	209
7.9	Properties of Correlation Ratio . . . . .	211
7.10	Coefficient of Concurrent Deviation . . . . .	211
7.11	Calculation of Correlation Coefficient Using MS Excel, SPSS, and SAS . . . . .	212
<b>8</b>	<b>Regression Analysis</b> . . . . .	223
8.1	Introduction . . . . .	223
8.2	Explanation of the Regression Equation . . . . .	224
8.3	Assumption of Linear Regression Model . . . . .	224
8.4	Simple Linear Regression Analysis . . . . .	225
8.5	Properties of Regression Coefficient . . . . .	230
8.5.1	Regression Coefficient . . . . .	230
8.5.2	The Sign of the Regression Coefficient . . . . .	231
8.5.3	Relation Between Correlation Coefficient and the Regression Coefficients . . . . .	231
8.5.4	Relation Between Regression Coefficients . . . . .	231
8.5.5	AM and GM of Regression Coefficients . . . . .	231
8.5.6	Range of Regression Coefficient . . . . .	231
8.5.7	Effect of Change of Origin and Scale on Regression Coefficient . . . . .	231
8.5.8	Angle Between Two Lines of Regression . . . . .	232
8.5.9	Regression with Zero Intercept . . . . .	232
8.6	Identification of the Regression Equations . . . . .	234
8.7	Expectations and Variances of the Regression Parameters . . . . .	235
8.8	Test of Significance for the Regression Coefficient . . . . .	236
8.9	Multiple Linear Regression Analysis . . . . .	236
8.10	Multiple Linear Regression Equation Taking Three Variables . . . . .	237

8.11	Estimation of the Parameters of Linear Regression Model Using OLS Technique in the Matrix Form . . . . .	238
8.12	Estimation of Regression Coefficients from Correlation Coefficients . . . . .	240
8.13	Multiple Correlations . . . . .	244
8.14	The Coefficient of Determination ( $R^2$ ) . . . . .	245
	8.14.1 Interpretation of $R^2$ . . . . .	246
	8.14.2 Adjusted $R^2$ . . . . .	247
8.15	Partial Correlation . . . . .	248
8.16	Some Other Measures of Association . . . . .	250
	8.16.1 Biserial Correlation . . . . .	250
	8.16.2 Tetrachoric Correlation . . . . .	251
	8.16.3 Part Correlation . . . . .	251
8.17	Worked-Out Example Using the Usual Method of Calculation and with the Help of the Software Packages . . . . .	251
	8.17.1 Calculation of All Possible Correlation Coefficients . . . . .	252
	8.17.2 Calculation of Partial Correlation Coefficients . . . . .	259
	8.17.3 Estimation of Simple Linear Regression . . . . .	263
	8.17.4 Estimation of Multiple Linear Regression Equation . . . . .	269
<b>9</b>	<b>Analysis of Variance . . . . .</b>	<b>277</b>
9.1	Introduction . . . . .	277
9.2	Linear Analysis of Variance Model . . . . .	278
9.3	Assumptions in Analysis Variance . . . . .	278
9.4	One-Way Classified Data . . . . .	279
	9.4.1 Analysis of One-Way Classified Data Using MS Excel . . . . .	284
9.5	Two-Way Classified Data . . . . .	286
	9.5.1 Two-Way Classified Data with One Observation per Cell . . . . .	288
	9.5.2 Analysis of Two-Way Classified Data with One Observation per Cell Using MS Excel . . . . .	293
9.6	Two-Way Classified Data with More Than One Observation per Cell . . . . .	296
	9.6.1 Analysis of Two-Way Classified Data with More than One Observation per Cell Using MS Excel . . . . .	301
9.7	Violation of Assumptions in ANOVA . . . . .	304
	9.7.1 Logarithmic Transformation . . . . .	305
	9.7.2 Square Root Transformation . . . . .	307
	9.7.3 Angular Transformation . . . . .	309
9.8	Effect of Change in Origin and Scale on Analysis of Variance . . . . .	311

<b>10</b>	<b>Basic Experimental Designs</b> . . . . .	319
10.1	Introduction . . . . .	319
10.2	Principles of Design . . . . .	322
10.3	Uniformity Trial . . . . .	323
10.4	Optimum Size and Shape of Experimental Units . . . . .	324
10.5	Layout . . . . .	325
10.6	Steps in Designing of Experiments . . . . .	325
10.7	Completely Randomized Design (CRD) . . . . .	326
10.7.1	Randomization and Layout . . . . .	326
10.7.2	Statistical Model and Analysis . . . . .	328
10.7.3	Merits and Demerits of CRD . . . . .	329
10.8	Randomized Block Design/Randomized Complete Block Design (RBD/RCBD) . . . . .	338
10.8.1	Experimental Layout . . . . .	338
10.8.2	Statistical Model and Analysis . . . . .	340
10.8.3	Merits and Demerits of RBD . . . . .	342
10.9	Latin Square Design (LSD) . . . . .	353
10.9.1	Randomization and Layout . . . . .	354
10.9.2	Statistical Model and Analysis . . . . .	354
10.9.3	Merits and Demerits of Latin Square Design . . . . .	356
10.10	Missing Plot Technique . . . . .	358
10.10.1	Missing Plot Technique in CRD . . . . .	359
10.10.2	Missing Plot Technique in RBD . . . . .	359
10.10.3	Missing Plot Technique in LSD . . . . .	361
<b>11</b>	<b>Factorial Experiment</b> . . . . .	365
11.1	Introduction . . . . .	365
11.1.1	Factor and Its Levels . . . . .	366
11.1.2	Type of Factorial Experiment . . . . .	366
11.1.3	Effects and Notations in Factorial Experiment . . . . .	366
11.1.4	Merits of Factorial Experiment . . . . .	367
11.1.5	Demerits of Factorial Experiment . . . . .	367
11.2	Two-Factor Factorial Experiments . . . . .	367
11.2.1	$2^2$ Factorial Experiment . . . . .	367
11.2.2	Two-Factor Asymmetrical ( $m \times n, m \neq n$ ) Factorial Experiment . . . . .	389
11.3	Three-Factor Factorial Experiments . . . . .	398
11.3.1	$2^3$ Factorial Experiment . . . . .	398
11.3.2	$m \times n \times p$ Asymmetrical Factorial Experiment . . . . .	422
11.4	Incomplete Block Design . . . . .	439
11.4.1	Split Plot Design . . . . .	440
11.5	Strip Plot Design . . . . .	452

---

<b>12</b>	<b>Special Experiments and Designs</b> . . . . .	467
12.1	Introduction . . . . .	467
12.2	Comparison of Factorial Effects vs. Single Control Treatment . . . . .	468
12.3	Augmented Designs for the Evaluation of Plant Germplasms . . . . .	472
12.3.1	Augmented Completely Randomized Design . . . . .	472
12.3.2	Augmented Randomized Block Design . . . . .	475
12.4	Combine Experiment . . . . .	482
12.5	Analysis of Experimental Data Measured Over Time . . . . .	500
12.5.1	Observations Taken Over Time in RBD . . . . .	500
12.5.2	Observations Taken Over Time in Two-Factor RBD . . . . .	500
12.5.3	Observations Taken Over Time in Split Plot Design . . . . .	501
12.6	Experiments at Farmers' Field . . . . .	501
12.6.1	Major Considerations During Experimentations at Farmers' Fields . . . . .	502
<b>13</b>	<b>Use-Misuse of Statistical Packages</b> . . . . .	507
	<b>References</b> . . . . .	521
	<b>Index</b> . . . . .	529