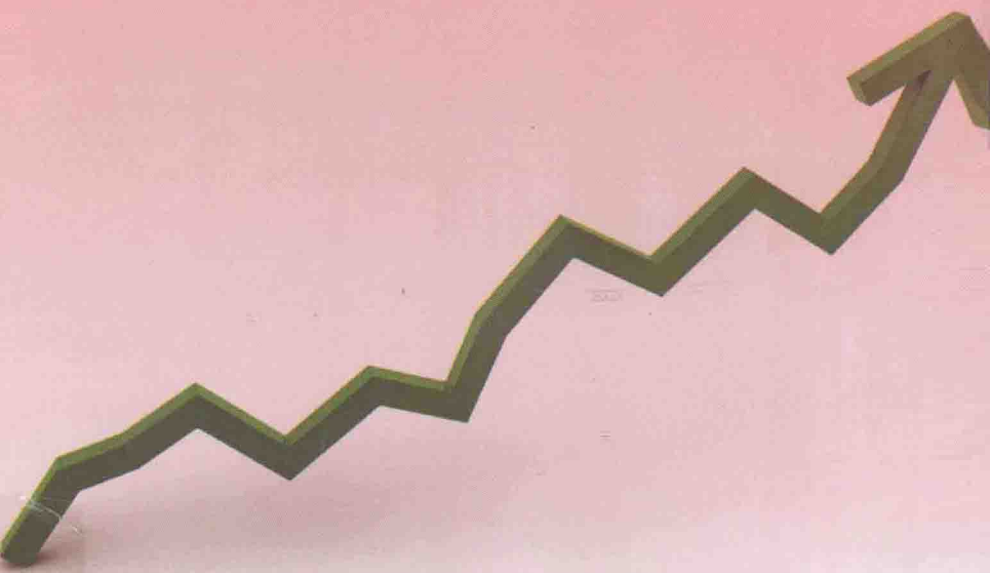
Applied Statistics in Physical Education



Dr. D.K. Maitai

Applied Statistics in Physical Education

Edited & Compiled By
Dr. D.K. Maitai

KSK Publishers
7/26, Basement, Ansari
New Delhi-

Published by:

KSK Publishers & Distributors

7/26, BASEMENT, Ansari Road, Darya Ganj, New Delhi-110002.

Ph.: (O) 011-65190343, 43551324, 42564726
(M) 9811088729, (Fax) 42564726, (Res.) 47091605

E-mail: *khelsahitya1@rediffmail.com*

Web: *www.khelsahitya.trade-india.com*

I.S.B.N. - 978-81-89983-40-6

PRINTED IN INDIA 2011

Laser Typeset by:
JAIN MEDIA GRAPHICS
Delhi-110035.

Printed by:
NEW GYAN OFFSET PRINTERS
Shahjada Bagh, Delhi-110035.

Price: Rs. 600/-

Applied Statistics in Physical Education

Preface

Statistics is the science of the collection, organization, and interpretation of data. It deals with all aspects of this, including the planning of data collection in terms of the design of surveys and experiments. Statistics is closely related to probability theory, with which it is often grouped. Applied Statistics is a field of mathematical statistical implications and practical uses of measurement and evaluation.

Applied Statistics in Physical Education is specially designed for social sciences research, education and physical education students. This book covers all -important concepts of research methodology and applied statistics. The sequence of chapters is in logical order and follow the syllabus of various universities. The distinctive feature of this book is that each chapter is exactly touching each and every concept of Statestical methods.

We hope this book will be very much helpful to research scholars as well as P.G. students though it is impossible to cover the entire study material in this short book. Any type of suggestions are always welcomed.

Author

CONTENTS

1

UNDERSTANDING STATISTICS

Statistics is the science of the collection, organization, and interpretation of data. It deals with all aspects of this, including the planning of data collection in terms of the design of surveys and experiments. Statistics is closely related to probability theory, with which it is often grouped. Statistics is considered by some to be a mathematical science pertaining to the collection, analysis, interpretation or explanation, and presentation of data, while others consider it a branch of mathematics concerned with collecting and interpreting data. Because of its empirical roots and its focus on applications, statistics is usually considered to be a distinct mathematical science rather than a branch of mathematics.

The word statistics has been derived from the Latin word. "Status" the Italian word "Statist" or German word Statistik. All these words mean a 'Political State'. It is also called as "Science of Statecraft" or "Political Arithmetic". Statistics can be understood in either of two sense *i.e.* singular sense or in plural sense. In plural sense, it means collection of numerical facts such as data relating to production of rice in different years, data relating to income, expenditure, prices *etc.* Percentages, averages, correlation, variation,

facts based upon time are all included in Statistics. In singular sense, statistics means the various techniques or methods which are adopted to collect, analyse and interpret the figure. This means that statistics in singular sense refers to statistical methods. Thus statistics means either the data or the numerical methods. In the words of *Croxton and Cowden,* "Statistics may be defined as the collection, presentation, analysis and interpretation of numerical data."

In the words of *Lovitt,* "Statistics is the science which deals with the collection, classification and tabulation of numerical data as a basis for the explanation, description and comparison of Phenomena".

Statistical Tools and Stages of Statistical Study

Several types of statistical/data presentation tools exist, including: (a) charts displaying frequencies (bar, pie, and Pareto charts, (b) charts displaying trends (run and control charts), (c) charts displaying distributions (histograms), and (d) charts displaying associations (scatter diagrams).Different types of data require different kinds of statistical tools. There are two types of data. *Attribute data* are countable data or data that can be put into categories: *e.g.*, the number of people willing to pay, the number of complaints, percentage who want blue/percentage who want red/percentage who want yellow. *Variable data* are measurement data, based on some continuous scale: *e.g.*, length, time, cost.

To Show	Use	Data Needed
Frequency of occurrence:	Bar chart	Tallies by category (data can be
Simple percentage or	Pie chart	attrubute data or variable data
comparisions of magnitude	Pareto chart	divided into categories)
Trends over time	Line graph	Measurements taken in
	Run chart	chronological order (attribute
	Control chart	or variable data cane be used)
Distribution: Variation not related to time (distributions)	Histograms	Forty or more measurements (not necessarily in chronological order, variable data)

Association: Looking for a correlation between two things	Scatter diagram	Forty or more paired measurements (measures of both things of interest, variable data)

Every stage of the statistical study involves the use of certain standard techniques or methods. These methods are called statistical tools. Thus, there are statistical tools used for the collection of data, like the 'Sample' and 'Census' techniques.

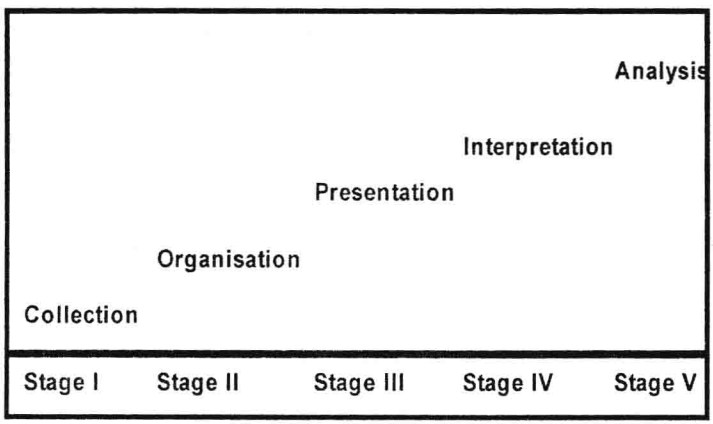


Fig. Stages of Statistical Study

Stages of Statistical Study and the Related Statistical Tools

1. **Identifying the question**
   - What is the question? (What are my hypotheses?)
   - Is it possible to answer the question with statistics?
   - Is the data obtainable? (birth weight, socio economic, drugs, alcohol)
   - Is it ethical to obtain such data?
   - If not, is there a reasonable substitute?
   - Are the assumptions reasonable?

2. Designing a Study

- Identify the population of interest
- Survey
- Obtain a representative sample of that population
- Simple Random Sampling
- Stratified Sampling (M-F, Age groups)
- Systematic Sampling (class roster, census list)
- Multi-Stage Sampling
- Sources of Bias
- Voluntary Response
- Non-response bias (day phone)
- Response bias (people lie)
- Undercoverage
- Observational Studies
- Used when a designed experiment is not ethical
- Subjects studied over a period of time in natural setting
- Case/Control ? Control must match
- Record Variables of interest
- Confounding is a major issue
- Designing an Experiment
- Researcher has control over the subjects or units in the study
- An intervention takes place that otherwise would not occur
- Randomization used to assign treatments
- Strongest case for causality
- EDA ? Exploratory Data Analysis (trends, relationships, differences)
- Pilot Study

3. **Collecting Data**
   - Identify variables
   - Identify types of variables
   - Qualitative
   - Quantitative
   - Identify Limits of measurement or observation
4. **Analyze the data**
   - Use proper procedures and techniques.
   - Check the assumptions behind the procedures and techniques.
5. **Make Conclusions and Discuss Limitations**
   - What are the answers to the original hypotheses?
   - What are the limitations of the study?
   - What conclusions does the study not make?
   - What new questions arise from this study?

Functions of Statistics

Due to increased functions performed by statistics scope of statistics has widened. Statistics make use of methods and theory to numerical data to derive out rational decisions in case of uncertainty. The broad functions performed by statistics are as under:

(i) *Statistics Presents the Complex Data in a Simple form*: So it becomes easy to comprehend. For example data relating to changes in prices of certain commodities between 1961 and 2004 may be so voluminous (large) that it becomes difficult to understand. But if the same data is presented in the form of percentage changes overtime, these data become simple and informative.

(ii) *Statistics Express the Facts in Numbers*: Qualitative form of presentation are not convincing, unless they are shown quantitatively. So, with the help of figures we can represent the data in true form. If it is said that prices are increasing, it is not so precise. But if the same thing is said or

expressed in percentage. Like inflation rate is 4% p.a in year 2002 and which increased to 5% in year 2005, it becomes easy to understand and interpret with certainty.

(iii) *Comparison of Facts*: Another function of statistical is to compare the data relating to facts. When it is said that per capita income of any particular individuals is Rs. 245. It is very difficult to say that whether it is low or high, but when this particular income is compared with other individuals income. We can say that it is low or high. So with the help of comparison, we can draw more meaningful conclusions.

(iv) *Facilitates Policy Formulation*: Statistical methods help in facilitating policy formulation. Nature of each problem can be as certained from the analysis.

(v) *Future Forecasting*: Statistics help in forecasting future. Future behaviour depends more or less on the performance of the past or according to the performance of the present situation. These performances are collected, organised and analysed properly and interpreted, so that proper forecasting can be done.

(vi) *Realisation of Magnitude of Problem*: Statistic helps is realizing the direction of problem of the country. Pries are rising rapidly. One cannot fully realize the gravity of the situation, but by the help of statistics it is quantitatively know that it is increasing at such a speed.

Subject Matter of Statistics

Subject matter of statistics includes two components. Descriptive Statistics and Inferential Statistics.

1. Descriptive Statistics
2. Inferential Statistics

(1) *Descriptive Statistics:* By descriptive statistics, we mean those methods which are used to collect data, to present them in the form of tables, diagrams *etc.* and to explain their characteristics.

These methods include measurement of central tendencies such as average mean, median and mode, measurement of dispersion *etc.* For example, if the number of students of class B.A. (III) is 100 and average marks obtained by them is calculated then we are making use of descriptive statistics.

(2) *Inferential Statistics:* By inferential statistics, we mean those methods by which, on the basis of a given sample, authentic conclusions are drawn relating to the universe or population. For example if the teacher of the class decides to find out the average marks secured by all the students on the basis of sample marks, he would be making use of inferential statistics.

Another important types of statistics are:

(i) Predictive Statistics
(ii) Comparative Statistics
(iii) Relationship Statistics

Data

Measurement results in data. Data means "expression of facts", quantitatively or quatitatively. In statistics, there are two types of data (i) dichotomous and (ii) Continuous

(i) ***Dichotomous or Discrete Data*:** The term dichotomous means 'cut into two parts'. Two fold classification can be athletic--non athletic, dropouts and stay ins, pass-fail, success-failure, married--single *etc.* Such data cannot be subjected to further division. A discrete variable is a variable which can take on numerical values that are specific distinct points on a scale: In discrete the gap is real and thus unbridgeable. No number in fraction. Sex of individual is either male or female, there is nothing in between. This means that fractional divisibility is neither acceptable nor possible.

(ii) ***Continuous Data*:** The continuous data can take any value between two points, on a scale. For example, weight,

height, time, body-fat all can be measured in fraction *i.e.* these variable can be measured in continuous form.

Social sciences, generally present the data in the discrete form, because of certain problem in measuring variables on continuous scales.

Classification of Data

The basic purpose of research is to collect, classify and verify facts so that realistic theories and principles could be established. Data collected from tests and other types of measuring tools are raw and does not have any meaning to the researcher, till they are properly arranged or classified into some systematic order. The method for classifying data into frequency distribution can be illustrated by finding the following marks of eighth grade students in a test of arithmetic.

Marks of the Students in a Test of Arithmetic

3,	4,	6,	8,	5,	7,	4,	3,	5,	3
4,	9,	5,	7,	6,	3,	1,	4,	5,	6,
2,	5,	8,	6,	10,	9,	3,	2,	1,	6,
7,	5,	4,	5,	0,	2,	5,	7,	6,	4
0,	4,	5,	9,	10,	7,	2,	3,	3,	1

By looking at the above marks nothing can be said. They are simply a raw data. By inspecting marks we come to notice that many data is repeated more than once. We see that there are 2 tens five 7's, three 9's and so on. Now we will organise the data in the table given below: Steps to be followed to organise the data in a form of frequency distribution:-

(1) Arrange the data in ascending or descending order.

(2) Tally the number of the students with marks obtained by him.

(3) To record five tallies, we mark four (IIII) and the fifth is crossed (IIII)

(4) Total the tallies in the next collumn.

TABLE: THE FREQUENCY DISTRIBUTION OF THE MARKS ON A TEST OF ARITHMETIC

Marks	Tallies	Frequency f
0	II	2
1	III	3
2					4
3	||| II	7
4	||| II	7
5	||| IIII	9
6	|||| I	6
7	IIII	5
8	II	2
9	III	3
10	II	2
		N = 50

(i) Frequency means: Number of times, a particular marks is repeated and arrangement of data in this form is known as frequency distribution.

(ii) Value of the marks is represented by symbol 'X' and frequency by the symbol 'f'

Arbitrarily Defined Classes of the Variable

When we have large number of classes. It becomes difficult to handle them, so, in such a cases we reduce the number of classes by arranging the data in arbitrarily defined classes of the variable. Frequency distribution consists of class intervals, each of them of same *size.* The number and size of each interval is determined by how best the presumption underlying group is met. Assumptions underlying group are:

(i) Frequencies of each interval is distributed uniformly about the mid points of each interval.

(ii) Arithmetic average of the frequencies is equal to the mid point.

Steps to be followed for construction of a frequency distribution in case of large data:

(i) Class intervals are arranged serially from the smallest number at the starting and highest number at the end. Each class intervals can covers any scores but they should be same size say 5, 10, 15 *etc.*

(ii) For each marks, tally is marked against the corresponding class intervals in which it lies.

(iii) Total number of tallies of each class in the form of frequency is written in the next column.

(iv) Total of 'f' will give total number of marks and it is denoted by 'N'.

Example: Following are the marks obtained by 40 students, present them in continuous series or group frequency table. Marks are as such:

89	58	80	67	70
79	85	80	67	70
59	85	87	75	70
89	58	87	73	60
90	82	87	72	60
88	82	77	76	57
78	82	80	76	82
78	80	80	70	90

Solution:– Marks grouped into a frequency table:

Class Interval	Tallies	Frequency
56-60	IIII I	6
61-65	–	–
66-70	IIII I	6
71-75	III	3
76-80	IIII IIII I	11
81-85	IIII I	6
86-90	IIII III	8
91-95	–	–
96-100	–	–
		N = 50

Cumulative Frequency Series

Cumulative frequency series is that series in which the frequencies are continuously added corresponding to each class interval in the series. If we denote individual class frequencies of the successive class intervals by f_1, f_2, f_3, ... fn, the cumulative frequencies will be f_1, $f_1 + f_2$, $f_1 + f_2 + f_3$ and so on.

Illustration

CUMULATIVE FREQUENCY SERIES

Marks	Frequency	Cumulative Frequency
5-10	3	3
10-15	8	3+8=11
15-20	9	11+9=20
20-25	4	20+4=24
25-30	4	24+4=28

Two ways of presenting cumulative frequency series are: There are two ways of presenting cumulative frequency series as under:-

(i) Cumulative frequencies may be expressed on the basis of upper limits of the class intervals, *e.g.* less than 15, less than 20, less than 25, when the class intervals are 5-10,10-15, 15-20, 20-25.

(ii) Cumulative frequencies may be expressed on the basis of lower class limits of the class intervals, *e.g.* more than 5, more than 10, more than 15, when the class intervals are 5-10, 10-15 and 15-20.

Following example, shows how a frequency distribution is converted into a cumulative frequency distribution.

TABLE: CUMULATIV FREQUENCY SERIES

Method I		Method II	
Marks	No. of students	Marks	No. of Students
less than 10	0+3=3	more than 5	= 24
less than 15	3+8=11	more than 10	24-3 = 21