ANIMAL GENETICS & BREEDING

BIO-STATISTICS AND COMPUTER APPLICATION

Theory

INTRODUCTION AND IMPORTANCE - MEANING OF STATISTICS AND THEIR FUNCTION

Dr Anil Meel
Department of Animal Genetics & Breeding
MJF Veterinary college

STATISTICS

- The word 'Statistics' has come from the Latin word 'status' ,the Italian word 'statista' or the German word 'statistik' , the French word 'statistique', each of which means political state.
- In early days, facts and figures about the financial resources, births and deaths, army strength and income were collected for the purpose of efficient administration which was called statistics i.e, anything pertaining to the state.

- Now a days Statistics is not only the science of state but it plays an important role in all walks of life and in all branch of scientific enquiry. In fact, statistics has become one of the essential tools in modern biology.

- Usually, the word 'statistics' carries different meanings depending on the occasion in which it is used.
  - For e.g., it may mean statistical data which refers to quantitative information, statistical method which means the methods dealing with quantitative information or statistical measures of a sample. i.e., Arithmetic mean, standard deviation etc. of a sample.

- By statistical data, we mean the aggregate of facts which are affected by multiplicity of causes, numerically expressed, estimated to a reasonable standard of accuracy and collected in a systematic manner for a pre - determined purpose.
- Statistical method includes collection, classification, tabulation, presentation, analysis and interpretation of data.
- Biostatistics is the application of statistical methods to the problems of biology including human biology, medicine and public health.
- Biostatistics is also called Biometry meaning "biological measurement".

DEFINATION

- Statistics is a branch of Applied Mathematics and is concerned with observational data.
- Statistics may be defined as the science and art of collection, organization, presentation, analysis and interpretation of numerical data. OR Statistics is concerned with scientific methods for collecting, organizing, summarizing, presenting and analyzing the data as well as drawing valid conclusions and making reasonable decisions on the basis of such analysis.

HISTORY

- Father of Biostatistics – Francis Galton (1822-1911) who was cousin of Charles Darwin.
- Before 19th Century statistics was mathematical, political and governmental.
- In early 19th Century - Belgian Statistician Quetelet combined theory and practical methods of statistics and applied to the problems of biology, education, medicine and sociology.

- Francis Galton – Introduction and use of statistics in biology. The inadequacy of Darwin's genetic theories stimulated by him to try and solve the problems of heredity. His contribution in heredity, eugenesis, psychology, anthropometry and statistics is notable. Present day correlation understanding is credited to him.
- Pearson collaborated with Galton in later years and was instrumental in developing many correlation and regression formulas that are in use today.
- The term biometry was coined by W.F.R. Weldon (1860-1906).

- An English statistician R.A. Fischer gave the major contribution in small sample theory. He developed most of his methods in agricultural and biological sciences. Fischer's methods are used in almost all the fields of sciences.

Functions of Statistics

- Presents facts in a definite form
- Simplifies mass of figures
- Facilitates comparison
- Helps in formulating hypothesis
- Helps in testing the hypothesis
- Helps in prediction
- Helps in the formulation of suitable policies

Limitations of Statistics

- Statistics does not deal with the individuals.
- It deals only with quantitative characters. However qualitative characters can be numerically expressed and analyzed.
- For eg. Intelligence of students by marks obtained, poverty by income received.
- Statistical results are true only on an average.
- Statistics is only one of the methods of studying a problem.
- Statistics may be sometimes misused, if not properly interpreted.

Doubts

- Anthropometry is the science of obtaining systematic measurements of the human body. Anthropometry first developed in the 19th century as a method employed by physical anthropologists for the study of human variation and evolution in both living and extinct populations. In particular, such anthropometric measurements have been used historically as a means to associate racial, cultural, and psychological attributes with physical properties. Specifically, anthropomorphic measurements involve the size (e.g., height, weight, surface area, and volume), structure (e.g., sitting vs. standing height, shoulder and hip width, arm/leg length, and neck circumference), and composition (e.g., percentage of body fat, water content, and lean body mass) of humans.

- To obtain anthropometric measurements, a variety of specialized tools (as depicted below) are used:
- Stadiometers: height
- Anthropometers: length and circumference of body segments
- Biocondylar calipers: bone diameter
- Skinfold calipers: skin thickness and subcutaneous fat
- Scales: weight
- **eugenics:** the selection of desired heritable characteristics in order to improve future generations, typically in reference to humans.

Biostatistics

- Biostatistics deals with the application of statistics to biological data. It involves collection, arrangement, analysis and interpretation of numerical facts so as to draw scientific conclusion or make effective decisions. It is also called Biometry.
- Father of Biostatistics - Francis Galton (1822-1911)
- Biostatistics can be divided in to two subcategories
  I. **Descriptive biostatistics**
  II. **Inferential biostatistics.**

I. **Descriptive biostatistics -** The descriptive biostatistics involves graphical and tabular approaches to describe, summarize and analyze the data. It makes more informative and comprehensible.

II. **Inferential biostatistics -** The procedures which serve to make generalization on the basis of the studies of sample, which is a part of the population of interest, constitute inferential biostatistics. e.g., weight gain in pigs after being feeding new diet.

Methods for collection data

1) Measurement: e.g. Length of pod, Height of plant, Area of leaf, Volume of mango fruit etc.

2) Scale in physical science: e.g. Temperature, Humidity, Wind velocity etc.

3) Score or ranks: e.g. Intelligence test by judging candidates interview.

4) Personal contact: e.g. By asking questions to the individual.

5) Questionnaire: e.g. Data collected by mailing a form called questionnaire consisting several questions.

Population

- A population is any actual or conceptual collection of individuals defined by stated characteristics. Population is a group of individuals about which the scientific inferences are made.

or

- A set or collection of objects pertaining to a phenomenon of statistical enquiry is referred to as universe or population or census.
- e.g. – all university students in Delhi, all progenies of particular rat, all persons who undergo heart surgery in 2007. On occasions biostatistician may deal with population of measurements.

Variable

- A variable is any property of an individual (person, tree, rat etc.) that can be expressed in numerical terms. The characteristics which show variation or variability are called variables or variates.

Or

- The quantitative or numerical characteristic of the data is called as a variable.
- e.g. – weight of an animal, Cabbage yield, wheat yield per hectares of the growers, height of 12 year old boy, % of calcium in blood of patient.

- **The variable can of two types**
- **(i) Qualitative**: - The characteristics which can not be measured numerically or in terms of magnitude e.g. flower color, nature of surface, eye color in drosophila fly .
- **(ii) Quantitative**: - The characteristics which can be measured in terms of magnitude e.g. yield of crop, height, weight. The quantitative characteristics are of two types.
- **(A) Discrete**: - Character which takes only integer values/or whole value. There is a definite gap between two values. e.g. Number of students in a class, Number of bacteria in given area.
- **(B) Continuous**: - The quantity which can take any numerical value within a certain range. Height, weight (They are in fraction and there is no definite gap between values).

Important Terminology

- **Attributes:** Qualitative characteristics of an individual which shows variability is known as attributes. e.g. Red or White, Live or Dead, Pregnant or non-pregnant.

Or

It refers to the qualitative character of the items chosen. (e.g.) breed of an animal.

- **Constant:** It is a particular type of variable which does not vary from other member of a group to another or within a particular set of defined conditions. e.g. age, sex, religion.

Or

It is a numerical value, which is same for all the units in the population. (e.g.) no. of credit hours for B.V.Sc students.

- **Sample:** A sample is any segment of the population that is examined and measured as an aid to investigating properties of population. A sample is representative of the population hence the selection of sample must be random. e.g. 5 ml milk from each can taken for fat test at dairy.

or

When a few units are selected from a population, it is called as a sample. (e.g.) animals of a particular breed in a farm.

- **Parameter:** A parameter is any numerical property descriptive of population or A statistical measure pertaining to a population is called as a parameter.

(e.g.) mean, standard deviation of the population, median, mode are parameters of location.

- **Accuracy:** Accuracy is the closeness of a measured or computed value to its true value.
- **Precision:** Precision is the closeness of repeated measurements of the same quantity.
- **Statistic:** A statistical measure pertaining to the sample is called as a statistic. (e.g.) mean, standard deviation of the sample.

THANK YOU