

ANIMAL GENETICS & BREEDING

UNIT - I

BIO-STATISTICS AND COMPUTER APPLICATION

Theory

COLLECTION AND CLASSIFICATION OF DATA

Dr Anil Meel

Department of Animal Genetics & Breeding
MJF Veterinary college

COLLECTION OF DATA

- A statistical investigation always begins with collection of data. One can collect the data either by himself or from available records.
 - The data collected by the investigator himself or by his agent from the sample or population are called as the **primary data**.
 - The source from which one gathers primary data is called as the **primary source**.
 - The data collected from the available sources is known as **secondary data**.
 - The source from which we are getting secondary data is known as **secondary source**.

PRIMARY DATA

- The data collected originally or the first hand information of facts.
- Methods of collecting Primary Data
 - *Direct personal observation*: The investigator himself goes to the field of enquiry and collects the data
 - *Indirect personal observation*: The investigator collects data from a third person (called as witness), who knows about the data being gathered.

- *Data collection through agents, local reporters etc:* Here the investigator appoints some person called agents or local reporters on his behalf to collect information.
- *Data collection through questionnaires:* The investigator prepares the needed information for the particular study in the form of questions, called questionnaires and sends the same to the respondents to collect data from the respondents.

Methods	Merits	Demerits
Direct personal observation	<ul style="list-style-type: none"> • It is very accurate. • Intensive details can be collected. 	<ul style="list-style-type: none"> • Expensive in terms of time and money. • Not suitable when the field of enquiry is large.
Indirect personal observation.	<ul style="list-style-type: none"> • It saves time. 	<ul style="list-style-type: none"> • Witness should possess thorough knowledge of the facts regarding the problem of investigation. • Witness must be willing to give information.

<p>Data collection through agents and local reporters etc.</p>	<ul style="list-style-type: none"> • It saves time. • Large area can be covered 	<ul style="list-style-type: none"> • The agents will collect information in their own fashion. • Only approximate results can be obtained. • It is expensive.
<p>Data collection through questionnaires</p>	<ul style="list-style-type: none"> • It saves time. • It is less expensive. • Geographically dispersed area can be covered 	<ul style="list-style-type: none"> • It cannot be used if the informants are illiterate. • Response may be poor. • Possibility of vague/inaccurate answers.

SECONDARY DATA

- The data collected from the available sources like published reports, documents, journals etc. are called secondary data.
- The source from which the secondary data are collected is called as secondary source of data.
- While the primary data are collected for a specific purpose, the secondary data are gathered from sources which were done for some other purpose.

Sources of obtaining secondary data

- Published reports/ documents of institutions, NGOs etc.
- Scientific journals
- Government reports
- Books and news papers

Merits and Demerits of Secondary Data

- Merits

- It saves time, labour and money.

- Demerits

- It may not be very accurate
- All the data needed may not be available
- It might have been collected by some improper methods and in some abnormal condition

Classification of data

- Classification is the process of arranging data into sequences and groups according to their common characteristics or separating them into different but related parts.
- Objectives of Classifications
 - To remove unnecessary details
 - To bring out explicitly the significant features in the data
 - To make comparisons and drawing inferences

Methods of Classification

- **Numerical Classification** - Classification of data according to quantitative characters. (e.g) classification of animals in a farm according to their weight
- **Descriptive Classification** - Classification according to attributes i.e, qualitative characters. (e.g). classification of animals according to breeds
- **Spatial or Geographical Classification** - Classification according to geographical area. (e.g) district-wise livestock population in Tamil Nadu

- ***Temporal or Chronological Classification*** - Classification according to time.
(e.g) livestock population in different years
- ***Classification according to class interval or frequency distribution*** - When the data are grouped into classes of appropriate interval, showing the number in each class, we get frequency distribution. This is called grouped data. The original data is called raw data.
(e.g) The following is the frequency table showing the distribution of chicks in different weight classes.

Weight(in gm.)	No. of Chicks
36-40	12
40-44	25
44-48	17
48-52	05
52-56	06
56-60	10
<i>Total</i>	<i>75</i>

Terms used in Frequency distribution

Class Interval and Class Limits - Data are classified or grouped into regular intervals with the range of values of the data (Class Interval) with the lower and upper limits which are known as Class Limits.

- *True Class Interval* - When the Class Intervals are continuous, it is called True or Inclusive Class Interval.
- *Apparent Class Interval* - When there is a small gap between the upper boundary of any class and lower boundary of successive class, then the Class Interval is called Apparent or Exclusive Class Interval.

- **Class Frequency or Frequency**

Class frequency or frequency is the number of observations in that class.

- **Width or Length of the Class Interval**

Width Or length of the Class Interval is the difference between the upper boundary and lower boundary of the same class.

- **Class Mark**

It is the midpoint of the class.

It is given by half of the sum of the lower limit and upper limit of any class.

TABULATION OF DATA

It is a systematic arrangement of statistical data in columns and rows. It is the next process of condensation of data after classification. Tabulations is a mechanical part of classification. **The objects of tabulation are -**

- Tables are more comprehensive and intelligible and carry a lasting impression on the mind of the reader.
- Tables facilitate quick comparisons.
- Tables facilitate economy of space (while presenting) and time (while reading)
- Relationship and other relevant characteristics of item can be easily marked out in tabulated data.

The following are the points to be considered carefully in preparing a table.

- The title should be short but clear and it should give a full idea of its contents.
- The column and row headings should be self explanatory.
- Footnotes may be given if absolutely necessary.
- Prominence may be given to important facts by different methods of mailing and spacing.
- To have better clarity, space should be left after every five to ten rows.
- If the table is taken from secondary data, it is advisable to give a source note for the table mentioning the source for which the data is collected.

Types of table

- **Reference table or General Table** - These table contain a great deal of summarized information. They appear usually at the end of the report in the form of appendixes.
- **Text or summary tables** - They are used to analyse or assist in the analysis of classified data. They are included in the discussion of the body of the report.
- **Statistical tables** - These are special tables used by statisticians in interpreting the results of statistical analysis. The commonly used tables are 't' tables, z table, F table etc.

What is frequency distribution ?

- A frequency distribution in statistics is a representation that displays the number of observations within a given interval. The representation of a frequency distribution can be graphical or tabular so that it is easier to understand.

What Is the Importance of a Frequency Distribution?

- A frequency distribution is a means to organize a large amount of data. It takes data from a population based on certain characteristics and organizes the data in a way that is comprehensible to an individual that wants to make assumptions about a given population.

Types of frequency distribution

- **Ungrouped frequency distribution:** It shows the frequency of an item in each separate data value rather than groups of data values or The ungrouped frequency distribution is a type of frequency distribution that displays the frequency of each individual data value instead of groups of data values. In this type of frequency distribution, we can directly see how often different values occurred in the table.
- **Grouped frequency distributions:** The number of observations of each class interval of a variable. Class intervals are ordered groupings of a variable's values. You can use this type of frequency distribution for quantitative variables.

- Example distribution of following data in ungrouped and grouped distribution
- 22, 24, 27, 35, 40, 32, 29, 22, 24, 27, 35, 40, 32, 29, 24, 27, 35, 40, 24, 22, 27

Ungrouped

Number	Frequency
22	3
24	4
27	4
29	2
32	2
35	3
40	3

Grouped

Category	Frequency
21 - 25	7
26 - 30	6
31 - 35	5
36 - 40	3
Total	21

GROUPED DATA VERSUS UNGROUPED DATA

Differences	Grouped	Ungrouped
Classification	Organized into classes	No form of organisation
Preference	Preferred when analyzing data	Preferred when collecting data
Accuracy	Has higher accuracy levels when calculating mean and median	Less accurate in determining mean and median
Presentation	Frequency tables are mostly used	Lists are used in this data type
Summary	Summarized in frequency distribution	No form of summarization

- Relative frequency distribution: A relative frequency distribution shows the proportion of the total number of observations associated with each value or class of values and is related to a probability distribution, which is extensively used in statistics.
- Cumulative frequency distribution: The cumulative frequency is calculated by adding each frequency from a frequency distribution table to the sum of its predecessors. The last value will always be equal to the total for all observations, since all frequencies will already have been added to the previous total.

Example of relative frequency distribution

Number of Pets	Frequency	Relative Frequency
1	150	37.5%
2	90	22.5%
3	110	27.5%
4	30	7.5%
5	20	5.0%

← $150/400 = 37.5\%$

← $90/400 = 22.5\%$

← $110/400 = 27.5\%$

← $30/400 = 7.5\%$

← $20/400 = 5.0\%$

Example of relative frequency distribution

School Grade	Frequency of Students	Cumulative Frequency
1	23	23
2	20	$23 + 20 = 43$
3	15	$43 + 15 = 58$
4	12	$58 + 12 = 70$
5	10	$70 + 10 = 80$
6	8	$80 + 8 = 88$

THANK YOU