Biostatistics Practical

Introduction

Statistics, as we know, is the branch of applied mathematics which deals with the collection of numerical observations and their expression in logical terms or interpretation of the results. The use of statistics in biology is known as Biostatistics or Biometry.

Therefore, we can define Biostatistics or Biometry as the branch of statistics which deals with the study of biological problems or the application of mathematical methods to the study of living things or organisms. The purpose of statistics is not only to collect numerical data, but is to provide a methodology for handling, analyzing and drawing valid inference for the data. It has wide application in almost all the branches of science-social as well as physical such as biology, psychology, education, economics, planning, business management, mathematics and so on.

The science of Biostatistics was founded by Francis Galton and later developed by Karl Pearson, RA Fisher and others. At present, it is an indispensable tool in all the branches of biology.

In biology, we generally use two types of statistics. These are sample (descriptive) statistics and test statistics.

1. **Sample Statistics**
   These are generated from data used to estimate population parameters (e.g. mean, standard deviation etc). Sample statistics are used to define the nature and distribution of the data. It often provides knowledge as to whether the results are statistically significant or not.
2. **Test Statistics**
   Test statistics are used to test the hypothesis about one or more samples of data. The statistical test, one chooses for analysis, is dictated to a great extent by his/her experimental design, and the type of analysis should be considered when designing the data collection formats. Neither the statistical test nor the experimental design should entirely dictate the others, but they should be coordinated (e.g. *Chi*-square, *t*-Test etc).

Estimation of mean, median and mode

Mean

Mean or arithmetic mean is the simple average of the different values of a variable. In case of a data series, mean is the sum of all the observations divided by number of observations. Thus, if variable has "n" different values, the mean ($\bar{X}$) is given by-

$$\text{Mean } (\bar{X}) = \sum \frac{f1x+f2x+\cdots+fnx}{f1+f2+\cdots fn}$$

$$= \sum \frac{fx}{N}$$

Where,

∑= summation

f= frequency

x= size of population

N= number of observations

Median

The median refers to the middle value in the given set of observations. In a series of observations arranged in ascending or descending order the middle observation is called median.

To calculate the median of a continuous data we use the formula as given below.

$$Median\ (m) = l_0 + \frac{\frac{N}{2} - fc}{f} \text{x}\ h$$

Where,

$l_0 = lower\ limit\ of\ class\ interval\ in\ which\ median\ value\ lies$

$fc = cumulitative\ frequency\ just\ before\ the\ class\ interval\ in\ which\ median\ value\ lies$

$N = total\ number\ of\ observations$

$f = frequency\ of\ class\ interval\ in\ which\ median\ value\ lies$

$h = length\ of\ class\ intervals$

Mode

The mode is defined as the value which has the maximum frequency of occurrence or which occurs more frequently. In other words, mode is the value of variable which is predominant in the data series.

The value of mode is calculated by the formula as given below.

$$Mode = L + \frac{f1 - f2}{f1 - f0 - f2} \text{x}\ h$$

Where,

$L = lower\ limit\ of\ class\ interval\ in\ which\ mode\ value\ lies$

$f0 = frequency\ of\ the\ class\ interval\ preceding\ the\ median\ class$

$f1 = frequency\ of\ modal\ class$

$f2 = frequency\ of\ the\ class\ interval\ succeding\ the\ modal\ class$

$h = length\ of\ class\ intervals$

Mode can also be calculated by the following formula if the mean and median have already been worked out.

$$Mode = 3\ median - 2\ mean$$

Steps for calculating mode of a continuous series:

1. Locate the modal class
2. Find the value of mode by the formulae given above

Experiment No: 01

Aim: To calculate mean, median and mode from the given data.

Problem: The following frequency table gives the height of 100 plants in centimeter. Find mean, median and mode.

Class intervals (CI)	Frequency (f)
00-10	3
10-20	3
20-30	3
30-40	7
40-50	6
50-60	5
60-70	9
70-80	4
80-90	5
90-100	5

Solution: From the given data following frequency table is prepared which can satisfy the requirement of the formula.

Class intervals (CI)	Frequency (f)	Mid value of CI (x)	Cumulative frequency (fc)	f.x
00-10	3	5	3	15
10-20	3	15	6	45
20-30	3	25	9	75
30-40	7	35	16	245
40-50	6	45	**22**	270
50-60	**5**	**55 (median)**	27	275
60-70	9	65	36	585
70-80	4	75	40	300
80-90	5	85	45	425
90-100	5	95	50	475
	$\sum f = \mathbf{50}$	$\sum X = \mathbf{500}$		$\sum fx = \mathbf{2710}$

Now, we know,

$$\text{Mean } (X^{-}) = \sum \frac{fx}{N}$$

$$= \frac{2710}{50}$$

$$= 54.2$$

Hence, mean height of 100 plants is 54.2 cm

Therefore,

Mode = 3(median)-2 (mean)
= 3 x 56 – 2 x 54.2
= 168-108.4
= 59.6

So, mode value in height of 100 plants is 59.6 cm.

Again, we have,

$$Median\ (m) = l_0 + \frac{\frac{N}{2} - fc}{f} \text{x } h$$

$$= 50 + \frac{\frac{50}{2} - 22}{5} \text{x } 10$$

$$= 50 + \frac{25 - 22}{5} \text{x } 10$$

$$= 50 + \frac{3}{5} \text{x } 10$$

$$= 50 + 6$$

$$= 56$$

Hence, median height of 100 plants is 56 cm