Statistical Analysis of Measurement Data

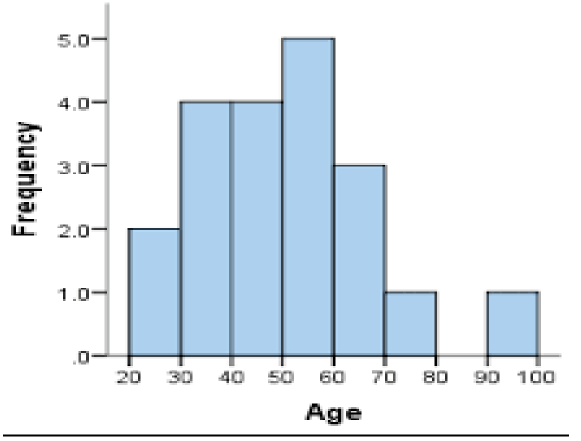
3.1 Introduction

After the gross and systematic errors are removed or minimized, there remain random error in the final result of measurement. The outcome of measurement with random error can be predicted by statistical analysis. The statistical analysis of experimental data is obtained by two forms of tests:

1. Single sample test.
2. Multi-sample test.

In single-sample test, the measurement is done under identical conditions but at different time. Multi-sample test involves repeated measurements of a given quantity using different test conditions, such as different instrument, different ways of measurement and different observers.

The collection of measured data is called the sample data. Many data may be repeated a number of times. The number of repetition of a datum is called its frequency. The sample data may be represented by a graph known as a histogram as shown in figure below.



3.2 Statistical Descriptors

The statistical analysis is done by calculating some numbers called statistical 04descriptors. Some of these descriptors are discussed here:

1. Arithmetic mean

The arithmetic mean of a number of readings gives the most probable value of the measured quantity. The arithmetic mean of number of readings is calculated as follows:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Or

$$\bar{X} = \frac{x_1 + x_2 + \dots\dots\dots\dots + x_n}{n}$$

Where,

$\bar{X}$ = is the mean value.

$x_1, x_2, \dots\dots\dots\dots, x_n$= are the readings of the measured quantity, and

n= is the number of readings.

If the sample data has a frequency distribution, the arithmetic mean will calculated as follows:

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i}$$

or

$$\bar{X} = \frac{x_1 f_1 + x_2 f_2 + \dots \dots \dots \dots + x_n f_n}{f_1 + f_2 + \dots \dots \dots + f_n}$$

Where,

$f_1, f_2, \dots \dots \dots \dots, f_n$= are the frequencies of the measured quantity, and

n= the number of groups of repeated data.

2. The Median

When the number of readings in the data set is very large, the calculation of the mean value becomes tedious and it is more convenient to use the median value. This being a close approximation to the mean value.

- For odd number of readings, the median is the middle value when the readings are arranged in ascending or descending order.
- For even number of readings, the median is the average of the two middle values when the readings are arranged in ascending or descending order.

3. The Mode

The mode of a set of readings, is the value which occurs with great frequency. The mode may not be unique or may not exist.

4. The average deviation

The deviation of reading from the mean value is a measure of error in measurement. The deviation from the mean value can be express as:

$$d_i = x_i - \bar{X}$$

Where, d is the deviation of the i^{th} value of readings from the mean. It should be noted that the deviation from the mean may be positive or negative and the algebraic sum of these deviations is always equal to zero.

The average deviation is defined as the mean of the absolute values of the deviations of readings and it is indicate the precision of instrument. A highly precise instrument gives a low average deviation of readings. The average deviation may be expressed as:

$$D = \frac{\sum_{i=1}^{n}|d_i|}{n}$$

Or

$$D = \frac{|d_1| + |d_2| + \dots\dots\dots\dots\dots + |d_n|}{n}$$

Where, D = is average deviation.

6. The Standard deviation

The root mean square of the deviation from the mean value is known as the standard deviation, and is a very useful aid in the analysis of random error in the measurement.

The standard deviation is expressed as:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}}$$

Where,

σ = is the standard deviation, (reading sigma).

6. The Variance

The deviation of readings can alternatively be expressed by the variance, which is the square of the standard deviation, i.e.

$$V = \sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}$$

Where,

V = is the variance.

The variance essentially gives the same information as can be obtained from the standard deviation, but the standard deviation however has the advantage of being of the same unit as variable.

3.3 Error Expression

There are two formulas used to express the errors of measurements:

1. Absolute Error

The absolute error is the difference between the true value and the measured value, i.e.

$$E_a = |x_T - x_m|$$

Where,

E_a = is the absolute error.
x_T = is the true value
x_m = is the measured value

2. Relative Error

Is the error when the measured value is related to the true value of the quantity under measurement and is given as a percentage value and expressed as:

$$E_r = \left(\frac{|x_T - x_m|}{x_T}\right) \times 100\ \%$$

Where,

E_r = is the relative error.

Examples – L3

Ex3.1 Six readings were taken to measure a displacement of 500 cm as follows: 500.8, 500.9, 500.6, 500.9, 500.7 and 500.9.

A. Calculate :
(1) The arithmetic mean (2) The median (3) The mode (4) The average deviation (5) The standard deviation (6) The variance.

B. For this measuring instrument and according to the fifth reading, determine:
(1) The absolute error (2) The relative error

Sol. (A)

(1) The arithmetic mean is given as:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Or

$$\bar{X} = \frac{x_1 + x_2 + \ \dots\dots\dots\dots + \ x_n}{n}$$

Where n = 6.
Therefore,

$$\bar{X} = \frac{500.8 + 500.9 + 500.6 + 500.9 + 500.7 + 500.9}{6}$$

Or

$$\bar{X} = 500.8 \ \ cm$$

Also, the arithmetic mean may be calculated as follows:

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i}$$

Or

$$\bar{X} = \frac{x_1 f_1 + x_2 f_2 + \ \dots\dots\dots\dots + \ x_n f_n}{f_1 + f_2 + \ \dots\dots\dots + f_n}$$

Therefore,

$$\bar{X} = \frac{500.8 \times 1 + 500.9 \times 3 + 500.6 \times 1 + 500.7 \times 1}{1 + 3 + 1 + 1}$$

Thus,

$$\bar{X} = 500.8 \ \ cm$$

(2) The median

By ascending arrangement of readings we get:

500.6, 500.7, 500.8, 500.9, 500.9, 500.9.

Since the number of readings is even, therefore

$$The\ Median = \frac{500.8 + 500.9}{2} = 500.85\ cm$$

(3) The Mode

The Mode = 500.9 cm

(4) The Average deviation

$$d_i = x_i - \bar{X}$$

Thus,

d1 = 500.6 – 500.8 = - 0.2
d2 = 500.7 – 500.8 = - 0.1
d3 = 500.8 – 500.8 = 0
d4 = 500.9 – 500.8 = 0.1
d5 = 500.9 – 500.8 = 0.1
d6 = 500.9 – 500.8 = 0.1

The average deviation is given as:

$$D = \frac{|d_1| + |d_2| + \dots\dots\dots\dots\dots + |d_n|}{n}$$

Where, n = 6.

Therefore,

$$D = \frac{|-0.2| + |-0.1| + |0| + |0.1| + |0.1| + |0.1|}{6} = \frac{0.6}{6} = 0.1\ cm$$

(5) The Standard deviation

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{n-1}}$$

Or,

$$\sigma = \sqrt{\frac{0.2^2 + 0.1^2 + 0^2 + 0.1^2 + 0.1^2 + 0.1^2}{6-1}}$$

Or,

$$\sigma = \sqrt{\frac{0.08}{5}} = 0.126\ cm$$

(6) The Variance

$$V = \sigma^2 = 0.016\ cm^2$$

(B)

(1) The absolute error

$$E_a = |x_T - x_m|$$

Therefore,

$$E_a = |500 - 500.7| = 0.7\ cm$$

(2) The relative error

$$E_r = \left(\frac{|x_T - x_m|}{x_T}\right) \times 100\ \%$$

Or,

$$E_r = \left(\frac{|500 - 500.7|}{500}\right) \times 100\ \%$$

Thus,

$$E_r = 0.14\ \%$$