Unit 7 APPLICATIONS IN PHYSICS

Structure

7.1 INTRODUCTION

Physics, as you know, is an experimental science. Any hypothesis or theory proposed in physics (as also in other sciences) has to stand the test of observation and experiment. The raw data obtained from observations and experiments has to be analysed and interpreted systematically. It is in this area that the concepts of probability theory and statistics which we developed in Units 5 and 6 find prolific applications.

In this unit you will learn about **correlation analysis** and **regression analysis.** These are the most notable tools used in the interpretation of statistical data, particularly when we have to investigate relationships between two or more variables. Using correlation analysis we can establish the strength of linear association between two variables. Regression analysis allows us to determine the quantitative relationship between variables. The method of least squares is a very useful technique for establishing such relationships. You will also learn about it in this unit.

As a student of physics you have performed many experiments. And you would know that in every measurement an element of error is involved. In the last section of the unit we have discussed the statistical theory of errors. You will learn to estimate errors in the measured value of a given physical quantity.

Objectives

After studying this unit you should be able to

- compute the correlation coefficient for given data
- use the method of least squares to fit the regression line for given data
- estimate the standard error of the mean for random errors and compute the best value of directly and indirectly measured quantities,

7.2 CORRELATION ANALYSIS

Can you recall some experiments from your school science courses? What was the objective of performing these experiments? You were essentially investigating some phenomena or verifying certain laws by establishing relationships between the variables involved. At some

stage in your curriculum you must have performed experiments with a simple pendulum. This experiment has also been included in our first level physics laboratory course PHE-03 (L). If you have taken this course, you would recall that in doing the experiment you are required to determine whether or not the time period of the pendulum depends on its length, mass of the bob and amplitude of the swing. And if it does, what is the exact quantitative relationship between the associated variables?

This situation can be generalised. In physics, and in nearly all other scientific studies, the phenomena to be investigated usually involve many variables. And the basic aim of an investigation is to find the quantitative relationship between these variables. For the simple pendulum, it is also possible to obtain a theoretical expression for its time period, in terms of its length and acceleration due to gravity. (You can apply Newton's laws of motion to the system and obtain it.) But at advanced levels, particularly in exploratory research, often we do not have a ready theory specifying the relationship between variables. In such a situation we would first like to know whether or not there is any association between the variables and then determine the exact quantitative relationship between them.

As an example, let us consider a problem from the ongoing research on disturbances in solar atmosphere. Observations of a series'of photographs of the sun's surface show that the pattern of brightness at the sun's surface changes drastically every 5 minutes or so. For measuring the brightness a line is specified on the sun's surface. The intensity of light is measured at certain points along this line on each photograph of the sun's surface. Table 7.1 gives these measurements at five points along one such line drawn on the photograph of the sun's surface. The first row of numbers gives the initial brightnesses. These are measured in a standardised manner so that the brightness at a point is measured as a departure from the mean. This standardising gives these numbers the property that their average is 0, and the average of their squares is 1. (In this table the numbers have been rounded off and the averages do not work out exactly.) The second and third rows give the brightnesses a few seconds and a few minutes later.

Table 7.1: Brlghtness measurements at various points on the sun's surface at various times

Time	Points				
	1	2	3	4	5
t_1	1.5	0.5	0	-0.5	-1.5
t_2	1.4	0.7	0	-0.7	-1.4
t_3	1.0	-1.4	0.4	-0.1	1.0

The question now is to find out whether the brightnesses measured at various points of time are related. How do we do it?

It is here that statistics comes to our aid. It provides us with the useful tool of **correlation analysis** to find out whether a relationship exists or not between a given set of variables. We can know the extent of the relationship between the variables by determining the correlation coefficient for the given data. In Unit 5 you have seen that the correlation coefficient is given by

$$r_{XY} = \frac{n\Sigma(xy) - \Sigma(x)\Sigma(y)}{\sqrt{n\Sigma(x^2) - (\Sigma x)^2}\sqrt{n\Sigma(y^2) - (\Sigma y)^2}} \qquad (7.1)$$

where each pair of values occurs only once.

As you know, the correlation coefficient is a measure of the degree of linear association between two variables. As given by Eq. (7.1), r_{XY} always has a value between +1 and -1. A value of r_{XY} close to +1 or -1 indicates high correlation. Correlation is considered low when r_{XY} is close to zero.

Let us now determine the correlation coefficient for the first and second set of brightnesses given in Table 7.1.

Since $\frac{1}{n}\sum x$ and $\frac{1}{n}\sum y$ are zero, $\frac{1}{n}\sum x^2$ and $\frac{1}{n}\sum y^2$ are 1 for the standardised brightness measurements, we obtain

$$r_{XY} = \frac{1}{n}\sum (xy)$$

$$= \frac{1}{5}[\,1.5\,(1.4) + (0.5)\,(0.7) + 0\,(0) + (-0.5)\,(-0.7) + (-1.5)\,(-1.4)\,]$$

$$= 0.98$$

Correlating the first set with the third gives

$$r_{XY} = \frac{1}{5}[\,1.5\,(1.0) + 0.5\,(-1.4) + 0\,(0.4) + (-0.5)\,(-1.0) + (-1.5)\,(1.0)\,]$$

$$= -0.04$$

You can see that the first and second set of measurements are highly correlated, whereas the correlation between the first and third set is negative but very close to zero. What does this correlation analysis of the measurements tell us about the solar atmosphere? Let us first interpret the values of the correlation coefficient. The high positive correlation tells us that the bright points on one photograph correspond to the bright points on the other. A high negative correlation means that the bright points on one correspond to darker points on the other. And zero correlation means we can say nothing about one set by looking at the other, That is, such sets are unrelated. From a larger set of these measurements it was found that the correlation coefficient decreases towards zero without becoming negative. Its value is reduced to about 1/2 in an interval of five minutes and to less than 1/4 in ten minutes.

We present here the possible explanation of these results, emerging from an analysis given by scientists working in this area: It is known that intensely hot vapour formations are spread unevenly in the solar atmosphere - the hot and cool areas cover hundreds of kilometers and their outlines are roughly hexagonal. This pattern is called granulation and the hexagonal shaped elements are called granules. Granules are, in fact, bubbles of hot gas welling up from the centre of the sun and disturbing the outer layers. The gas is so hot that it emits radiation in the visible region. The above correlation analysis of the brightness measurements of the emitted radiation suggests that, at random times averaging about 5 min apart, new granules appear and gradually cool. Each granule is replaced by another at the next random time. If the cooling is slow, and the replacement is slow in coming, the correlation coefficient goes to zero slowly. If a replacement comes faster, it goes to zero rapidly. This explanation is in close agreement with other available facts and measurements.

Do you now appreciate how this simple statistical tool finds application in areas of advanced research? Would you like to apply it to some measurements? However, before you compute correlation coefficient for a given set of data, you must understand some basic concepts about it.

Firstly, you must understand that r_{XY} is an index number whose value indicates the strength of only the linear relationship between the two random variables X and Y. Curvilinear relationship between variables, however strong it may be, need not be reflected in the correlation. The sign of r_{XY} indicates the direction of the relationship: a high positive value of r_{XY} indicates that high values of one variable correspond to high values of the other variable, and low values of one correspond to low values of the other. A negative r_{XY} means just the opposite: as the values of one variable increase, the values of the other variable decrease (or vice-versa).

An easy way to get an idea of the type of relationship and the correlation between two variables is to plot one against the other. Such plots are called scatter diagrams. Scatter diagrams exhibiting various degrees of correlation are shown in Fig, 7.1.

Although correlation coefficient is usually the first number that is calculated when we are given several sets of measurements, we must take great care in interpreting the results. You must always remember that the correlation coefficient indicates only the extent or degree of linear relationship between two variables. It does not tell us anything about the causal relationship. Thus, a high correlation between two sets of numbers does not indicate a cause-effect relationship. Two variables that may appear to have a high positive correlation may not even be directly associated. So even if you find a high correlation between two sets of variables X and Y you should not immediately conclude that change in one variable influences (or causes) change in the other. In scientific experiments, you will have to take Into consideration all factors that influence Ybefare arriving at any wnclusion.

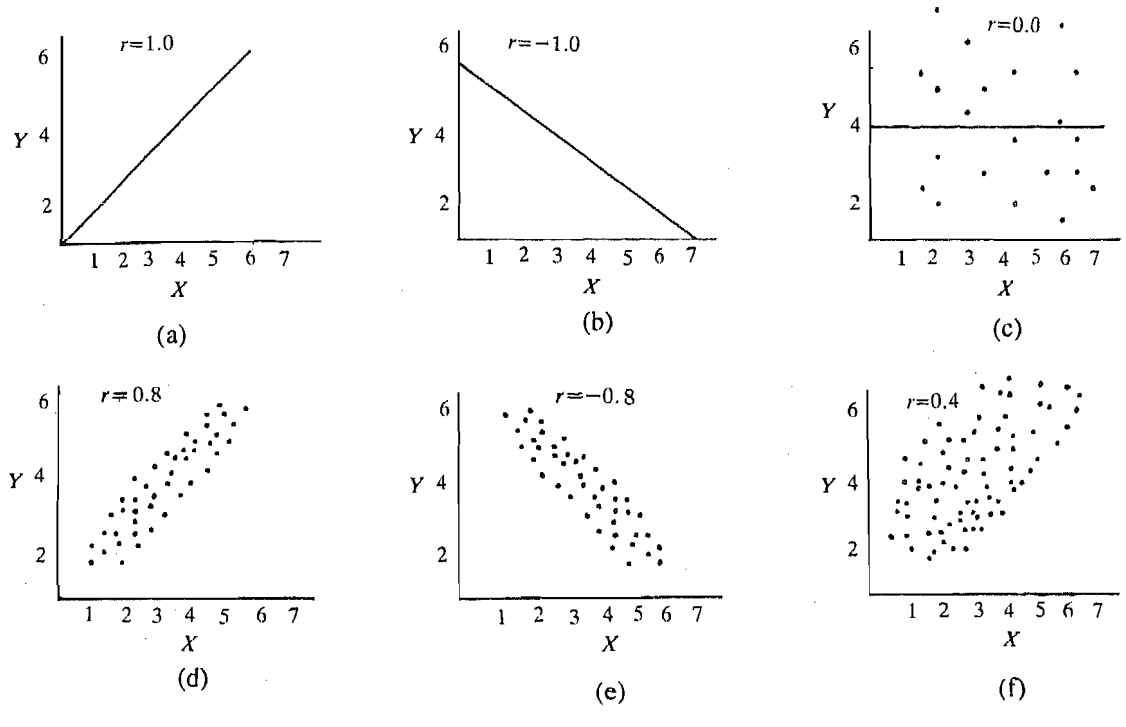

Fig. 7.1 : Scatter diagrams showing various degrees of correlation. (a) For $r_{XY} = +1.00$, all values fall on a straight line showing a direct relationship; (b) for $r_{XY} = -1.00$, the values again fall on a straight line showing an inverse relationship. For both (a) and (b), knowing one variable, we can predict the other; (c) no trend is found in the variation of X and Y; (d) and (e) there is a substantial correlation between X and Y; (f) a low degree of relationship yields low r_{XY}.

Let us now work out another example on computation of correlation coefficient.

Example 1

The heat capacity of liquid sulphuric acid was measured at various temperatures yielding the following set of data:

heat capacity (in cal/°C) :	0.38	0.39	0.40	0.41	0.45	0.46
temperature (in °C)	50	100	150	200	250	300

Compute the correlation coefficient r_{XY}.

Solution

Let X represent the temperature and Y, the heat capacity. We arrange the data in a tabular form and compute r_{XY} as follows:

x (°C)	y (cal/°C)	x^2	y^2	xy
50	0.38	2500	0.1444	19.00
100	0.39	10000	0.1521	39.00
150	0.40	22500	0.1600	60.00
200	0.41	40000	0.1681	82.00
250	0.45	62500	0.2025	112.50
300	0.46	90000	0.2116	138.00
$\sum x = 1050$	$\sum y = 2.49$	$\sum x^2 = 227500$	$\sum y^2 = 1.0387$	$\sum xy = 450.50$

Using Eq. (7.1) we have

$$r_{XY} = \frac{6\,(450.50) - (1050)\,(2.49)}{\sqrt{6\,(2,\,27,\,500) - (1050)^2}\,\sqrt{6\,(1.0387) - (2.49)^2}}$$

$$= \frac{88.5}{\sqrt{(512.35)}\,\sqrt{(0.179)}}$$

$$= 0.96$$

This suggests a high degree of positive correlation.

You must now solve an SAQ to practice computing the correlation coefficient.

SAQ 1

Spend 10 *min*

The measurements of surface tension of water at various temperatures are given in the following table.

T(°C)	10	20	30	40	50	60
Y (dynes cm^{-1})	74.0	73.0	71.0	70.0	68.0	66.0

.Calculate the correlation coefficient for this data.

Once we have determined the correlation coefficient for a given set of data, we know that the variables are related. But the correlation coefficient does not specify how the variables are related. To determine the quantitative relationship between variables, we make use of another tool of statistics - the **regression analysis.** Once a relationship is found in the form of an equation, we can use it to **predict** the value of one of the variables if the value of the other variable is known.

7.3 REGRESSION ANALYSIS

We will restrict our discussion to the case of two variable linear regression. The term linear signifies that a straight line describes the **average relationship** between the two variables. Our aim is to **predict** or **estimate** the value of one variable for a given value of another variable using regression analysis. The variable to be estimated is called the dependent variable. The variable that presumably causes variations in the **dependent variable** is termed the **independent variable.** Let us denote the dependent variable by y and the independent variable by x. As the first objective of linear regression, we must obtain the equation of a straight line (called the **regression line)** that describes the average relationship between the dependent and the independent variables, In regression analysis we use the notation

$$\hat{y} = a + bx \qquad (7.2)$$

for the equation of the regression line. The constants a and b are called the **regression coefficients.**

You may wonder why we use the notation $\hat{y}$ instead ofy. To find the answer, let us consider the data given in Example 1. When we obtain the regression equation for the data, we may wish to estimate the heat capacity of sulphuric acid at 220°C. We can obtain the required estimate by substituting x = 220°C in the corresponding regression equation. However, it is quite possible that the value of y obtained from Eq. (7.2)does not agree with its observed value, if a measurement is made at 220°C. Thus a different symbol y is used to denote the estimated (or predicted) value of the dependent variable.

The question now is: How do we draw the regression line? For example, consider the scatter diagram shown in Fig. 7.2. We can draw more than one line to fit the data. How do we know which one of them gives the best fit to the data?

Statisticians use a method for determining the equation which best fits a given set of data. Let us study this method, known as the method of **least squares.** It applies not only to linear relationship but to curvilinear relationship also. This is the most widely used curve-fitting technique in statistics.

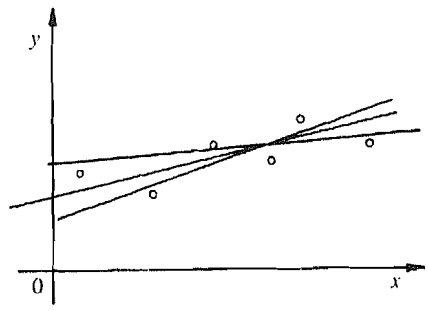

Fig. 7.2: More than one straight line can be drawn to fit this data.

7.3.1 The Method of Least Squares

Let us first consider the problem of fitting a regression line to a given set of data. Whenever we draw an estimated regression line, not all points will necessarily lie on it. Some may lie above it and some may be below. The difference between any point and the corresponding point on the regression linc is called the (vertical) **deviation** from the linc. It represents the difference between what we predicted and what was actually observed or measured.

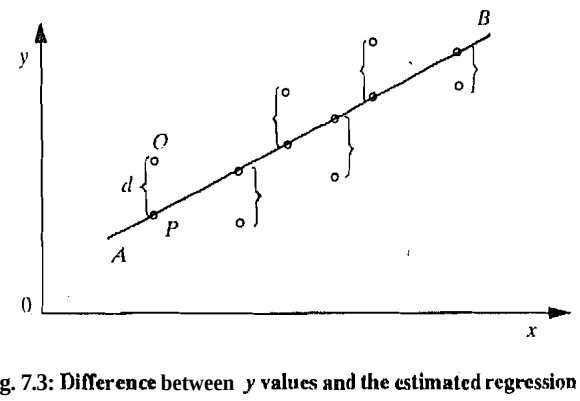

Fig. 7.3: Difference between y values and the estimated regression line.

The method of least squares determines the regression line in such a way that the **sum of the squares of these vertical deviations is as small as possible.**

Study Fig. 7.3 to understand this statement. Here AB represents the estimated regression linc for y and x. *Let* O be the observed value corresponding to the predicted value P on the line. Then the curve for which the sum of tbe squares of the difference d between the observed and the predicted value is a minimum, is called the best- fitting curve.

Since we are considering a linear relationship, the regression line is the ***straight line that best fits a set of data points nccording to the least-squares criterion.*** We state here the value of the regression coefficients obtained from the least-squares method.

Regression Equation

The equation of the estimated regression line is

$$\hat{y} = a + bx \qquad (7.3a)$$

where

$$b = \frac{n\sum xy - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}, \quad a = \frac{1}{n}(\sum y - b\sum x) \qquad (7.3b)$$

and n is the number of pairs of data points

Note: The equation of the regression line is called the **regression equation.** The proof of Eq.(7.3) is given in the margin along its side.

We can use Eqs. (7.3a) and (7.3b) to estimate the values of y for given values ofx. However, remember that ***the values of y can be estimated only for those values of x which lie within the range of the given set of data. We cannot use these equations to extrapolate y values.***

The derivation of Eqs.(7.3a,b) is as follows: Let F (a, b) denote the sum of squared deviations:

$$F(a, b) = \sum (y - \hat{y})^2$$

Substituting $a + bx$ for $\hat{y}$ in the above equation gives

$$F(a, b) = \sum (y - a - bx)^2$$ Since $F(a,b)$ should bc a minimum, we impose the conditionsthat

$$\frac{\partial F}{\partial a}(a, b) = 0, \frac{\partial F}{\partial b}(a, b) = 0$$

which gives

$$-2\sum (y - a - bx) = 0$$

and

$$-2\sum x(y - a - bx) = 0$$

Simplyfing these equations we get

$$\sum y = na + b\sum x$$

and

$$\sum yx = a\sum x + b\sum x^2$$

These are termed the normal equations.

Solving for a and b we get the required result.

$$b = \frac{6(450.5) - (1050)(2.49)}{6(227500) - (1050)^2} = .0003$$

and

$$a = \frac{1}{6}[2.49 - (0.0003)\,1050] = 0.36$$

Thus, the equation of the regression line for the data given in Example 1 is

$$\hat{y} = 0.36 + 0.0003\,x$$

You may like to quickly work out the regression equation for a given set of data.

Spend 5 min

SAQ 2

Determine the regression equation for the data of SAQ 1.

We hope that you have not formed the impression that the method of least squares applies only to straight lines. We can even fit a polynomial of any degree m ($< n$) to a set of n data points using this method. We can also fit linear functions of the form $y = \sum a_i x_i$ and the non-linear functions which are reducible to linear forms. It is not our intention here to go into the details of these applications. If you are interested in these details, you may study the mathematics courses entitled 'Probability and Statistics' and 'Numerical Analysis'.

If the data seems to fit a curve other than a straight line we can use some transformations to obtain a linear regression, We suminarise in the table below some transformations which you can carry out to fit a given set of data to a straight line.

Table 7.2: Some useful transformations to linearise.

The form of the function relating y and x	Transformation	Form of linear regression equation
Exponential : $y = ce^{bx}$	$y' = \ln y$	$y^* = a + bx$
Power : $y = cx^d$	$y^* = \log y;\ x^* = \log x$	$y' = a + bx^*$
Reciprocal : $y = a + \frac{b}{x}$	$x^* = \frac{1}{x}$	$y = a + bx^*$
Hyperbolic function : $y = \frac{x}{c + dx}$	$Y = \frac{1}{y},\ x^* = \frac{1}{x}$	$y^* = a + bx^*$

We will now illustrate the method of least squares with the help of an example.

Example 2

The pressure of a gas corresponding to various volumes V is measured, yielding the following data

V (cm^3)	50	60	70	90	100
p (kg cm^{-2})	64.7	51.3	40.5	25.9	7.8

Fit the data to the equation $pV^\gamma = C$.

Solution

Let us take the natural log of both sides of the equation:

$$pv^\gamma = C$$

We obtain the linear model:

$$\ln p = \ln C - \gamma \ln V,$$

We can draw up a table in the following form:

V	p	ln V	ln p	$(\ln V)^2$	$(\ln V)(\ln p)$
50	64.7	3.91	4.17	15.288	16.305
60	51.3	4.09	3.94	16.728	16.115
70	40.5	4.25	3.70	18.062	15.725
90	25.9	4.50	3.25	20.250	14.625
100	7.8	4.61	2.05	21.252	9.451
		$\sum(\ln V) = 21.36$	$\sum(\ln p) = 17.11$	$\sum(\ln V)^2 = 91.580$	$\sum(\ln V)(\ln p) = 72.221$

From Eq. (7.3) the regression coefficients are

$$\gamma = \frac{5(72.211) - (21.36)(17.11)}{5(91.580) - (21.36)^2} = -2.66$$

and

$$\ln C = \frac{1}{5}[17.11 + (2.66)(21.36)] = 14.8$$

or $C = 2676445$

Thus, the regression equation is $pV^{-2.66} = 2676445$

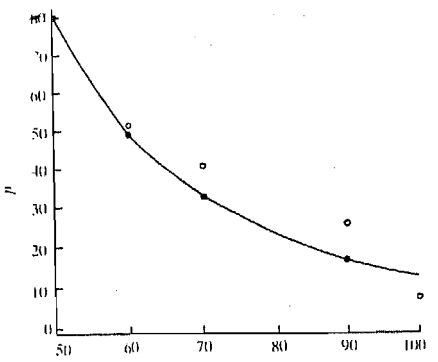

Fig. 7.4: Pressure and vdume data of Example 2 and (he regression curve fitting the data.

The pressure, volume and the fitted regression curve are shown in Fig. 7.4.

So far we have discussed how the regressioit equation is fitted to a set of data and can be used to get an estimate of the dependent variable. When we predict the value of the dependent variable from a regression equation, the question naturally arises: "How dependable is the estimate?" That is, how close or how far the value of the dependent variable is from its actually observed value. For this we need to know the standard error of estimate which determines the differences between these values. Let us discuss it now.

7.3.2 The Standard Error of Estimate

We can get an idea about the error by looking at the scatter diagram and the correlation coefficient When the data points are closely spaced around the regression line (i.e., correlation is high), it is logical to assume that the estimate will be more dependable. An estimate based on a regression line where the spread is greater, or correlation is low, will be less reliable. So we need a measure of the extent of the spread or scatter of the points around the regression line. The smaller this measure is, the more dependable will the estimate be. This measure is termed the standard error of estimate. It is, in reality, the standard deviation that measures the scatter of the observed values around the regression line. The only difference is that the standard deviation is measured from the mean, while the standard error is measured from the regression line. There are two basic assumptions underlying this idea:

(i) For each value of x, there is a set of y values that is normally distributed about the regression line (see Fig. 7.5).

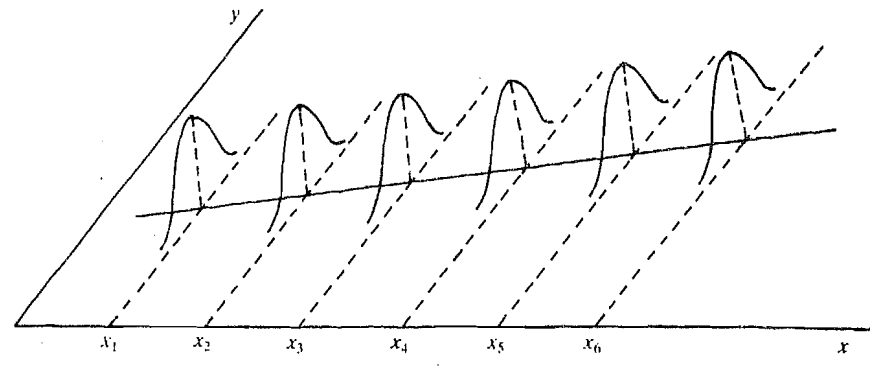

Fig. 7.5 : The assumptions underlying the concept of standard error of estimate.

(ii) There is the same amount of variability around the regression line at each x, i.e., the standard deviation of y values is the same for all x values.

The standard error of estimate is defined as

$$s_e = \sqrt{\frac{\sum (y-\hat{y})^2}{n-2}} \qquad (7.4a)$$

where; is the estimated value, y is the observed value and n is the total number of points in the set of data. In computing s_e, it is easier to use the following formula:

$$s_e = \left[\frac{\sum y^2 - a\sum y - b\sum xy}{(n-2)}\right]^{1/2} \qquad (7.4b)$$

where a and b are the regression coefficients given by Eq.(7.3b). You can readily verify that s_e = 0.04 for the data of Example 1 which indicates a good fit.

You should now attempt an SAQ to check whether you have grasped these concepts or not.

Spend
10 min

SAQ 3

A mass M is suspended from the centre of a steel bar supported at its ends. The depression y is measured by means of a dial height indicator. The following readings are obtained:

M (kg)	0	1	2	3	4
y (μm)	1600	1300	950	600	250

Calculate the best value of the slope and the standard error of estimate for y.

Before proceeding further you may like to stop for a while and reflect on what you have studied so far. Thus far we have introduced you to two tools of statistical analysis: correlation analysis and regression analysis, which go hand in hand in statistical interpretation of data. These two tools have been developed to study and measure the statistical relationship between two or more variable. We have restricted our study lo the linear relationship between two variables. Using correlation analysis, we aim to measure the strength or closeness of the linear relatioliship between the variables. Regression analysis enables us to describe the quantitative relationship between the variables. Thus, it enables us to estimate the values of one variable from those of the other. The method of least squares is a very useful statistical technique for this purpose. By computing the standard error of estimate we can also know how dependable the estimated value is, In summary, wc can say that regression analysis asks "What is the pattern of the relationship between the variables?" And correlation analysis asks "How strong is the relatioliship between the variables?"

Having introduced you briefly to correlation and regression analysis, we now turn our attention to another application: the statistical theory of errors.

7.4 THE STATISTICAL THEORY OF ERRORS

Scientific investigations rely heavily on experiments. As you know experimental investigations essentially involve measurement. Most of the magnitudes used in scientific investigations result, directly or indirectly, from measurements. But measurements are never perfect. For example, look at the following table. It lists the measured values of the speed of light by various workers at different times:

Year	Investigator	Speed of light (c) in km s^{-1}
1875	Cornu	299 990 ± 200
1880	Michelson	299 910 ± 50
1883	Newcomb	299 860 ± 30
1883	Michelson	299 850 ± 60
1926	Michelson	299 796 ± 4

What do these results tell us? These results tell us that no experimental measurement yields the exact value of the measured quantity. All measurements are inaccurate to some extent. It is, however, important to make a reasonable assessment of accuracy of the measurement. We would, after all, like the difference between the measured (or observed) value and the 'true value' to be as small as possible.

You will note from the table that Michelson's experimental resull of 1926 only tells us that the value of the speed of light lies somewhere between 299, 792 km s^{-1} and 299, 800 km s^{-1}. The present day accepted value of velocity of light is 299,792,459.0 ± 0.08 m s^{-1}.

Notice that the result is still not exact, though the estimated error of observation (± 0.8 m s^{-1}) is very small. By definition, the error of observation is the difference between the true value and the observed (measured) value of a physical quantity

$$\text{Error} = (\text{True value}) - (\text{Observed value}) \qquad (7.5)$$

You may be wondering: How do errors arise in experimental investigations? The next section will help us answer this question.

7.4.1 Types of Errors

Errors arise from a variety of reasons. They may be due to the limitations of the observer, the measuring instruments used, the method used or some changes in external factors, such as atmospheric temperature, vibrations, or disturbance in power supply etc. It is customary to classify errors of observation as either systematic or random, although it is not always easy to distinguish between the two.

Systematic error

A systematic error is one that is due to a definitely identifiable cause. Under given conditions it affects a measurement in a regular way. If it is constant it affects all measurements and shifts them by the same magnitude. For example, if the zero reading in a measuring instrument (say, an ammeter) has been wrongly set, there will be a systematic error in the measurements. Again a stop-watch running slow or fast will produce a systematic error. Systematic errors may be subdivided into three kinds:

1. Instrument errors are those resulting from imperfections in a measuring instrument, as due to the stop-watch mentioned above.
2. Natural errors arise from change in ambient conditions. For example, temperature changes may cause expansion of metals, electrical instruments may be affected by external magnetic fields, and so on.
3. Personal errors result from the physical limitations of an observer. For example, while measuring the time period of a pendulum, an observer might press the stop-watch a bit early or a bit late.

Systematic errors need not be revealed by repealed observations. The safest course for an experimenter is to regard them as effects to be discovered and eliminated.

There are quite a few instances of systematic errors introduced in otherwise well performed experments. A classic example was provided in the measurement of electronic charge by Millikan. In Millikan's oil drop experiment, it is necessary to know the viscosity of air. Millikan used the value of the viscosity of air measured by Harrington which contained a systematic error. As a result, the value of e obtained by Millikan was

$e = (1.591 \pm 0.002) \times 10^{-19}$ C

Compare it with the present value

$e = (1.602189 \pm 0.000005) \times 10^{-19}$ C

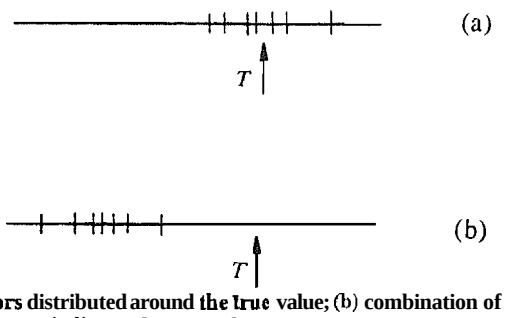

Fig. 7.6 : (a) Random errors distributed around the true value; (b) combination of systematic and random errors. T indicates the true value.

Random error

Random errors arise due to unknown reasons, which cannot be controlled by the experimenter. For example, small irregularities in the motion of the pendulum may cause variations in successive measurements and may be regarded as random errors. Random errors are just as likely to be positive as negative and their magnitudes can vary. In the absence of systematic errors, the random errors cause successive readings to spread about the true value of the quantity (Fig. 7.6a).If, in addition, a systematic error is present, the readings spread about some displaced value. (Fig. 7.6b)

In the context of errors, you should also learn to distinguish between the terms accuracy and precision. A measurement is said to be accurate if it is relatively free from systematic errors; it is said to be precise if the random error is small.

Having learnt about various types of errors and how they arise, you would wish to know how to minimise them and estimate their magnitude. You have just studied that systematic errors can be reduced only by recognising their causes and eliminating them. This, of course, requires considerable skill and experience on the experimenter's part. However, we can detect and minimise random errors by repeating the measurements of a quantity a number of times. Further, random errors can be estimated by statistical methods. Let us now learn these methods. Yo11 will first learn how to estimate random errors when a single quantity is being measured in experiment.

7.4.2 Estimation of Random Errors for a Single Variable

Suppose a physical quantity is measured n times yielding the values

$$x_1, x_2, \ldots\ldots x_n$$

At the outset let us assume that there is no systematic error in the measurements. Although we have only n actual measurements, let us imagine that we go on making the measurements so that we end up with a large number N. We call this hypothetical set of a very large number of readings a distribution. Our actual set of n measurements should be regarded as a random sample taken from the distribution of N measurements. We then define a function $f(x)$, known as the distribution function, such that $f(x)\,dx$ gives the probability that a single measurement taken at random from the distribution will lie in the interval x and $x + dx$. In other words $f(x)$ is a probability distribution.

From Unit 6 you would recall that most experimental observations obey the normal (or the Gaussian) distribution given by Eq.(6. 7):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\bar{x})^2}{2\sigma^2}\right] \qquad (7.6)$$

wberex represents the value obtained from a single measurement, $\bar{x}$ represents the mean of the distribution and σ is the standard deviation of the distribution.

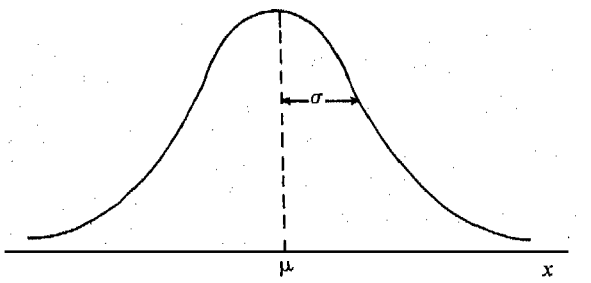

Fig. 7.7: The Gaussian distribution obeyed by most experimental observations

You would recall that the normal curve is symmetric about the mean and its maximum value occurs at $x = \bar{x}$. In other words, the most probable value of x is $\bar{x}$. Thus, the best value of the

quantity being measured is the arithmetic mean $\bar{x}$ of the measurements defined as

$$\bar{x} = \frac{1}{n}\sum x \qquad (7.7)$$

The standard deviation **a** is a measure of the spread of the observations (Fig. 7.8). It is called the **standard error.**

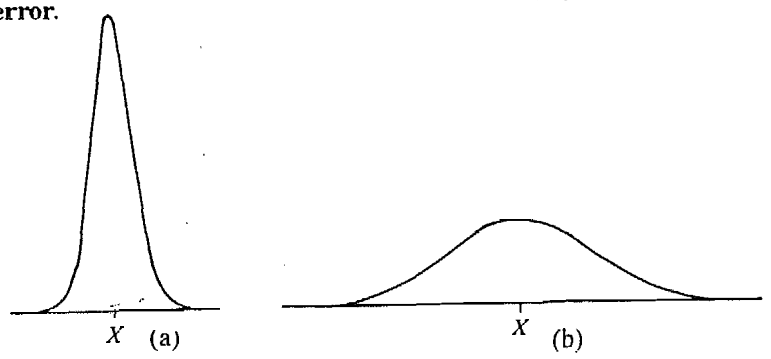

Fig.7.8: Distribution function for (a) a precise set of measurements has a smaller value of a; and (b) an imprecise set of measurements has a large value of a

According to the normal distribution for a set of readings with mean $\bar{x}$, roughly two thirds of the readings should lie between $\bar{x} \pm a$. However, you would note that we, are considering a random sample of a limited number (n) of measurements from a large (infinite) number of measurements to which the normal distribution refers. And it is unlikely that the mean of only n values would exactly equal the mean of an infinite number. Moreover, a different sample of n values would in general have a different mean. We, therefore, also need to ask how the means of various samples of n values are distributed. Thus, we define another quantity called the **standard error in the mean a:**

$$\sigma_m = \frac{\sigma}{\sqrt{n}} \qquad (7.8)$$

where σ_m is a measure of how reliably the sample mean indicates the mean of the infinite number. Whereas σ represents the error in a single measurement, **a** represents the error in the mean of n measurements. The final result of the measurement is written as

$$x = \bar{x} \pm \sigma_m \qquad (7.9)$$

This result implies that, in the absence of systematic error, the probability that the true value of the quantity lies in the range ($\bar{x} \pm a$) is about 2/3.

So far we have, theoretically defined the quantities σ and **a** that are the measures of the error in the single measurement and in the mean. The next question to ask is: How do we determine them from actual measurements? We will now discover the answer to this question.

Determination of σ and a

Consider a set of n measurements $x_1, x_2, ..., x_n$. Let the error in the i^{th} reading be

$$e_i = x_i - x_t \qquad (7.10a)$$

where x_t is the true value of the quantity, which is unknown. The error in the mean E_m is

$$E_m = \bar{x} - x_t = \left(\frac{1}{n}\sum x_i\right) - x_t = \frac{1}{n}\sum_i (x_i - x_t) = \frac{1}{n}\sum_i e_i \qquad (7.10b)$$

The best estimate of σ is $[\,(1/n)\sum e_i^2\,]^{1/2}$. But since the errors e_i come from the true value x_t they are not known. So how do we find σ? A way around this difficulty is provided by introducing the **residuals:**

The residual d_i of the i^{th} measurement is defined by

$$d_i = x_i - \bar{x} \qquad (7.11)$$

Unlike the error, the residual is a known quantity and we can relate it to **a** and σ_m.

From Eqs. (7.10a) and (7.10b) we have that

$$x_i - \bar{x} = e_i - E_m \qquad (7.12)$$

It follows that

$$\sum d_i^2 = \sum (x_i - \bar{x})^2 = \sum (e_i - E_m)^2$$

$$= \sum e_i^2 - 2E_m \sum e_i + nE_m^2$$

$$\text{or} \quad \sum d_i^2 = n\sigma^2 - nE_m^2 \quad \left[\because \sum e_i = \sum (x_i - x_t) = n(\bar{x} - x_t) = nE_m \right] \tag{7.13}$$

Now from Eq. (7.10b)

$$E_m^2 = \frac{1}{n^2}\left(\sum e_i\right)^2 = \frac{1}{n^2}\left[\sum e_i^2 + \sum_{i \neq j} e_i e_j\right]$$

$$= \frac{\sigma^2}{n} + \frac{1}{n^2}\sum_{i \neq j} e_i e_j$$

Thus, Eq. (7.13) yields

$$\sum d_i^2 = (n-1)\sigma^2 - \frac{1}{n}\sum_{i \neq j} e_i e_j$$

For large values of n, the sum of the cross products of the errors becomes very small compared to $(n-1)\sigma^2$. Hence, we get

$$\sigma^2 = \frac{1}{n-1}\sum d_i^2 \tag{7.14}$$

From Eq. (7.8) we obtain

$$\sigma_m^2 = \frac{1}{n(n-1)}\sum d_i^2 \tag{7.15}$$

So when we are given a sample set of ineasurements for a single variable, we can reasonably assume that the larger set, from which the sample is taken, follows the normal distribution. Then the mean of the sample (Eq. 7.7) is the best value of the unknown magnitude and σ_m gives the standard error of the mean. The final result is quoted as

$$x = \bar{x} \pm \sigma_m \tag{7.16}$$

Remember that in arriving at Eq. (7.16) we have assumed that there are no systematic errors. Let us now compute $\bar{x}$ and σ_m for a given set of data.

Example 3

The resistance of a coil of wire (in Ω) was measured to be 4.615, 4.638, 4.597, 4.634, 4.613, 4.623, 4.659, 4.623. Calculate the best value of the resistance and the standard error of the mean.

Solution

We calculate the mean of the sample and the residual of each measurement and use Eqs. (7.14) and (7.15). The calculations are tabulated below:

x_i (Ω)	$d_i = x_i - \bar{x}$	d_i^2
4.615	− 0.010	0.000100
4.638	+ 0.013	0.000169
4.597	− 0.028	0.000784
4.634	0.009	0.000081
4.613	− 0.012	0.000144
4.623	− 0.002	0.000004
4.659	0.034	0.001156
4.623	− 0.002	0.000004
$\bar{x} = 4.625$		$\sum d_i^2 = 0.002442$

$$\sigma_m^2 = \frac{1}{56} \times 0.002442 = 0.000044, \text{ and } \sigma_m = .007$$

Therefore, the best value of the resistance is

$$(4.625 \pm 0.007)\ \Omega$$

You should now work out the following SAQ to compute the error in a given set of measurements.

SAQ 4

Spend 10 min

The number of a -particles, N, emitted per minute from a uranium sample were measured to be 36076,35753,36116,35884,36136,35741,35640,36124,36076.

Calculate the best value of N and the standard error in the mean.

In the previous section you have learnt to estimate the random error in the measurement of a single variable. In most experiments we do not measure a physical quantity directly. The desired quantity is a function of other quantities and only they are measured. For example, you may have determined the Young's modulus for the material of a beam in the first physics laboratory course. Recall that it is given by $Y = \frac{WL^3}{4\delta bd^3}$, where W is the load on the beam, 6 the depression at the centre of the beam, L , B and d are the beam's length, width and depth of its cross-section, respectively. You must have directly measured δ, l, b and d to know Y. Now, as you know, each of the directly measured quantities possesses some error. So the question now is to estimate the error in a quantity, which is a function of several directly measured quantities, if we know the errors in their measurements. This is what you are going to learn in the next section.

7.4.2 Random Error Estimation for Indirect Measurements

Let us consider the problem of estimating error in a quantity Q which is a function of n other quantities $u_1, u_2, u_3, \ldots, u_n$.

$$Q = Q(u_1, u_2, u_3, \ldots, u_n)$$

The quantities $u_1, u_2, \ldots u_n$ are directly measured. Let the random errors in these quantities be $\pm A u_1, \pm A u_2, \ldots \pm A u_n$. These errors will give rise to a random error, say $Q \pm \Delta Q$ in Q. If the random errors are small, the best value of Q (say $\overline{Q}$) will be that calculated from the mean values of the measured quantities $\overline{u}_1, \overline{u}_2, \ldots \overline{u}_n$. Then, provided that the function $Q(u_1, u_2, \ldots, u_n)$ is continuous, we can make use of the result

$$\Delta Q = \frac{\partial Q}{\partial u_1}\Delta u_1 + \frac{\partial Q}{\partial u_2}\Delta u_2 + \ldots + \frac{\partial Q}{\partial u_n}\Delta u_n \qquad (7.17)$$

where $A u_1, A u_2, \ldots, A u_n$, represent small deviations from $u_1, u_2, \ldots u_n$, respectively, and ΔQ represents the corresponding deviation of Q from $\overline{Q}$. The partial derivatives $\frac{\partial Q}{\partial u_1}$ etc. arc evaluated at the mean values $\overline{u}_1, \overline{u}_2, \ldots$ If a large number N of sets of values of $u_1, u_2, \ldots u_n$ are measured, the standard deviation σ_Q of the resulting N values of Q can be obtained from Eq. (7.14) as

$$\sigma_Q^2 = \frac{1}{N-1}\sum (\Delta Q_i)^2$$

where $\Delta Q_i = Q_i - \overline{Q}, \quad i = 1, 2, \ldots N$

Using Eq. (7.17) we can write

$$\sigma_Q^2 = \frac{1}{N-1}\sum \left[\frac{\partial Q}{\partial u_1}\Delta u_1 + \frac{\partial Q}{\partial u_2}\Delta u_2 + \ldots + \frac{\partial Q}{\partial u_n}\Delta u_n\right]^2 \qquad (7.18)$$

When the quantity within the brackets on the right hand side is squared, we get two types of terms.

One kind of terms are squares like

$$\left[\frac{\partial Q}{\partial u_1}(\Delta u_1)\right]^2$$

The other kind are 'cross' terms like

$$2\frac{\partial Q}{\partial u_1}\frac{\partial Q}{\partial u_2}\Delta u_1 \Delta u_2$$

Since Δu_1, Δu_2 etc. represent normal errors they are just as likely to be positive as negative. So the sum of the large number of cross terms will be close to zero, or in any case much smaller than the sum of the square terms. We can prove this result rigorously for normal distribution provided the variables are independent. Without going into these details here, we drop the cross terms and write

$$\sigma_Q^2 = \frac{1}{N-1}\sum\left[\left(\frac{\partial Q}{\partial u_1}\right)^2(\Delta u_1)^2 + \left(\frac{\partial Q}{\partial u_2}\right)^2(\Delta u_2)^2 + \ldots + \left(\frac{\partial Q}{\partial u_n}\right)^2(\Delta u_n)^2\right]$$

$$= \left(\frac{\partial Q}{\partial u_1}\right)^2\frac{1}{N-1}\sum(\Delta u_1)^2 + \left(\frac{\partial Q}{\partial u_2}\right)^2\frac{1}{N-1}\sum(\Delta u_2)^2 + \ldots + \left(\frac{\partial Q}{\partial u_n}\right)^2\frac{1}{N-1}\sum(\Delta u_n)^2$$

Since the partial derivatives are computed at the respective mean values $\bar{u}_1, \bar{u}_2, \ldots$, they are constants and have been taken out of the summation. Now, $\frac{1}{N-1}\sum(\Delta u_i)^2$ is simply $\sigma^2_{\bar{u}_i}$. Thus, we obtain

$$\sigma_Q^2 = \left(\frac{\partial Q}{\partial u_1}\right)^2\sigma_{u_1}^2 + \left(\frac{\partial Q}{\partial u_2}\right)^2\sigma_{u_2}^2 + \ldots + \left(\frac{\partial Q}{\partial u_n}\right)^2\sigma_{u_n}^2 \qquad (7.19)$$

We can express this result in a compact form as

$$\sigma_Q^2 = \sum\left(\frac{\partial Q}{\partial u_i}\right)^2\sigma_{u_i}^2 \qquad (7.20)$$

where the partial derivatives are computed at the mean values $\bar{u}_1, \bar{u}_2, \ldots \bar{u}_n$. In terms of the standard error in the mean, we have

$$\sigma_{\bar{Q}}^2 = \sum\left(\frac{\partial Q}{\partial u_i}\right)^2\sigma_{\bar{u}_i}^2 \qquad (7.21)$$

where $\sigma_{\bar{u}_i}$ is the standard error in the mean of u_i. The best value of Q is written as

$$Q = \bar{Q} \pm \sigma_{\bar{Q}} \qquad (7.22)$$

Let us consider an example to illustrate the use of Eq. (7.22).

Example 4

The combined resistance R of two resistors (of resistance R_1 and R_2) connected in parallel is given by

$$R = \left(\frac{1}{R_1} + \frac{1}{R_2}\right)^{-1}$$

Determine the best value of R, given that $R_1 = (200 \pm 5)\ \Omega$ and $R_2 = (600 \pm 10)\ \Omega$. The errors in R_1 and R_2 are the standard errors in the mean.

Solution

The mean value of R is

$$\bar{R} = \left(\frac{1}{\bar{R}_1} + \frac{1}{\bar{R}_2}\right)^{-1} = \left(\frac{1}{200} + \frac{1}{600}\right)^{-1} = 150\ \Omega$$

To use Eq. (7.22) we must calculate $\frac{\partial R}{\partial R_1}$ and $\frac{\partial R}{\partial R_2}$ at $R_1 = \bar{R}_1$ and $R_2 = \bar{R}_2$:

$$\frac{\partial R}{\partial R_1} = \frac{R_2^2}{(R_1+R_2)^2} = \frac{(600\,\Omega)^2}{(800\,\Omega)^2} = \frac{9}{16}$$

$$\frac{\partial R}{\partial R_2} = \frac{R_1^2}{(R_1+R_2)^2} = \frac{(200\,\Omega)^2}{(800\,\Omega)^2} = \frac{1}{16}$$

Hence Eq. (7.21) gives

$$\sigma_R^2 = \left(\frac{9}{16}\right)^2 (5)^2 + \left(\frac{1}{16}\right)^2 (10)^2 = 8$$

or $\sigma_R = 3$

Therefore, the best value of R is $(150 \pm 3)\ \Omega$.

We end this section with an exercise for you.

SAQ 5

Sped
10 min

The modulus of rigidity of a wire is given by

$$n = \frac{2LN}{\pi r^4 \theta}$$

where L and r are the length and radius, respectively, of the wire. The angle θ is the magnitude of angular displacement when the wire is subjected to a torque N. The following measurements are made for r, L and θ/N:

$r = 1.00 \pm 0.02$ mm

$L = 500 \pm 2$ mm

$\theta/N = 4.00 \pm 0.10$ rad $N^{-1}\,m^{-1}$

Obtain the best value of n.

To sum up, in this section you have studied the types of errors that arise in experimental measurements. You have also learnt to estimate the standard error in the mean for random errors arising in directly and indirectly measured variables.

Let us now summarise the contents of this unit.

7.5 SUMMARY

- The **correlation coefficient** determines the strength of linear association between two variables. It is given as

$$r_{XY} = \frac{n\sum(xy) - \sum(x)\sum(y)}{\sqrt{n\sum(x^2) - (\sum x)^2}\ \sqrt{n(\sum y^2) - (\sum y)^2}}$$

  where n is the number of pairs of data points. A value of r_{XY} close to +1 or –1 indicates high correlation. When r_{XY} is close to zero, correlation is low. However, a high correlation between two variables does not indicate a cause-effect relationship.

- **Regression analysis** is used to estimate the value of one variable from that of the other. For linear relationship between two variables, the method of least squares can be used to obtain the equation of the regression line, which is

$$\hat{y} = a + bx$$

  where

$$b = \frac{n\sum xy - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}, \quad a = \frac{1}{n}(\sum y - b\sum x)$$

and n is the number of pairs of data points. The regression equation allows us to predict the value of y for any value of x which lies within the range of the x values in the given data.

- The regression equation can be used to fit data which suggest a non-linear relationship between variables, provided the relationship is reducible to linear forms. Some useful transformations to linearise relationships are given below:

function	transformation	form of linear regression equation
$y = ce^{bx}$	$y^* = \ln y$	$y^* = a + bx$
$y = cx^d$	$y^* = \log y, x^* = \log x$	$y^* = a + bx^*$
$y = a + \frac{b}{x}$	$x^* = \frac{1}{x}$	$y = a + bx^*$
$y = \frac{x}{c + dx}$	$y^* = \frac{1}{y}, x^* = \frac{1}{x}$	$y^* = a + bx^*$

- The standard error of estimate which measures the scatter of the observed values around the regression line is given by

$$s_e = \left[\frac{\sum y^2 - a\sum y - b\sum xy}{n-2} \right]^{1/2}$$

- Errors in measurement can be classified into two categories - systematic errors and random errors. Systematic errors arise from imperfection in measuring instruments, change in ambient conditions or the physical limitations of an observer. Systematic errors can be eliminated, if discovered, by careful experimentation. Random errors arise due to unknown reasons which cannot be controlled by the experiments. It is possible to estimate random en-ors statistically. The underlying assumption in random error estimation is that most experimental observations obey the Gaussian distribution.

- The standard error in a set of n measured values of a single variable is given by

$$\sigma = \left[\frac{1}{n-1} \sum d_i^2 \right]^{1/2}$$

where $d_i = x_i - \bar{x}$ is the residual.

Here $\bar{x}$ is the mean value of the measured variable. It is usual to quote the best value of the result as

$$x = \bar{x} \pm \sigma_m$$

where $\sigma_m = \sigma/\sqrt{n}$ is the standard error in the mean.

The standard error in the mean for a quantity Q which is a function of n directly measured quantities ($u_1, u_2, \ldots, u_n$) is given by

$$\sigma_Q = \left[\sum \left(\frac{\partial Q}{\partial u_i} \right)^2 \sigma_{u_i}^2 \right]^{1/2}$$

where $\sigma_{\bar{u}_i}$ is the standard error in the mean of u_i. The best value of Q is written as

$$Q = \bar{Q} \pm \sigma_Q$$

7.6 TERMINAL QUESTIONS

Spend 30 min

1. In an experiment to determine the Young's modulus of a steel bar by the optical lever method, the following readings were recorded

Load x (g)	0	10	20	30	40	50	60	70	80	90	100	120
Shift (cm)	0	0.55	1.20	1.95	2.60	3.30	4.00	4.75	5.45	6.05	6.75	8.10

Obtain the correlation coefficient, the regression equation and the standard error of estimate for the data.

2. The refractive index of a glass prism is given by

$$\mu = \frac{\sin\frac{A+D}{2}}{\sin\frac{A}{2}}$$

where A is the angle of the prism and D is the angle of minimum deviation. The results of several measurements of A and D for sodium light are tabulated below.

$2A$	120° 10′	119° 58′	120° 20′	120° 12′	119° 56′
D	48° 21′	48° 18′	48° 19′	48° 20′	48° 22′

Calculate the best value of μ for sodium light.

7.7 SOLUTIONS AND ANSWERS

SAQs (Self-assessment Questions)

1)

T (°C)	Y (dynes cm^{-1})	T^2 (°C)^2	Y^2 (dynes^2 cm^{-2})	TY (dynes cm^{-1} °C)
10	74.0	100	5476.0	740.0
20	73.0	400	5329.0	1460.0
30	71.0	900	5041.0	2130.0
40	70.0	1600	4900.0	2800.0
50	68.0	2500	4624.0	3400.0
60	66.0	3600	4356.0	3960.0
$\sum T = 210$	$\sum Y = 422$	$\sum T^2 = 9100$	$\sum Y^2 = 29726.0$	$\sum TY = 14490.0$

With the above results we can compute r_{XY} for the given data:

$$r_{XY} = \frac{6(14490.0) - (210)(422)}{\sqrt{6(9100) - (210)^2}\sqrt{6(29726.0) - (422)^2}}$$

$$= \frac{-1680}{(102.5)(16.5)}$$

$$= -0.99$$

2) The regression equation is

$$\hat{Y} = a + bT$$

where $b = \frac{6(14490.0) - (210)(422)}{6(9100) - (210)^2}$

$$= \frac{-1680}{10500} = -0.16$$

and $a = \frac{1}{6}[422 + (0.16)(210)] = 75.9$

Thus, the equation that fits the given data is

$$\hat{Y} = 75.9 - 0.16\,T$$

3. Let us draw up the table for the calculations and use Eq. (7.3b) to estimate the slope which is nothing but the regression coefficient b.

M (kg)	y (μm)	My (kg μm)	M^2 (kg^2)	y^2 (μm)^2
0	1600	0	0	2560000
1	1300	1300	1	1690000
2	950	1900	4	902500
3	600	1800	9	360000
4	250	1000	16	62500
$\sum M = 10$	$\sum y = 4700$	$\sum My = 6000$	$\sum M^2 = 30$	$\sum y^2 = 5575000$

The slope of y vs. M line is

$$b = \frac{5 \times 6000 - 47000}{5 \times 30 - (10)^2} = -\frac{17000}{50} = -340 \ \mu \text{m kg}^{-1}$$

The standard error of estimate in the value of y is given by Eq. (7.4b). For calculating s_e we also require the value of the regression coefficient a, which is given by Eq. (7.3b)

$$a = \frac{1}{5}(4700 + 340 \times 10) = 1620$$

The standard error of estimate is given from

$$s_e^2 = \frac{1}{3}[5575000 - (1620)(4700) + (340)(6000)] = 333.3$$

or $$s_e = 18$$

4. We will calculate the mean of the sample and the residual of each measurement, and use Eqs. (7.14) and (7.15)

N_i	$d_i = N_i - \bar{N}$	d_i^2
36076	126	15876
35753	−197	38809
36116	166	27556
35884	−66	4356
36136	186	34596
35741	−209	43681
35640	−310	96100
36124	174	30276
36076	126	15876
$\bar{N} = 35950$		$\sum d_i^2 = 307126$

$$\therefore \ \sigma_m^2 = \frac{1}{72} \times 307126 = 4266$$

and the standard error of the mean is

$$\sigma_m = 65$$

Thus the best value of the number of a –particles emitted per minute is

$$N = (35950 \pm 65)$$

5. The mean value of n is

$$\bar{n} = \frac{2\bar{L}}{\pi \bar{r}^4 (\bar{\theta}/\bar{N})} = \frac{2 \ x \ 0.5 \ m}{\pi \times (10^{-3} \text{ m})^4 \times 4 \text{ rad N}^{-1} \text{ m}^{-1}}$$

or $\bar{n} = 8.0 \times 10^{10} \text{ Nm}^{-2}$

We now have to obtain the values of $\frac{\partial n}{\partial L}$, $\frac{\partial n}{\partial r}$ and $\partial(\theta/N)$ at the mean values $\bar{L}, \bar{r}, \bar{\theta}/\bar{N}$:

$$\frac{\partial n}{\partial L} = \frac{2}{\pi r^4 (\theta/N)} = 1.6 \times 10^{11}$$

$$\frac{\partial n}{\partial r} = -\frac{8L}{\pi r^5 (\theta/N)} = -3.2 \times 10^{14}$$

$$\frac{\partial n}{\partial (\theta/N)} = -\frac{2}{\pi r^4 (\theta/N)^2} = -2 \times 10^{10}$$

$$\therefore \sigma_n^2 = (1.6 \times 10^{11})^2 \times (2 \times 10^{-3})^2 + (-3.2 \times 10^{14})^2 \times (0.02 \times 10^{-3})^2 + (-2 \times 10^{10})^2 \times (0.1)^2 = 4.5 \times 10^{19}$$

$$\therefore \sigma_n = 7.0 \times 10^9$$

The best value of n is $(8.0 \pm 0.7) \times 10^{10}$ N m^{-2}.

Terminal Questions

1. Let us construct the relevant table for the calculations.

Load x (g)	Shift y (cm)	x^2	xy	y^2
0	0	0	0	0
10	0.55	100	5.50	0.3025
20	1.20	400	24.00	1.4400
30	1.95	900	58.50	3.8025
40	2.60	1600	104.00	6.7600
50	3.30	2500	165.00	10.8900
60	4.00	3600	240.00	16.0000
70	4.75	4900	332.50	22.5625
80	5.45	6400	436.00	29.7025
90	6.05	8100	544.50	36.6025
100	6.75	10000	675.00	45.5625
120	8.10	14400	972.00	65.6100
$\sum x = 670$	$\sum y = 44.7$	$\sum x^2 = 52900$	$\sum xy = 3557.00$	$\sum y^2 = 239.235$

The correlation coefficient $r_{xy} = \frac{12(3557.00) - (670)(44.7)}{\sqrt{12(52900) - (670)^2}\ \sqrt{12(239.235) - (44.7)^2}}$

$$= \frac{12735}{\sqrt{185900}\ \sqrt{872.73}} = \frac{12735}{12736.5} = 0.9999$$

This indicates a very high linear correlation. The linear regression coefficients are

$$b = \frac{12(3557.00) - (670)(44.7)}{12(52900) - (670)^2} = \frac{12735}{185900} = 0.0685$$

$$a = \frac{1}{12}[44.7 - 0.0685(670)] = -0.0998$$

The regression equation for the data is

$$\hat{y} = -0.0998 + 0.0685x$$

The standard error of estimate is

$$s_e = \left[\frac{239.235 + 0.0998\,(44.7) - 0.0685\,(3557.0)}{10}\right]^{1/2} = 0.06$$

2. We have to first determine the standard errors in the mean for the measurements of A and L.

Standard error in the mean for A

$2A_i$	A_i	$d_i = A_i - \overline{A}$	d_i^2 in $(')^2$
120° 10′	60° 5′	0° 1′	1
119° 58′	59° 59′	−0° 5′	25
120° 20′	60° 10′	0° 6′	36
120 12′	60° 6′	0° 2′	4
119 56′	59° 58′	−0° 6′	36
	$\overline{A}$ = 60° 4′		$\sum d_i^2 = 102$

$$\sigma_{\overline{A}}^2 = \frac{1}{20} \times 102 = 5 \quad \text{and} \quad \sigma_{\overline{A}} = 0°\,2'$$

Standard error in the mean for D

D_i	$d_i = D_i - \overline{D}$	d_i^2 in $(')^2$
48° 21′	0° 1′	1
48° 18′	−0° 2′	4
48° 19′	−0° 1′	1
48° 20′	0° 0′	0
48° 22′	0° 2′	4
$\overline{D}$ = 48° 20′		$\sum d_i^2 = 10$

$$\sigma_{\overline{D}}^2 = \frac{1}{20} \times 10 = 0.5 \quad \text{and} \quad \sigma_{\overline{D}} = 0°\,1'$$

Now the best value of μ is given by

$$\mu = \mu \pm \sigma_{\overline{\mu}}$$

where $\sigma_{\overline{\mu}}$ is to be computed using Eq. (7.21) and

$$\overline{\mu} = \frac{\sin\dfrac{\overline{A} + \overline{D}}{2}}{\sin\dfrac{\overline{A}}{2}} = \frac{\sin\dfrac{60° 4' + 48° 20'}{2}}{\sin\,(30''\,2')}$$

or $$\overline{\mu} = \frac{\sin\,(54°\,12')}{\sin\,(30°\,2')} = 1.6205$$

From Eq.(7.21)

$$\sigma_{\overline{\mu}}^2 = \left(\frac{\partial \mu}{\partial A}\right)^2 \sigma_{\overline{A}}^2 + \left(\frac{\partial \mu}{\partial D}\right)^2 \sigma_{\overline{D}}^2$$

$$\frac{\partial \mu}{\partial A} = \frac{1}{2}\,\frac{\cos\left(\dfrac{A+D}{2}\right)}{\sin\left(\dfrac{A}{2}\right)} - \frac{1}{2}\,\frac{\sin\left(\dfrac{A+D}{2}\right)}{\sin\left(\dfrac{A}{2}\right)\tan\left(\dfrac{A}{2}\right)}$$

and

$$\frac{\partial \mu}{\partial D} = \frac{1}{2}\,\frac{\cos\left(\dfrac{A+D}{2}\right)}{\sin\left(\dfrac{A}{2}\right)}$$

The values of $\frac{\partial \mu}{\partial A}$ and $\frac{\partial \mu}{\partial D}$ have to be evaluated at $A = \overline{A}$ and $D = \overline{D}$:

$$\frac{\partial \mu}{\partial A} = \frac{1}{2}\frac{\cos(54^\circ 12')}{\sin(30^\circ 2')} - \frac{1}{2}\frac{\sin(54^\circ 12')}{\sin(30^\circ 2')\tan(30^\circ 2')}$$

$$= -0.80$$

$$\frac{\partial \mu}{\partial D} = \frac{1}{2}\frac{\cos(54^\circ 12')}{\sin(30^\circ 2')} = 0.58$$

For the remaining calculations, we have to convert to radian units in which

$$\sigma_{\overline{A}} = 0.00058 \text{ rad} \quad \text{and} \quad \sigma_{\overline{D}} = 0.00029 \text{ rad}$$

$$\therefore \sigma_{\mu}^2 = (0.64)(0.00058)^2 + (0.34)(0.00029)^2$$

$$\therefore \sigma_{\overline{\mu}} = 0.0005$$

Thus, the best value of μ is (1.6205 ± 0.0005).