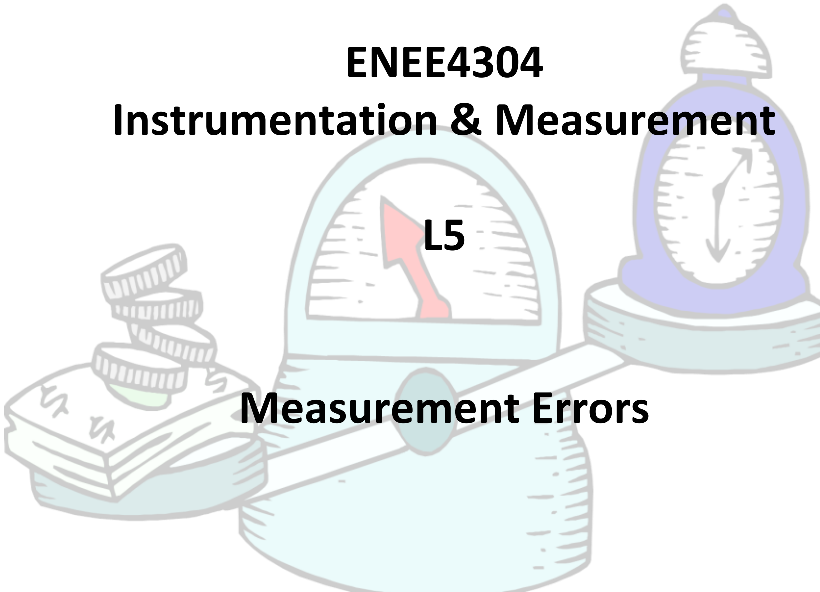
ENEE4304
Instrumentation & Measurement

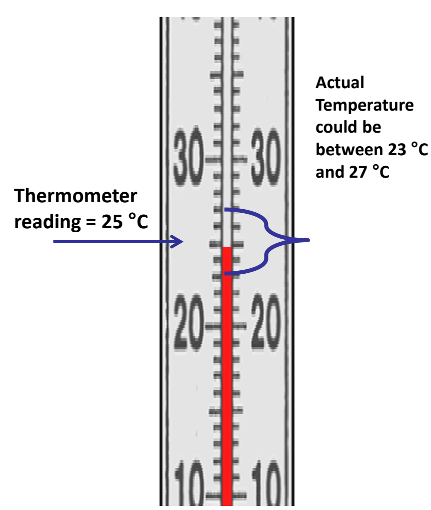


L5

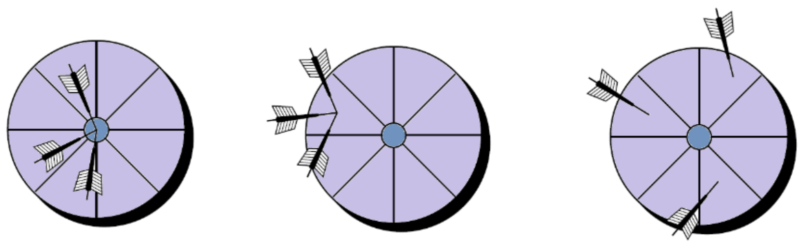
Measurement Errors

Systematic and Random Errors

- Errors are effects that cause a measured value to differ from its true value.
- Systematic error causes an offset between the mean value of the data set and its true value.
- Random error causes a random variation in measured values found during repeated measurements of a variable.
- Both random and systematic errors affect a system's accuracy.

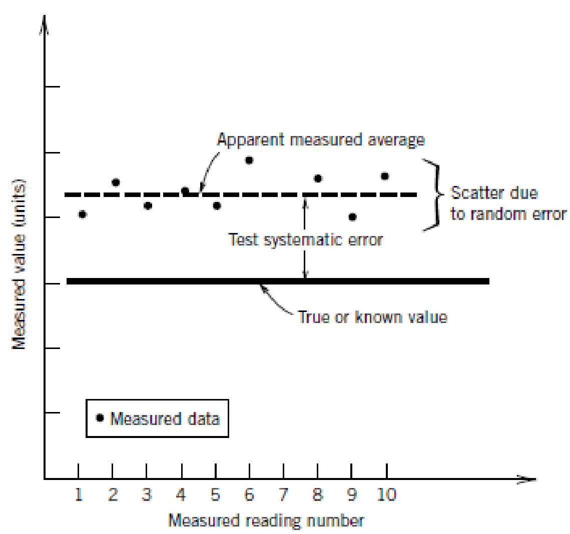


Systematic and Random Errors



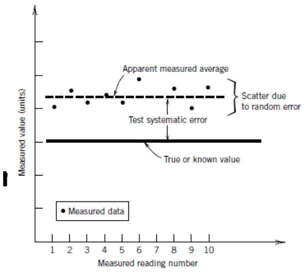
Low random error **Low systematic errors**	**Low random error** **High systematic error**	**High random error,** **High systematic error**
High Precision **High accuracy**	**High precision** **Low accuracy**	**Low precision** **Low accuracy**

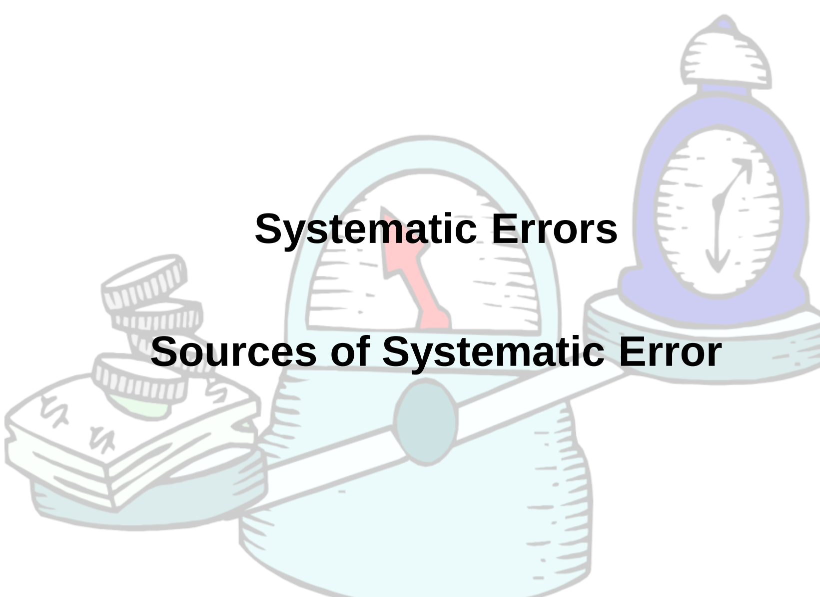
Systematic and Random Errors



Random Errors

- Random errors in measurements are caused by unpredictable variations in the measurement system.
- **They are usually observed as small perturbations of the measurement to either side of the correct value**, i.e. positive errors and negative errors occur in approximately equal numbers for a series of measurements made of the same constant quantity.
- Therefore, random errors can largely be eliminated by calculating the average of a number of repeated measurements, provided that the measured quantity remains constant during the process of taking the repeated measurements.





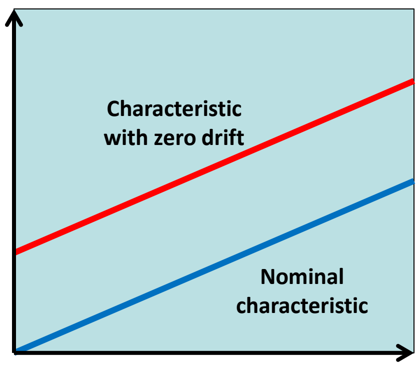
Systematic Errors

Sources of Systematic Error

Sources of Systematic Error: 1) Zero Drift

- Zero drift (zero offset) (or bias) causes a constant error over the full range of measurement. Zero drift is normally removable by calibration.

A scale giving a reading when no mass is placed has a zero drift



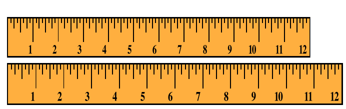
Sources of Systematic Error: 2) Scale Error

- Scale error produces an error that is a percentage of the measured quantity

$e = kx$

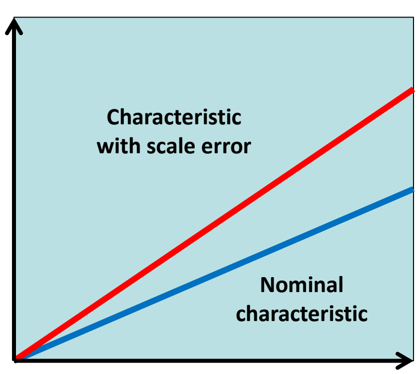
e : Scale Error

x : Measured Variable

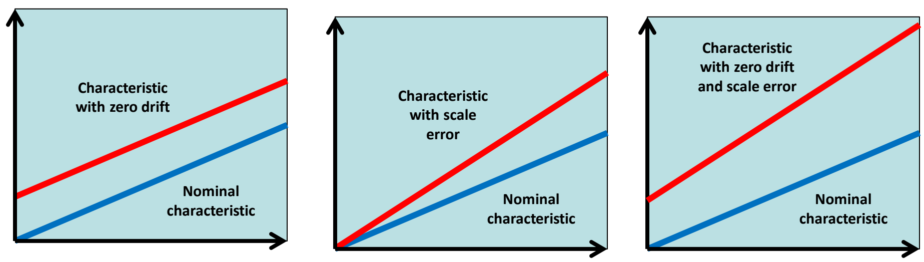


Expansion or contraction of the ruler due to temperature

Hardening of softening of the spring used in the scale

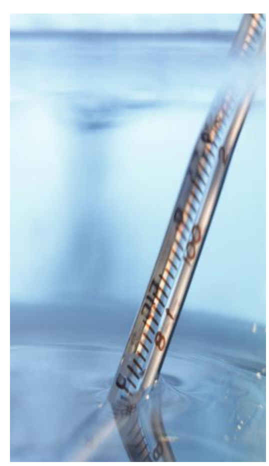


Combined Zero Drift and Scale Error



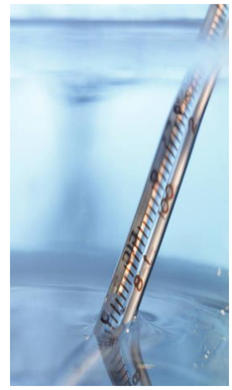
Sources of Systematic Error:
3) System disturbance due to measurement

- Disturbance of the measured system by the act of measurement is a common source of systematic error.
- A mercury-in-glass thermometer, initially at room temperature, and used to measure the temperature of a hot water beaker, would introduce a disturbance (heat capacity of the thermometer) into the hot water and lower the temperature of the water.
- In nearly all measurement situations, the process of measurement disturbs the system and alters the values of the physical quantities being measured.



Example 1: System Disturbance in temperature Measurement

- A liquid in glass is used to measure the temperature of a water in a hot water beaker.
- Determine the parameters affecting the disturbance error. Estimate the disturbance error in term of these parameters.
- Hint: Use an energy balance on the overall thermometer-beaker system

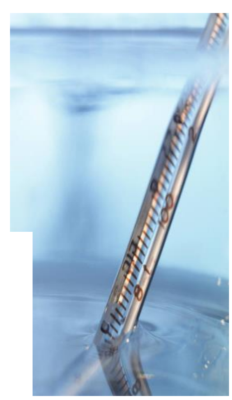


Example 1: Solution

Heat loss from the water = Heat gain by the thermometer

$$C_w(T_a - T_m) = C_t(T_m - T_r)$$

$$(T_a - T_m) = \frac{C_t}{C_w}(T_m - T_r)$$



T_a :Actual temperature of the water in the beaker

T_m :Measured temperature of the water in the beaker

T_r : Initial temperature of the thermometer (room temperature.

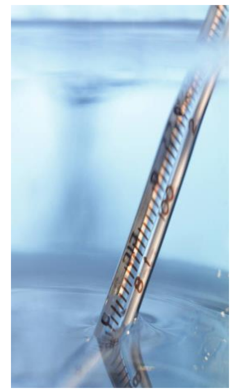
C_t : Heat capacity of the thermometer

C_W:Heat capacity of the water in the beaker = $m_w c_w$

Sources of Systematic Error:

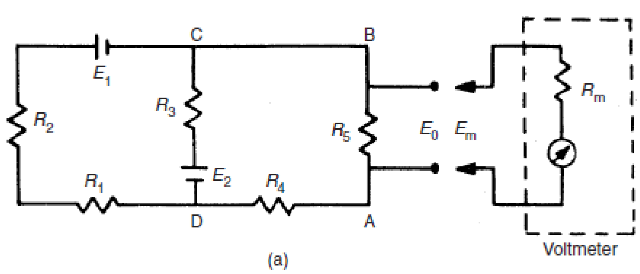
- If we consider the thermometer to be measuring the amount by which the temperature in the beaker deviates from the current temperature of the thermometer.
- then the disturbance error can be considered as a scale error

$$\underbrace{(T_a - T_m)}_{\text{Scale Error}} = \underbrace{\frac{C_t}{C_w}}_{\text{Scale factor}} \underbrace{(T_m - T_r)}_{\text{Measured Quantity}}$$



Example 2: System Disturbance in electric circuit measurements

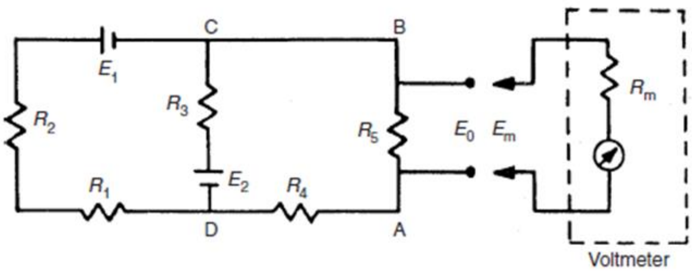
- In the circuit shown, the voltage across resistor R_5 is to be measured by a voltmeter with resistance R_m. Estimate the disturbance error in the measured voltage in term of R_m.



(a)

Example 2: Solution

- Here, R_m acts as a shunt resistance across R_5, decreasing the resistance between points AB and so disturbing the circuit.
- Therefore, the voltage E_m measured by the meter is not the value of the voltage E_o that existed prior to measurement.
- The extent of the disturbance can be assessed by calculating the open circuit voltage E_o and comparing it with E_m.



Example 2: Solution

$$\frac{1}{R_{CD}} = \frac{1}{R_1 + R_2} + \frac{1}{R_3}$$

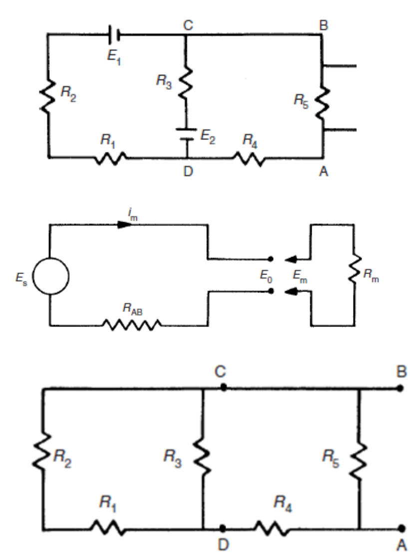
or

$$R_{CD} = \frac{(R_1 + R_2)R_3}{R_1 + R_2 + R_3}$$

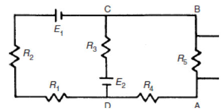
$$\frac{1}{R_{AB}} = \frac{1}{R_{CD} + R_4} + \frac{1}{R_5}$$

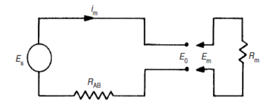
or

$$R_{AB} = \frac{(R_{CD} + R_4)R_5}{R_{CD} + R_4 + R_5}$$



Example 2: Solution



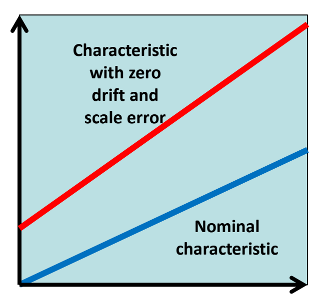


$$E_m = R_m I = \frac{R_m E_o}{R_{AB} + R_m}$$

- It is thus obvious that as R_m gets larger, the ratio E_m/E_0 gets closer to unity, showing that the design strategy should be to make R_m as high as possible to minimize disturbance of the measured system.
- Disturbance errors are usually present in passive instruments where energy needs to be withdrawn from the system in the measurement process.
- It is often the reason for the use of alternative active instruments such as digital voltmeters, where the inclusion of auxiliary power greatly improves performance.

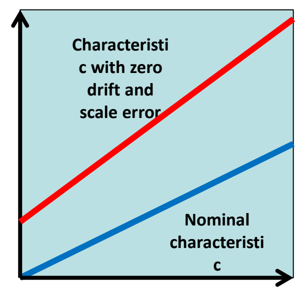
4) Errors due to environmental inputs

- An environmental input is defined as an apparently real input to a measurement system that is actually caused by a change in the environmental conditions surrounding the measurement system.
- The magnitude of environment-induced variation is quantified by the two constants known as sensitivity drift and zero drift, both of which are generally included in the published specifications for an instrument.



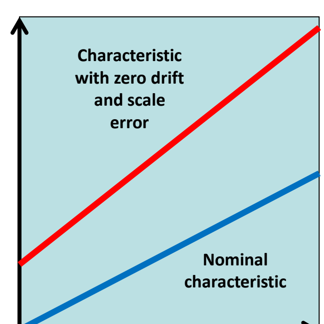
Errors due to environmental inputs

- Variations of environmental conditions away from the calibration conditions are sometimes described as modifying inputs to the measurement system.
- When such modifying inputs are present, it is often difficult to determine how much of the output change in a measurement system is due to a change in the measured variable and how much is due to a change in environmental conditions.



Errors due to environmental inputs

- In any general measurement situation, it is very difficult to avoid environmental inputs, because it is either impractical or impossible to control the environmental conditions surrounding the measurement system.
- System designers are therefore charged with the task of either reducing the susceptibility of measuring instruments to environmental inputs or, alternatively, quantifying the effect of environmental inputs and correcting for them in the instrument output reading.



5) Wear in instrument components

- ❑ Systematic errors can frequently develop over a period of time because of wear in instrument components.
- ❑ Recalibration often provides a full solution to this problem.

Hardening of softening of the spring used in the scale

6) Connecting leads

- ❑ In connecting together the components of a measurement system, a common source of error is the failure to take proper account of the resistance of connecting leads (or pipes in the case of pneumatically or hydraulically actuated measurement systems).
- ❑ For instance, in typical applications of a resistance thermometer, it is common to find that the thermometer is separated from other parts of the measurement system by perhaps 100 metres.
- ❑ The resistance of such a length of 20 gauge copper wire is 7Ω, and there is a further complication that such wire has a temperature coefficient of 1mΩ /°C.
- ❑ Therefore, careful consideration needs to be given to the choice of connecting leads.

Reduction of systematic errors

- The prerequisite for the reduction of systematic errors is a complete analysis of the measurement system that identifies all sources of error.
- Simple faults within a system, such as bent meter needles and poor cabling practices, can usually be readily and cheaply rectified once they have been identified.
- However, other error sources require more detailed analysis and treatment.
- Various approaches to error reduction are considered in the next slides.

Reduction of systematic errors

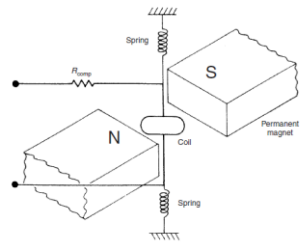
1) Careful instrument design

- Careful instrument design is the most useful method in dealing with environmental inputs.
- This aims at reducing the sensitivity of an instrument to environmental inputs to as low a level as possible.
- For instance, in the design of strain gauges, the element should be constructed from a material whose resistance has a very low temperature coefficient (i.e. the variation of the resistance with temperature is very small).
- However, errors due to the way in which an instrument is designed are not always easy to correct, and a choice often has to be made between the high cost of redesign and the alternative of accepting the reduced measurement accuracy if redesign is not undertaken.

Reduction of systematic errors

2) Method of opposing inputs

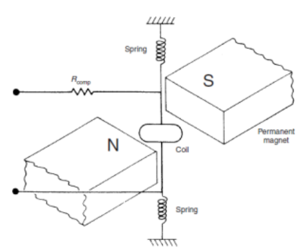
- ❑ The method of opposing inputs compensates for the effect of an environmental input in a measurement system by introducing an equal and opposite environmental input that cancels it out.



- ❑ One example of how this technique is applied is in the type of millivolt meter shown.
- ❑ It consists of a coil suspended in a fixed magnetic field produced by a permanent magnet.
- ❑ When an unknown voltage is applied to the coil, the magnetic field due to the current interacts with the fixed field and causes the coil (and a pointer attached to the coil) to turn.

Method of opposing inputs

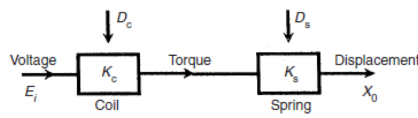
- ❑ If the coil resistance R_{coil} is sensitive to temperature, then any temperature change in the environment will alter the value of the coil current for a given applied voltage and so alter the pointer output reading.



- ❑ Compensation for this is made by introducing a compensating resistance R_{comp} into the circuit, where R_{comp} has a temperature coefficient that is equal in magnitude but opposite in sign to that of the coil.
- ❑ Thus, the total resistance remains approximately the same in response to a temperature change.

Reduction of systematic errors

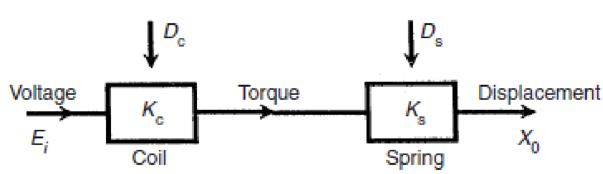
3)High-gain feedback



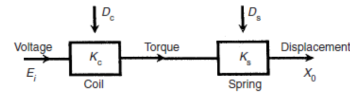
- The benefit of adding high-gain feedback to many measurement systems is illustrated by considering the case block diagram of the millivoltmeter shown.
- In this system, the unknown voltage E_i is applied to a coil of torque constant K_c, and the induced torque turns a pointer against the restraining action of a spring with spring constant K_s.
- The effect of environmental disturbance on the motor and spring constants is represented by variables D_c and D_s.

High-gain feedback

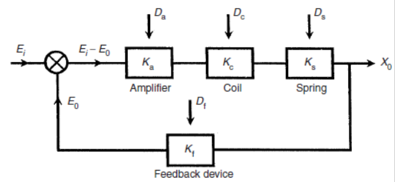
- In the absence of environmental inputs, the displacement of the pointer X_0 is given by: $X_0 = K_c K_s E_i$.
- However, in the presence of environmental inputs, both K_c and Ks change, and the relationship between X_0 and E_i can be affected greatly. Therefore, it becomes difficult or impossible to calculate E_i from the measured value of X_0 .



High-gain feedback



- Consider now what happens if the system is converted into a high-gain, closed-loop one, as shown in the, by adding an amplifier of gain constant Ka and a feedback device with gain constant Kf.



- Assume also that the effect of environmental inputs on the values of and Kf are represented by Da and Df.
- The feedback device feeds back a voltage E_0 proportional to the pointer displacement X_0.
- This is compared with the unknown voltage Ei by a comparator and the error is amplified.

- Writing down the equations of the system, we have:

$$E_o = K_f X_o$$

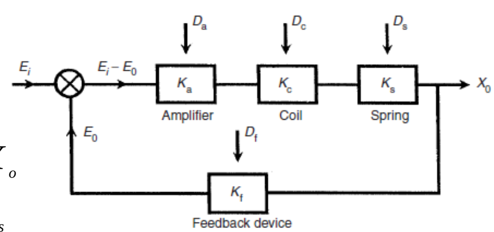
$$X_o = (E_i - E_o)K_a K_c K_s$$

$$X_o = (E_i - K_f X_o)K_a K_c K_s$$

$$X_o = E_i K_a K_c K_s - K_f K_a K_c K_s X_o$$

$$X_o(1 + K_f K_a K_c K_s) = E_i K_a K_c K_s$$

$$X_o = \frac{K_a K_c K_s}{(1 + K_f K_a K_c K_s)} E_i$$



If Ka is made very large (it is a high-gain amplifier),

$$K_f K_a K_c K_s >> 1$$

$$X_o \cong \frac{1}{K_f} E_i$$

High-gain feedback

$$X_o \cong \frac{1}{K_f} E_i$$

- This important result shows that the relationship between the output, X_0 , and the input, E_i , has been reduced to one that involves only K_f.
- The sensitivity of the gain constants K_a, K_c and K_s to the environmental inputs D_a, D_m and D_s has thereby been rendered irrelevant, and we only have to be concerned with one environmental input D_f.

High-gain feedback

$$X_o \cong \frac{1}{K_f} E_i$$

- It is usually easy to design a feedback device that is insensitive to environmental inputs: this is much easier than trying to make a coil or spring insensitive.
- Thus, high gain feedback techniques are often a very effective way of reducing measurement system's sensitivity to environmental inputs.

- One potential problem, however, is that there is a possibility that high-gain feedback will cause instability in the system.
- Therefore, any application of this method must include careful stability analysis of the system.

Reduction of systematic errors
4)Intelligent Instruments

- Intelligent instruments contain extra sensors that measure the value of environmental inputs and automatically compensate the value of the output reading.

- They have the ability to deal very effectively with systematic errors in measurement systems, and errors can be attenuated to very low levels in many cases

Reduction of systematic errors
5)Calibration

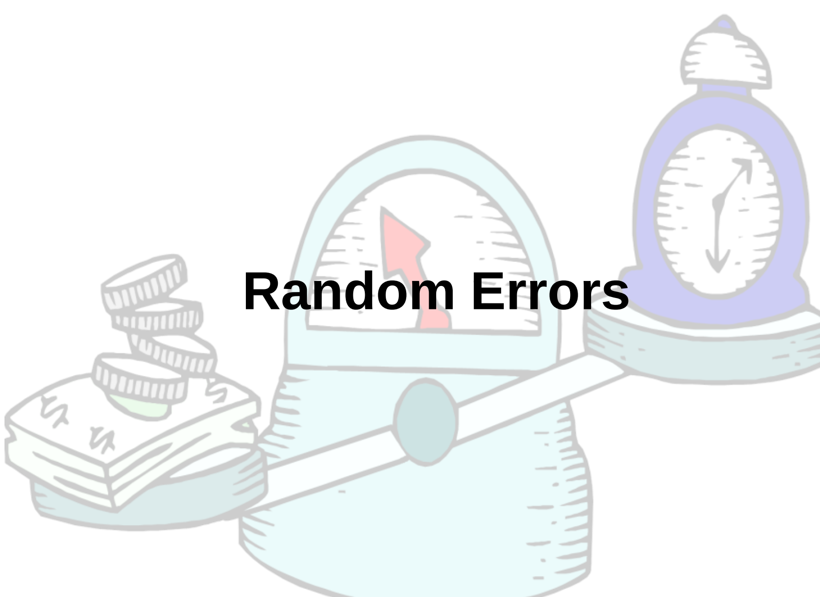
- Instrument calibration is a very important consideration in measurement systems as all instruments suffer drift in their characteristics, and the rate at which this happens depends on many factors, including environmental conditions in which instruments are used and the frequency of their use.

- Thus, errors due to instruments being out of calibration can usually be rectified by increasing the frequency of recalibration.

Quantification of systematic errors

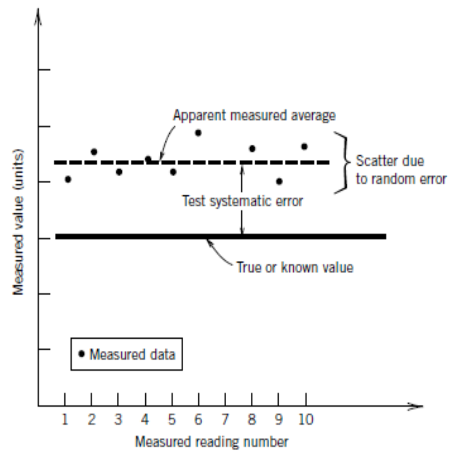
- Once all practical steps have been taken to eliminate or reduce the magnitude of systematic errors, the final action required is to estimate the maximum remaining error that may exist in a measurement due to systematic errors.
- The usual course of action is to assume mid-point environmental conditions and specify the maximum measurement error as ±x% of the output reading to allow for the maximum expected deviation in environmental conditions away from this mid-point.
- Data sheets supplied by instrument manufacturers usually quantify systematic errors in this way, and such figures take account of all systematic errors that may be present in output readings from the instrument.

Random Errors



Statistical Analysis of Measurements Subject to Random Errors

- The average value of a set of measurements of a constant quantity can be expressed as either the mean value or the median value
- The degree of confidence in the calculated ***mean/median*** values can be quantified by calculating the standard deviation or variance of the data.



Mean and Median Values

- As the number of measurements increases, the difference between the mean value and median values becomes very small.
- For any set of n measurements, $x_1, x_2, \ldots, x_n$ of a constant quantity, the mean given by:

$$x_{mean} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum x_i$$

- When the measurement errors are distributed equally about the zero error value for a set of measurements, the most likely true value is the mean value.

Mean and Median Values

- The median is an approximation to the mean and it is the middle value when the measurements in the data set are written in ascending order of magnitude.
- For a set of n measurements, $x_1, x_2, \dots, x_n$ of a constant quantity, written down in ascending order of magnitude, the median value is given by:

$$x_{median} = \begin{cases} x_{(n+1)/2} & n \text{ is odd} \\ \dfrac{x_{n/2} + x_{(n+2)/2}}{2} & n \text{ is even} \end{cases}$$

- Thus, for a set of 9 measurements $x_1, x_2, \dots, x_9$ arranged in order of magnitude, the median value is x_5. For an even number of measurements, the median value is midway between the two centre values, i.e. for 10 measurements $x_1, x_2, \dots, x_{10}$, the median value is given by: $(x_5+x_6)/2$

Example 3

- The length of a steel bar is measured by two sets of observers and the following two sets measurements were recorded (units mm). Find the mean and median for each data set.

Measurement set A, 11 observers

398 420 394 416 404 408 400 420 396 413 430

Measurement set B, 14 observers

409 406 402 407 405 404 407 404 407 407 408 405 412

Example 3 Solution

Measurement set A, 11 observers

398 420 394 416 404 408 400 420 396 413 430

Write in ascending order

394 396 398 400 404 408 413 416 420 420 430

Mean = 409

Median = 408

Measurement set B, 14 observers

409 406 402 407 405 404 407 404 407 407 408 405 412 403

Write in ascending order

402 403 404 404 405 405 406 407 407 407 407 408 409 412

Mean = 406.1429

Median = 406.5

Note that as the number of observer increases, the median normally gets closer to the mean.

Confidence in Mean and Median Values

- Which of the two measurement sets A and B, and the corresponding mean and median values, should we have most confidence in?

  Measurement set A, 11 observers written in ascending order

  394 396 398 400 404 408 413 416 420 420 430

  Measurement set B, 14 observers written in ascending order

  402 403 404 404 405 405 406 407 407 407 407 408 409 412

- Intuitively, we can regard measurement set B as being more reliable since the measurements are much closer together. In set A, the spread between the smallest (394) and largest (430) value is 36, whilst in set B, the spread is only 10.
- **Thus, the smaller the spread of the measurements, the more confidence we have in the mean or median value calculated.**

Standard Deviation and Variance

- Instead of expressing the spread of measurements simply as the difference between the largest and smallest value, a much better way of is to calculate the variance or standard deviation of the measurements. We start by calculating the deviation (error) d_i of each measurement x_i from the mean value x_{mean}

$$d_i = x_i - x_{mean}$$

The variance *V* is then given by:

$$V = \frac{d_1^2 + d_2^2 + \cdots + d_n^2}{n-1} = \frac{1}{n-1}\sum (x_i - x_{mean})^2$$

and the standard deviation

$$\sigma = \sqrt{V} = \sqrt{\frac{d_1^2 + d_2^2 + \cdots + d_n^2}{n-1}} = \sqrt{\frac{1}{n-1}\sum (x_i - x_{mean})^2}$$

Example 4

- Calculate the variance *V* and the standard deviation σ for the data sets A and B of example 3.

Measurement set A, 11 observers

398 420 394 416 404 408 400 420 396 413 430

Measurement set B, 14 observers

402 403 404 404 405 405 406 407 407 407 407 408 409 412

Example 4. Solution

- As V and σ decrease for a measurement set, we are able to express greater confidence that the calculated mean or median value is close to the true value,
- i.e. that the averaging process has reduced the random error value close to zero.

Data Set A

Xi	Xi-Xmean
398	-11
420	11
394	-15
416	7
404	-5
408	-1
400	-9
420	11
396	-13
413	4
430	21

variance	137
stdev	11.7047

Example 4. Solution

- V and σ normally get smaller as the number of measurements increases, confirming that confidence in the mean value increases as the number of measurements increases.

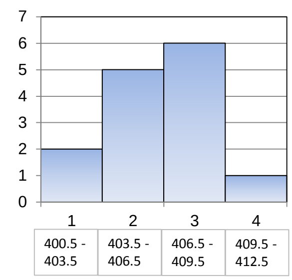
Data Set B

Xi	Xi-Xmean
402	-4.14286
403	-3.14286
404	-2.14286
404	-2.14286
405	-1.14286
405	-1.14286
406	-0.14286
407	0.857143
407	0.857143
407	0.857143
407	0.857143
408	1.857143
409	2.857143
412	5.857143

variance	6.747253
stdev	2.597547

Graphical data analysis techniques – frequency distributions

- Graphical techniques are a very useful way of analyzing the way in which random measurement errors are distributed. The simplest way of doing this is to draw a histogram, in which bands of equal width across the range of measurement values are defined and the number of measurements within each band is counted.

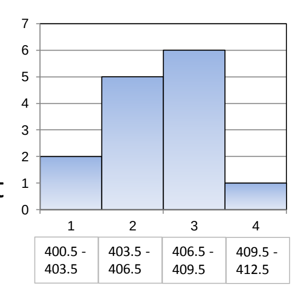


- The figure shows a histogram for set B of the length measurement data given in example 3, in which the bands chosen are 3mm wide.

402 403 404 404 405 405 406 407 407 407 407 408 409 412

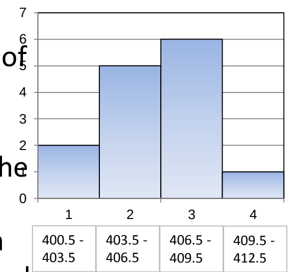
Graphical data analysis techniques – frequency distributions

- For instance, there are 6 measurements in the range between 406.5 and 409.5 and so the height of the histogram for this range is 6 units.
- Also, there are 5 measurements in the range from 403.5 to 406.5 and so the height of the histogram over this range is 5 units.
- The rest of the histogram is completed in a similar fashion.
- The scaling of the bands was deliberately chosen so that no measurements fell on the boundary between different bands and caused ambiguity about which band to put them in.



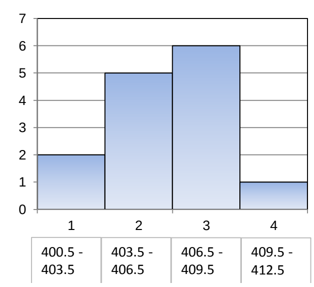
Graphical data analysis techniques – frequency distributions

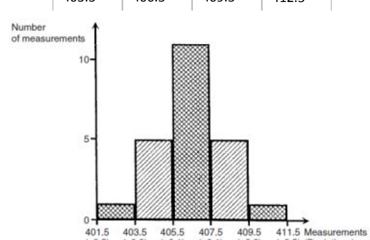
- It is often useful to draw a histogram of the deviations of the measurements from the mean value rather than to draw a histogram of the measurements themselves.
- The starting point for this is to calculate the deviation of each measurement away from the calculated mean value.
- Then a histogram of deviations can be drawn by defining deviation bands of equal width and counting the number of deviation values in each band.
- This histogram has exactly the same shape as the histogram of the raw measurements except that the scaling of the horizontal axis has to be redefined in terms of the deviation values



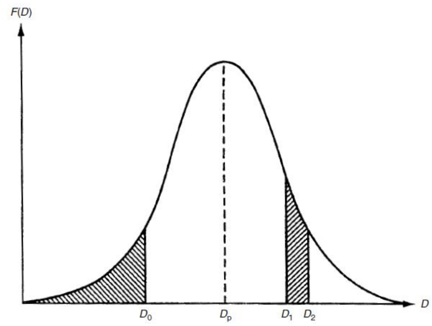
Graphical data analysis techniques – frequency distributions

- As the number of measurements increases, smaller bands can be defined for the histogram, which retains its basic shape but then consists of a larger number of smaller steps on each side of the peak.
- For example, the histogram shown below is for a sample of a total of 23 length measurements of the bar in example 3. The bands chosen are 2mm wide



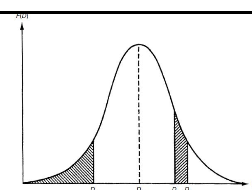


Graphical data analysis techniques – frequency distributions



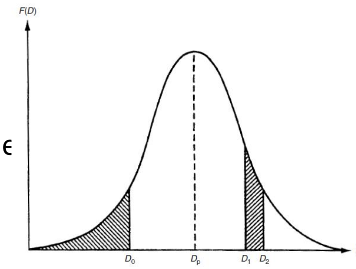
- As the number of measurements approaches infinity, the histogram becomes a smooth curve known as a *frequency distribution curve*.
- The ordinate of this curve is the frequency of occurrence of each deviation value, *F(D)*, and the abscissa is the magnitude of deviation, *D*.

Graphical data analysis techniques – frequency distributions



- The **symmetry** of the figure about the zero deviation value is very useful for showing graphically that the measurement data only has **random errors and are free from systematic error.**
- If the height of the frequency distribution curve is normalized such that the area under it is unity, then the curve in this form is known as a probability curve, and the height F(D) at any particular deviation magnitude D is known as the *probability density function* (p.d.f.).

Graphical data analysis techniques – frequency distributions



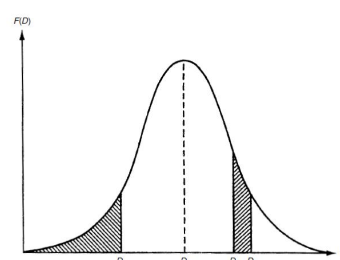
- The condition that the area under the curve is unity can be expressed mathematically as:

$$\int_{-\infty}^{\infty} F(D)dD = 1$$

- The probability that the error in any one particular measurement lies between two levels D_1 and D_2 can be calculated by measuring the area under the curve contained between two vertical lines drawn through D_1 and D_2. This can be expressed mathematically as:

$$P(D_1 \leq D \leq D_2) = \int_{D1}^{D2} F(D)dD$$

Graphical data analysis techniques – frequency distributions

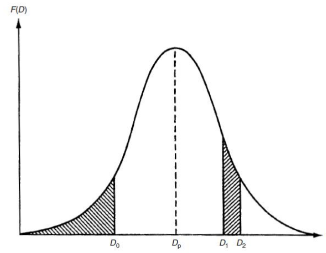


- Of particular importance for assessing the maximum error likely in any one measurement is the *cumulative distribution function* (c.d.f.).
- This is defined as the probability of observing a value less than or equal to D_0, and is expressed mathematically as:

$$P(D \leq D_0) = \int_{-\infty}^{D0} F(D)dD$$

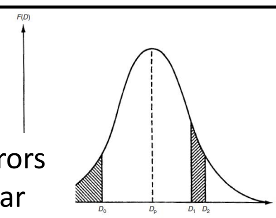
- Thus, the c.d.f. is the area under the curve to the left of a vertical line drawn through D_0, as shown by the left-hand hatched area

Graphical data analysis techniques – frequency distributions



- The deviation magnitude Dp corresponding with the peak of the frequency distribution curve is the value of deviation that has the greatest probability.
- If the errors are entirely random in nature, then the value of Dp will equal zero. Any non-zero value of Dp indicates systematic errors in the data, in the form of a bias that is often removable by recalibration.

Gaussian Distribution



- Measurement sets that only contain random errors usually conform to a distribution with a particular shape that is **called Gaussian**.
- The shape of a Gaussian curve is such that the frequency of small deviations from the mean value is much greater than the frequency of large deviations.
- This coincides with the usual expectation in measurements subject to random errors that the number of measurements with a small error is much larger than the number of measurements with a large error.
- Alternative names for the Gaussian distribution are the Normal distribution or Bell-shaped distribution.

Gaussian Distribution

- A Gaussian curve is formally defined as a normalized frequency distribution that is symmetrical about the line of zero error and in which the frequency and magnitude of quantities are related by the expression:

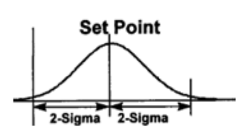
$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\left(-(x-m)^2/2\sigma^2\right)}$$

where *m* is the mean value of the data set *x* and σ is the standard deviation of the set.

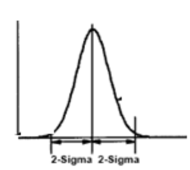
- This equation is particularly useful for analyzing a Gaussian set of measurements and predicting how many measurements lie within some particular defined range.

Gaussian Distribution

- If the measurement deviations *D* are calculated for all measurements such that *D* = *x* - *m*, then the curve of deviation frequency *F(D)* plotted against deviation magnitude *D* is:



$$F(D)=\frac{1}{\sigma\sqrt{2\pi}}e^{\left(-D^2/2\sigma^2\right)}$$



- The shape of a Gaussian curve is strongly influenced by the value of σ, with the width of the curve decreasing as σ becomes smaller.
- A smaller σ corresponds with the typical deviations of the measurements from the mean value becoming smaller.

Gaussian Distribution

- If the standard deviation is used as a unit of error, the Gaussian curve can be used to determine the probability that the deviation in any particular measurement in a Gaussian data set is greater than a certain value. By substituting the expression for F(D) from the previous equation into the probability equation

$$P(D_1 \le D \le D_2) = \int_{D1}^{D2} F(D)dD$$

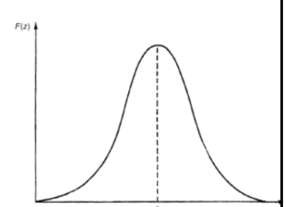
- The probability that the error lies in a band between error levels D1 and D2 can be expressed as:

$$P(D_1 \le D \le D_2) = \int_{D_1}^{D_2} \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-D^2/2\sigma^2\right)} dD$$

Gaussian Distribution

- Solution of this expression is simplified by the substitution:

$$z = D/\sigma$$

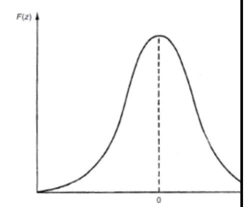


- The effect of this is to change the error distribution curve into a new Gaussian distribution that has a standard deviation of one (σ = 1), and a mean of zero.
- This new form is known as a standard Gaussian curve, and the dependent variable is now z instead of D.
- The equation can now be re-expressed as:

$$P(z_1 \le z \le z_2) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-z^2/2} dz$$

Standard Gaussian tables

- The previous equation can not be integrated analytically and numerical integration provides the only method of solution.

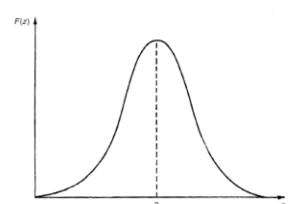


- In practice, the numerical integration can be avoided when analyzing data because the standard form of equation, and its independence from the particular values of the mean and standard deviation of the data, means that standard Gaussian tables that tabulate G(z) for various values of z can be used.

$$P(z_1 \le z \le z_2) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-z^2/2} dz$$

Standard Gaussian tables

- A standard Gaussian table tabulates F(z) for various values of z, is given by:

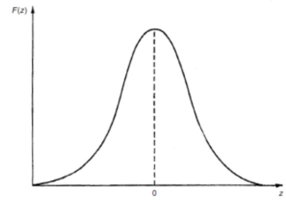


$$G(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{(-z^2/2)} dz$$

- Thus, G(z) gives the proportion of data values that are less than or equal to z.
- This proportion is the area under the curve of F(z) against z that is to the left of z.

Standard Gaussian tables

- To evaluate the probability that the error lies in a band between error levels D1 and D2, the expression has to be evaluated as



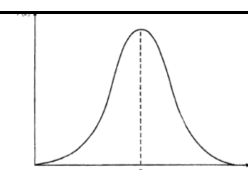
$$P(z_1 \le z \le z_2) = G(z_2) - G(z_1)$$

where

$$z_1 = D_1/\sigma, \quad z_2 = D_2/\sigma$$

- The table shows that G(z) = 0.5 for z = 0. This confirms that, as expected, the number of data values ≤ 0 is 50% of the total.
- This must be so if the data only has random errors. It will also be observed that Gaussian tables only gives G(z) for positive values of z.
- For negative values of z, we can make use of the following relationship: $G(-z) = 1 - G(z)$

Example 5. Solution



- The example above shows that, for Gaussian-distributed data values, 68% of the measurements have deviations that lie within the bounds of ±σ.
- Similar analysis shows that that boundaries of ±2σ contain 95.4% of data points, and extending the boundaries to ±3σ encompasses 99.7% of data points.
- The probability of any data point lying outside particular deviation boundaries can therefore be expressed by the following table.

Deviation boundaries	*% of data points within boundary*	*Probability of any particular data point being outside boundary*
$\pm\sigma$	68.0	32.0%
$\pm 2\sigma$	95.4	4.6%
$\pm 3\sigma$	99.7	0.3%

Example 8

- An integrated circuit chip contains 10^5 transistors. The transistors have a mean current gain of 20 and a standard deviation of 2.
- Calculate the number of transistors with a current gain between 19.8 and 20.2

Example 8. Solution

- An integrated circuit chip contains 10^5 transistors. The transistors have a mean current gain of 20 and a standard deviation of 2. Calculate the number of transistors with a current gain between 19.8 and 20.2

$$z_1 = D_1 / \sigma = -0.2 / 2 = -0.1$$
$$z_2 = D_2 / \sigma = 0.2 / 2 = 0.1$$
$$P(z_1 \le z \le z_2) = G(z_2) - G(z_1)$$
$$P(-0.1 \le z \le 0.1) = G(0.1) - G(-0.1)$$
$$P(-0.1 \le z \le 0.1) = G(0.1) - (1 - G(0.1))$$
$$P(-0.1 \le z \le 0.1) = 2G(0.1) - 1 = 2 \times 0.5398 - 1 = 0.0796$$

- Thus 0.0796 x 10^5 = 7960 transistors have a current gain in the range from 19.8 to 20.2.

Standard error of the mean

- The foregoing analysis has examined the way in which measurements with random errors are distributed about the mean value.
- However, we have already observed that some error remains between the mean value of a set of measurements and the true value, i.e. averaging a number of measurements will *only* yield the true value if the number of measurements is infinite.

Standard error of the mean

- If several subsets are taken from an infinite data population, then, by the central limit theorem, the means of the subsets will be distributed about the mean of the infinite data set. **The error between the mean of a finite data set and the true measurement value (mean of the infinite data set) is defined as the standard error of the mean, α, This is calculated as:**

$$\alpha = \sigma / \sqrt{n}$$

- The value of α approaches zero if the number of measurements in the data set expands towards infinity, or if σ approaches 0 . The measurement value obtained from a set of *n* measurements, $x_1, x_2, \ldots, x_n.$ measurement can then be expressed as: $x = x_{mean} \pm \alpha$

Example 6

- A set of length measurements consisting of 23 data points has a mean length value x_{mean} = 406.5 with a standard deviation σ = 1.88. Assuming normal distribution of data, express the value of the length as

$$x = x_{mean} \pm e$$

with a confidence limit of 68% (± σ boundaries)

Example 6. Solution

- A set of length measurements consisting of 23 data points has a mean length value x_{mean} = 406.5 with a standard deviation σ = 1.88. Assuming normal distribution of data, express the value of the length as

$$x = x_{mean} \pm e$$

with a confidence limit of

a) 68% (± σ boundaries)

$$x = x_{mean} \pm \alpha$$
$$\alpha = \sigma / \sqrt{n} = 1.88 / \sqrt{23} = 0.39$$
$$x = 406.5 \pm 0.4$$

b) 95.4% (± 2σ boundaries)

$$x = x_{mean} \pm \alpha$$
$$\alpha = 2\sigma / \sqrt{n} = 3.76 / \sqrt{23} = 0.78$$
$$x = 406.5 \pm 0.8$$

Estimation of random error in a single measurement

- In many situations where measurements are subject to random errors, it is not practical to take repeated measurements and find the average value.
- Also, the averaging process becomes invalid if the measured quantity does not remain at a constant value, as is usually the case when process variables are being measured.
- Thus, if only one measurement can be made, some means of estimating the likely magnitude of error in it is required.
- The normal approach to this is to calculate the error within 95% confidence limits, i.e. to calculate the value of the deviation D such that 95% of the area under the probability curve lies within limits of $\pm D$.
- These limits correspond to a deviation of $\pm 1.96\sigma$.

Estimation of random error in a single measurement

- Thus, it is necessary to maintain the measured quantity at a constant value whilst a number of measurements are taken in order to create a reference measurement set from which σ can be calculated.
- Subsequently, the maximum likely deviation in a single measurement can be expressed as: Deviation $D = \pm 1.96\sigma$.
- However, this only expresses the maximum likely deviation of the measurement from the calculated mean of the reference measurement set, which is not the true value as observed earlier.
- Thus the calculated value for the standard error of the mean has to be added to the likely maximum deviation value.
- Thus, the maximum likely error in a single measurement can be expressed as: **Error = $\pm(1.96\sigma + \alpha)$**

Example 7

- Suppose that a standard mass is measured 30 times with the same instrument to create a reference data set, and the calculated values of σ is σ = 0.43.
- If the instrument is then used to measure an unknown mass and the reading is 105.6 kg, express the mass value be with 95% confidence limits.

$$Error = \pm(1.96\sigma + \alpha)$$
$$Error = \pm\left(1.96\sigma + \sigma/\sqrt{n}\right)$$
$$Error = \pm\left(1.96 \times 0.43 + 0.43/\sqrt{30}\right) = 0.92$$

Mass = 105.6 ± 0.92 kg

Distribution of manufacturing tolerances

- Many aspects of manufacturing processes are subject to random variations caused by factors that are similar to those that cause random errors in measurements.
- In most cases, these random variations in manufacturing, which are known as tolerances, fit a Gaussian distribution, and the previous analysis of random measurement errors can be applied to analyse the distribution of these variations in manufacturing parameters.

Aggregation of measurement system errors

- Errors in measurement systems often arise from two or more different sources, and these must be aggregated in the correct way in order to obtain a prediction of the total likely error in output readings from the measurement system.

- Two different forms of aggregation are required.
- Firstly, a single measurement component may have both systematic and random errors and,
- secondly, a measurement system may consist of several measurement components that each have separate errors.

Combined effect of systematic and random errors

- If a measurement is affected by both systematic and random errors that are quantified as ±x (systematic errors) and ±y (random errors), some means of expressing the combined effect of both types of error is needed.
- One way of expressing the combined error would be to sum the two separate components of error, i.e. to say that the total possible error is e = ± (x+y). However, a more usual course of action is to express the likely maximum error as follows:

$$e = \sqrt{x^2 + y^2}$$

- It can be shown that this is the best expression for the error statistically, since it takes account of the reasonable assumption that the systematic and random errors are independent and so are unlikely to both be at their maximum or minimum value at the same time.

Aggregation of errors from separate measurement system components

- A measurement system often consists of several separate components, each of which is subject to errors. Therefore, what remains to be investigated is how the errors associated with each measurement system component combine together, so that a total error calculation can be made for the complete measurement system.
- All four mathematical operations of addition, subtraction, multiplication and division may be performed on measurements derived from different instruments/transducers in a measurement system. Appropriate techniques for the various situations that arise are covered below.

Error in a sum

- If the two outputs y and z of separate measurement system components are to be added together, we can write the sum as S = y + z.
- If the maximum errors in y and z are ± ay and ± bz respectively, one way to express the maximum and minimum possible values of S as:

$$S_{\max} = (y + ay) + (z + bz); \quad S_{\min} = (y - ay) + (z - bz); \quad \text{or } S = y + z \pm (ay + bz)$$

Error in a sum

- This relationship for S is not convenient because in this form the error term cannot be expressed as a fraction or percentage of the calculated value for S. Fortunately, statistical analysis can be applied that expresses S in an alternative form such that the most probable maximum error in S is represented by a quantity e, where e is calculated in terms of the absolute errors as:

$$e = \sqrt{(ay)^2 + (bz)^2}$$

- Thus. $S = (y + z) \pm e.$ This can be expressed in the alternative form

$$S = (y + z)(1 \pm f) \quad \text{where } f = e/(y + z)$$

Example 9

- A circuit requirement for a resistance of 550 Ω is satisfied by connecting together two resistors of nominal values 220 Ω and 330 Ω in series.
- If each resistor has a tolerance of ±2%, calculate the tolerance of the resulting resistance.

Example 9. Solution

- A circuit requirement for a resistance of 550 Ω is satisfied by connecting together two resistors of nominal values 220 Ω and 330 Ω in series. If each resistor has a tolerance of ±2%, calculate the tolerance of the resulting resistance.

$$e = \sqrt{(0.02 \times 220)^2 + (0.02 \times 330)^2} = 7.93$$

$$f = 7.93/550 = 0.0144$$

- Thus the total resistance S can be expressed as:

$S = 550\,\Omega \pm 7.93\,\Omega$ or $S = 550\ (1 \pm 0.0144)\,\Omega$, i.e. $S = 550\,\Omega \pm 1.4\%$

Error in a difference

- If the two outputs y and z of separate measurement systems are to be subtracted from one another, and the possible errors are ±ay and ±bz, then the difference S can be expressed (using statistical analysis as for calculating the error in a sum and assuming that the measurements are uncorrelated) as:

$$S = (y - z) \pm e \quad \text{or} \quad S = (y - z)(1 \pm f)$$

- where e and f are calculated as

$$e = \sqrt{(ay)^2 + (bz)^2}$$

$$f = e/(y - z)$$

Example 10

- A fluid flow rate is calculated from the difference in pressure measured on both sides of an orifice plate.
- If the pressure measurements are 10.0 bar and 9.5 bar and the error in the pressure measuring instruments is specified as ±0.1%, calculate the tolerance of the resulting flow rate measurement.

Example 10

- A fluid flow rate is calculated from the difference in pressure measured on both sides of an orifice plate. If the pressure measurements are 10.0 bar and 9.5 bar and the error in the pressure measuring instruments is specified as ±0.1%, calculate the tolerance of the resulting flow rate measurement.

$$e = \sqrt{(0.001 \times 10)^2 + (0.001 \times 9.5)^2} = 0.0138; \quad f = 0.0138/0.5 = 0.0276$$

The resulting flow rate has an error tolerance of 2.76 %

This example illustrates the relatively large error that can arise when calculations are made based on the difference between two measurements.

Error in a product

- If the outputs y and z of two measurement system components are multiplied together, the product can be written as P = yz. If the possible error in y is ± ay and in z is bz, then the maximum and minimum values possible in P can be written as:

$$P_{max} = (y + ay)(z + bz) = yz + ayz + byz + aybz;$$
$$P_{min} = (y - ay)(z - bz) = yz - ayz - byz + aybz$$

- For typical measurement system components with output errors of up to one or two per cent in magnitude, both a and b are very much less than one in magnitude and thus terms in aybz are negligible compared with other terms.
- Therefore, we have Pmax = yz(1 + a + b); Pmin = yz(1 - a – b). Thus the maximum error in the product P is ±(a+b).

Error in a product

- Whilst this expresses the maximum possible error in P, it tends to overestimate the likely maximum error since it is very unlikely that the errors in y and z will both be at the maximum or minimum value at the same time. A statistically better estimate of the likely maximum error e in the product P, provided that the measurements are uncorrelated, is given by:

$$e = \sqrt{a^2 + b^2}$$

- Note that in the case of multiplicative errors, e is calculated in terms of the fractional errors in y and z (as opposed to the absolute error values used in calculating additive errors).

Example 11

- If the power in a circuit is calculated from measurements of voltage and current in which the calculated maximum errors are respectively ±1% and ±2%, what is the maximum likely error in the calculated power value?

Example 11. Solution

- If the power in a circuit is calculated from measurements of voltage and current in which the calculated maximum errors are respectively ±1% and ±2%, what is the maximum likely error in the calculated power value?

$$e = \pm\sqrt{a^2 + b^2}$$
$$e = \pm\sqrt{0.01^2 + 0.02^2}$$
$$e = \pm 0.022$$
$$e = \pm 2.2\%$$

Error in a quotient

- If the output measurement y of one system component with possible error šay is divided by the output measurement z of another system component with possible error ±bz, then the maximum and minimum possible values for the quotient can be written as:

$$Q_{\max} = \frac{y+ay}{z-bz} = \frac{(y+ay)(z+bz)}{(z-bz)(z+bz)} = \frac{yz+ayz+byz+aybz}{z^2-b^2z^2};$$

$$Q_{\min} = \frac{y-ay}{z+bz} = \frac{(y-ay)(z-bz)}{(z+bz)(z-bz)} = \frac{yz-ayz-byz+aybz}{z^2-b^2z^2}$$

For a << 1 and b << 1, terms in ab and b^2 are negligible compared with the other terms. Hence:

$$Q_{\max} = \frac{yz(1+a+b)}{z^2}; \quad Q_{\min} = \frac{yz(1-a-b)}{z^2}; \quad \text{i.e. } Q = \frac{y}{z} \pm \frac{y}{z}(a+b)$$

Error in a quotient

- Thus the maximum error in the quotient is ±(a + b). However, using the same argument as made above for the product of measurements, a statistically better estimate of the likely maximum error in the quotient Q, provided that the measurements are uncorrelated, is that given as:

$$e = \sqrt{a^2 + b^2}$$

Example 12. Solution

- If the density of a substance is calculated from measurements of its mass and volume where the respective errors are ±2% and ± 3%, what is the maximum likely error in the density value?

$$e = \pm\sqrt{a^2 + b^2}$$
$$e = \pm\sqrt{0.02^2 + 0.03^2}$$
$$e = \pm 0.036$$
$$e = \pm 3.6\%$$

Total error when combining multiple measurements

- The final case to be covered is where the final measurement is calculated from several measurements that are combined together in a way that involves more than one type of arithmetic operation.
- For example, the density of a rectangular-sided solid block of material can be calculated from measurements of its mass divided by the product of measurements of its length, height and width.
- The errors involved in each stage of arithmetic are cumulative, and so the total measurement error can be calculated by adding together the two error values associated with the two multiplication stages involved in calculating the volume and then calculating the error in the final arithmetic operation when the mass is divided by the volume.

Example 12.

- A rectangular-sided block has edges of lengths a, b and c, and its mass is m. If the values and possible errors in quantities a, b, c and m are as shown below, calculate the value of density and the possible error in this value.

  a = 100mm ± 1%, b = 200mm ± 1%, c = 300mm ± 1%, m = 20 kg ± 0.5%.

  Value of ab = 0.02 m^2 ± 2% (possible error = 1% + 1% = 2%)
  Value of (ab)c = 0.006 m^3 ± 3% (possible error = 2% + 1% = 3%)
  Value of m/(abc) = 20/0.006 = 3330 kg/ m^3 ± 3.5%
  (possible error = 3% + 0.5% = 3.5%)

  Compare with

$$e = \pm\sqrt{a^2 + b^2 + c^2 + d^2}$$
$$e = \pm\sqrt{0.01^2 + 0.01^2 + 0.01^2 + 0.005^2}$$
$$e = \pm 0.018$$