31

Measures of central tendency and dispersion

31.1 Measures of central tendency

A single value, which is representative of a set of values, may be used to give an indication of the general size of the members in a set, the word **'average'** often being used to indicate the single value.

The statistical term used for 'average' is the arithmetic mean or just the **mean**.

Other measures of central tendency may be used and these include the **median** and the **modal** values.

31.2 Mean, median and mode for discrete data

Mean

The **arithmetic mean value** is found by adding together the values of the members of a set and dividing by the number of members in the set. Thus, the mean of the set of numbers: {4, 5, 6, 9} is:

$$\frac{4+5+6+9}{4}, \quad \text{i.e.} \quad 6$$

In general, the mean of the set: $\{x_1, x_2, x_3, \ldots x_n\}$ is

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}, \text{ written as } \frac{\sum x}{n}$$

where $\sum$ is the Greek letter 'sigma' and means 'the sum of', and $\bar{x}$ (called x-bar) is used to signify a mean value.

Median

The **median value** often gives a better indication of the general size of a set containing extreme values. The set: {7, 5, 74, 10} has a mean value of 24, which is not really representative of any of the values of the members of the set. The median value is obtained by:

(a) **ranking** the set in ascending order of magnitude, and

(b) selecting the value of the **middle member** for sets containing an odd number of members, or finding the value of the mean of the two middle members for sets containing an even number of members.

For example, the set: $\{7, 5, 74, 10\}$ is ranked as $\{5, 7, 10, 74\}$, and since it contains an even number of members (four in this case), the mean of 7 and 10 is taken, giving a median value of 8.5. Similarly, the set: $\{3, 81, 15, 7, 14\}$ is ranked as $\{3, 7, 14, 15, 81\}$ and the median value is the value of the middle member, i.e. 14.

Mode

The **modal value**, or **mode**, is the most commonly occurring value in a set. If two values occur with the same frequency, the set is 'bi-modal'. The set: $\{5, 6, 8, 2, 5, 4, 6, 5, 3\}$ has a modal value of 5, since the member having a value of 5 occurs three times.

Problem 1. Determine the mean, median and mode for the set:

$$\{2, 3, 7, 5, 5, 13, 1, 7, 4, 8, 3, 4, 3\}$$

The mean value is obtained by adding together the values of the members of the set and dividing by the number of members in the set.

Thus, **mean value**,

$$\bar{x} = \frac{\begin{array}{c}2+3+7+5+5+13+1\\+7+4+8+3+4+3\end{array}}{13} = \frac{65}{13} = \mathbf{5}$$

To obtain the median value the set is ranked, that is, placed in ascending order of magnitude, and since the set contains an odd number of members the value of the middle member is the median value. Ranking the set gives:

$$\{1, 2, 3, 3, 3, 4, 4, 5, 5, 7, 7, 8, 13\}$$

The middle term is the seventh member, i.e. 4, thus the **median value is 4**. The **modal value** is the value of the most commonly occurring member and is **3**, which occurs three times, all other members only occurring once or twice.

Problem 2. The following set of data refers to the amount of money in £s taken by a news vendor for 6 days. Determine the mean, median and modal values of the set:

$$\{27.90, 34.70, 54.40, 18.92, 47.60, 39.68\}$$

Mean value

$$= \frac{27.90 + 34.70 + 54.40 + 18.92 + 47.60 + 39.68}{6}$$

$$= \mathbf{£37.20}$$

The ranked set is:

$$\{18.92, 27.90, 34.70, 39.68, 47.60, 54.40\}$$

Since the set has an even number of members, the mean of the middle two members is taken to give the median value, i.e.

$$\textbf{median value} = \frac{34.70 + 39.68}{2} = \mathbf{£37.19}$$

Since no two members have the same value, this set has **no mode**.

Now try the following exercise

Exercise 112 Further problems on mean, median and mode for discrete data (Answers on page 282)

In Problems 1 to 4, determine the mean, median and modal values for the sets given.

1. $\{3, 8, 10, 7, 5, 14, 2, 9, 8\}$
2. $\{26, 31, 21, 29, 32, 26, 25, 28\}$
3. $\{4.72, 4.71, 4.74, 4.73, 4.72, 4.71, 4.73, 4.72\}$
4. $\{73.8, 126.4, 40.7, 141.7, 28.5, 237.4, 157.9\}$

31.3 Mean, median and mode for grouped data

The mean value for a set of grouped data is found by determining the sum of the (frequency × class mid-point values) and dividing by the sum of the frequencies,

i.e. mean value $\bar{x} = \dfrac{f_1x_1 + f_2x_2 + \cdots f_nx_n}{f_1 + f_2 + \cdots + f_n} = \dfrac{\sum(fx)}{\sum f}$

where f is the frequency of the class having a mid-point value of x, and so on.

Problem 3. The frequency distribution for the value of resistance in ohms of 48 resistors is as shown. Determine the mean value of resistance.

20.5–20.9	3,	21.0–21.4	10,	21.5–21.9	11,
22.0–22.4	13,	22.5–22.9	9,	23.0–23.4	2

The class mid-point/frequency values are:

20.7 3, 21.2 10, 21.7 11, 22.2 13, 22.7 9 and 23.2 2

For grouped data, the mean value is given by:

$$\bar{x} = \frac{\sum(fx)}{\sum f}$$

where f is the class frequency and x is the class mid-point value. Hence mean value,

$$\bar{x} = \frac{(3 \times 20.7) + (10 \times 21.2) + (11 \times 21.7) + (13 \times 22.2) + (9 \times 22.7) + (2 \times 23.2)}{48}$$

$$= \frac{1052.1}{48} = 21.919\ldots$$

i.e. **the mean value is 21.9 ohms**, correct to 3 significant figures.

Histogram

The mean, median and modal values for grouped data may be determined from a **histogram**. In a histogram, frequency values are represented vertically and variable values horizontally. The mean value is given by the value of the variable corresponding to a vertical line drawn through the centroid of the histogram. The median value is obtained by selecting a variable value such that the area of the histogram to the left of a vertical line drawn through the selected variable value is equal to the area of the histogram on the right of the line. The modal value is the variable value obtained by dividing the width of the highest rectangle in the histogram in proportion to the heights of the adjacent rectangles. The method of determining the mean, median and modal values from a histogram is shown in Problem 4.

Problem 4. The time taken in minutes to assemble a device is measured 50 times and the results are as shown. Draw a histogram depicting this data and hence determine the mean, median and modal values of the distribution.

14.5–15.5	5,	16.5–17.5	8,	18.5–19.5	16,
20.5–21.5	12,	22.5–23.5	6,	24.5–25.5	3

The histogram is shown in Fig. 31.1. The mean value lies at the centroid of the histogram. With reference to any arbitrary axis, say YY shown at a time of 14 minutes, the position of the horizontal value of the centroid can be obtained from the relationship $AM = \sum(am)$, where A is the area of the histogram, M is the horizontal distance of the centroid from the axis YY, a is the area of a rectangle of the histogram and m is the distance of the centroid of the rectangle from YY. The areas of the individual rectangles are shown circled on the histogram giving a total area of 100 square units. The positions, m, of the centroids of the individual rectangles are $1, 3, 5, \ldots$ units from YY. Thus

$$100M = (10 \times 1) + (16 \times 3) + (32 \times 5) + (24 \times 7) + (12 \times 9) + (6 \times 11)$$

i.e. $$M = \frac{560}{100} = 5.6 \text{ units from } YY$$

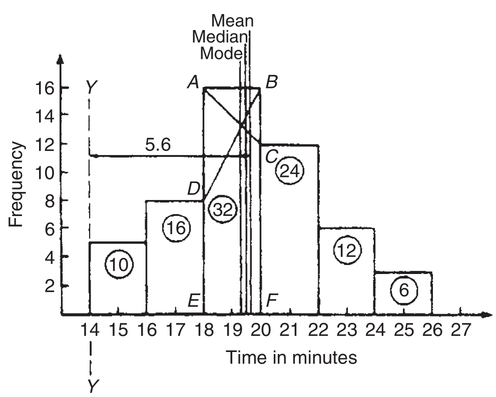

Fig. 31.1

Thus the position of the **mean** with reference to the time scale is $14 + 5.6$, i.e. **19.6 minutes**.

The median is the value of time corresponding to a vertical line dividing the total area of the histogram into two equal parts. The total area is 100 square units, hence the vertical line must be drawn to give 50 units of area on each side. To achieve this with reference to Fig. 31.1, rectangle $ABFE$ must be split so that $50 - (10 + 16)$ units of area lie on one side and $50 - (24 + 12 + 6)$ units of area lie on the other. This shows that the area of $ABFE$ is split so that 24 units of area lie to the left of the line and 8 units of area lie to the right, i.e. the vertical line must pass through 19.5 minutes. Thus the **median value** of the distribution is **19.5 minutes**.

The mode is obtained by dividing the line AB, which is the height of the highest rectangle, proportionally to the heights of the adjacent rectangles. With reference to Fig. 31.1, this is done by joining AC and BD and drawing a vertical line through the point of intersection of these two lines. This gives the **mode** of the distribution and is **19.3 minutes**.

Now try the following exercise

Exercise 113 Further problems on mean, median and mode for grouped data (Answers on page 282)

1. 21 bricks have a mean mass of 24.2 kg, and 29 similar bricks have a mass of 23.6 kg. Determine the mean mass of the 50 bricks.

2. The frequency distribution given below refers to the heights in centimetres of 100 people. Determine the mean value of the distribution, correct to the nearest millimetre.

150–156	5,	157–163	18,	164–170	20
171–177	27,	178–184	22,	185–191	8

3. The gain of 90 similar transistors is measured and the results are as shown.

83.5–85.5	6,	86.5–88.5	39,	89.5–91.5	27,
92.5–94.5	15,	95.5–97.5	3		

By drawing a histogram of this frequency distribution, determine the mean, median and modal values of the distribution.

4. The diameters, in centimetres, of 60 holes bored in engine castings are measured and the results are as shown. Draw a histogram depicting these results and hence determine the mean, median and modal values of the distribution.

2.011–2.014	7,	2.016–2.019	16,
2.021–2.024	23,	2.026–2.029	9,
2.031–2.034	5		

31.4 Standard deviation

(a) Discrete data

The standard deviation of a set of data gives an indication of the amount of dispersion, or the scatter, of members of the set from the measure of central tendency. Its value is the root-mean-square value of the members of the set and for discrete data is obtained as follows:

(a) determine the measure of central tendency, usually the mean value, (occasionally the median or modal values are specified),

(b) calculate the deviation of each member of the set from the mean, giving

$$(x_1 - \bar{x}), (x_2 - \bar{x}), (x_3 - \bar{x}), \ldots,$$

(c) determine the squares of these deviations, i.e.

$$(x_1 - \bar{x})^2, (x_2 - \bar{x})^2, (x_3 - \bar{x})^2, \ldots,$$

(d) find the sum of the squares of the deviations, that is

$$(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2, \ldots,$$

(e) divide by the number of members in the set, n, giving

$$\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \cdots}{n}$$

(f) determine the square root of (e).

The standard deviation is indicated by σ (the Greek letter small 'sigma') and is written mathematically as:

$$\textbf{standard deviation}, \boldsymbol{\sigma} = \sqrt{\left\{\frac{\sum(x-\bar{x})^2}{n}\right\}}$$

where x is a member of the set, $\bar{x}$ is the mean value of the set and n is the number of members in the set. The value of standard deviation gives an indication of the distance of the members of a set from the mean value. The set: $\{1, 4, 7, 10, 13\}$ has a mean value of 7 and a standard deviation of about 4.2. The set $\{5, 6, 7, 8, 9\}$ also has a mean value of 7, but the standard deviation is about 1.4. This shows that the members of the second set are mainly much closer to the mean value than the members of the first set. The method of determining the standard deviation for a set of discrete data is shown in Problem 5.

Problem 5. Determine the standard deviation from the mean of the set of numbers: $\{5, 6, 8, 4, 10, 3\}$, correct to 4 significant figures.

The arithmetic mean, $\bar{x} = \frac{\sum x}{n}$

$$= \frac{5+6+8+4+10+3}{6} = 6$$

Standard deviation, $\sigma = \sqrt{\left\{\frac{\sum(x-\bar{x})^2}{n}\right\}}$

The $(x-\bar{x})^2$ values are: $(5-6)^2$, $(6-6)^2$, $(8-6)^2$, $(4-6)^2$, $(10-6)^2$ and $(3-6)^2$.

The sum of the $(x-\bar{x})^2$ values,

i.e. $\sum(x-\bar{x})^2 = 1+0+4+4+16+9 = 34$

and $\frac{\sum(x-\bar{x})^2}{n} = \frac{34}{6} = 5.\dot{6}$

since there are 6 members in the set.

Hence, **standard deviation**,

$$\sigma = \sqrt{\left\{\frac{\sum(x-\bar{x}^2}{n}\right\}} = \sqrt{5.\dot{6}} = \mathbf{2.380}$$

correct to 4 significant figures.

(b) Grouped data

For **grouped data, standard deviation**

$$\boldsymbol{\sigma} = \sqrt{\left\{\frac{\sum\{f(x-\bar{x})^2\}}{\sum f}\right\}}$$

where f is the class frequency value, x is the class mid-point value and $\bar{x}$ is the mean value of the grouped data. The method of determining the standard deviation for a set of grouped data is shown in Problem 6.

Problem 6. The frequency distribution for the values of resistance in ohms of 48 resistors is as shown. Calculate the standard deviation from the mean of the resistors, correct to 3 significant figures.

20.5–20.9	3,	21.0–21.4	10,	21.5–21.9	11,
22.0–22.4	13,	22.5–22.9	9,	23.0–23.4	2

The standard deviation for grouped data is given by:

$$\sigma = \sqrt{\left\{\frac{\sum\{f(x-\bar{x})^2\}}{\sum f}\right\}}$$

From Problem 3, the distribution mean value, $\bar{x} = 21.92$, correct to 4 significant figures.

The 'x-values' are the class mid-point values, i.e. 20.7, 21.2, 21.7, ….

Thus the $(x-\bar{x})^2$ values are $(20.7–21.92)^2$, $(21.2–21.92)^2$, $(21.7–21.92)^2, \ldots$,

and the $f(x-\bar{x})^2$ values are $3(20.7–21.92)^2$, $10(21.2–21.92)^2$, $11(21.7–21.92)^2, \ldots$.

The $\sum f(x-\bar{x})^2$ values are

$$4.4652 + 5.1840 + 0.5324 + 1.0192$$
$$+5.4756 + 3.2768 = 19.9532$$

$$\frac{\sum\{f(x-\bar{x})^2\}}{\sum f} = \frac{19.9532}{48} = 0.41569$$

and **standard deviation,**

$$\boldsymbol{\sigma} = \sqrt{\left\{\frac{\sum\{f(x-\bar{x})^2\}}{\sum f}\right\}} = \sqrt{0.41569}$$

$$= \mathbf{0.645}, \text{ correct to 3 significant figures}$$

Now try the following exercise

Exercise 114 Further problems on standard deviation (Answers on page 283)

1. Determine the standard deviation from the mean of the set of numbers:

   $\{35, 22, 25, 23, 28, 33, 30\}$

   correct to 3 significant figures.

2. The values of capacitances, in microfarads, of ten capacitors selected at random from a large batch of similar capacitors are:

   34.3, 25.0, 30.4, 34.6, 29.6, 28.7,
   33.4, 32.7, 29.0 and 31.3

   Determine the standard deviation from the mean for these capacitors, correct to 3 significant figures.

3. The tensile strength in megapascals for 15 samples of tin were determined and found to be:

   34.61, 34.57, 34.40, 34.63, 34.63, 34.51, 34.49, 34.61, 34.52, 34.55, 34.58, 34.53, 34.44, 34.48 and 34.40

   Calculate the mean and standard deviation from the mean for these 15 values, correct to 4 significant figures.

4. Calculate the standard deviation from the mean for the mass of the 50 bricks given in Problem 1 of Exercise 113, page 237, correct to 3 significant figures.

5. Determine the standard deviation from the mean, correct to 4 significant figures, for the heights of the 100 people given in Problem 2 of Exercise 113, page 237.

6. Calculate the standard deviation from the mean for the data given in Problem 3 of Exercise 113, page 237, correct to 3 decimal places.

31.5 Quartiles, deciles and percentiles

Other measures of dispersion which are sometimes used are the quartile, decile and percentile values. The **quartile values** of a set of discrete data are obtained by selecting the values of members which divide the set into four equal parts. Thus for the set: $\{2, 3, 4, 5, 5, 7, 9, 11, 13, 14, 17\}$ there are 11 members and the values of the members dividing the set into four equal parts are 4, 7, and 13. These values are signified by Q_1, Q_2 and Q_3 and called the first, second and third quartile values, respectively. It can be seen that the second quartile value, Q_2, is the value of the middle member and hence is the median value of the set.

For grouped data the ogive may be used to determine the quartile values. In this case, points are selected on the vertical cumulative frequency values of the ogive, such that they divide the total value of cumulative frequency into four equal parts. Horizontal lines are drawn from these values to cut the ogive. The values of the variable corresponding to these cutting points on the ogive give the quartile values (see Problem 7).

When a set contains a large number of members, the set can be split into ten parts, each containing an equal number of members. These ten parts are then called **deciles**. For sets containing a very large number of members, the set may be split into one hundred parts, each containing an equal number of members. One of these parts is called a **percentile**.

Problem 7. The frequency distribution given below refers to the overtime worked by a group of craftsmen during each of 48 working weeks in a year.

25–29	5,	30–34	4,	35–39	7,	40–44	11,
45–49	12,	50–54	8,	55–59	1		

Draw an ogive for this data and hence determine the quartile values.

The cumulative frequency distribution (i.e. upper class boundary/cumulative frequency values) is:

29.5	5,	34.5	9,	39.5	16,	44.5	27,
49.5	39,	54.5	47,	59.5	48		

The ogive is formed by plotting these values on a graph, as shown in Fig. 29.2. The total frequency is divided into four equal parts, each having a range of 48/4, i.e. 12. This gives cumulative frequency values of 0 to 12 corresponding to the first quartile, 12 to 24 corresponding to the second quartile, 24 to 36 corresponding to the third quartile and 36 to 48 corresponding to the fourth quartile of the distribution, i.e. the distribution is divided into four equal parts. The quartile values are those of the variable corresponding to cumulative frequency values of 12, 24 and 36, marked Q_1, Q_2 and Q_3 in Fig. 31.2. These values, correct to the nearest hour, are **37 hours, 43 hours and 48 hours**, respectively. The Q_2 value is also equal to the median value of the distribution. One measure of the dispersion of a distribution is called the **semi-interquartile range** and is given by $(Q_3 - Q_1)/2$, and is $(48 - 37)/2$ in this case, i.e. $\mathbf{5\frac{1}{2}}$ **hours**.

Problem 8. Determine the numbers contained in the (a) 41st to 50th percentile group, and (b) 8th decile group of the set of numbers shown below:

14	22	17	21	30	28	37	7	23	32
24	17	20	22	27	19	26	21	15	29

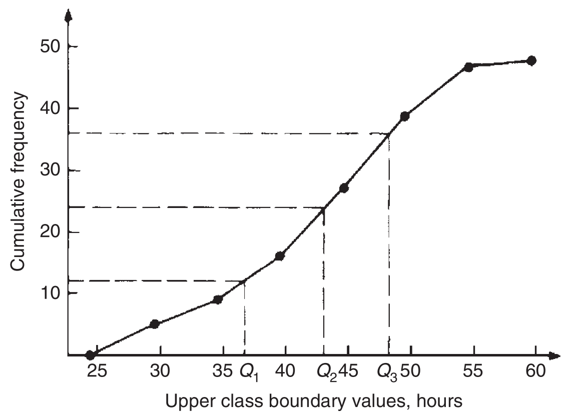

Fig. 31.2

The set is ranked, giving:

7	14	15	17	17	19	20	21	21	22
22	23	24	26	27	28	29	30	32	37

(a) There are 20 numbers in the set, hence the first 10% will be the two numbers 7 and 14, the second 10% will be 15 and 17, and so on. Thus the 41st to 50th percentile group will be the numbers **21 and 22**.

(b) The first decile group is obtained by splitting the ranked set into 10 equal groups and selecting the first group, i.e. the numbers 7 and 14. The second decile group are the numbers 15 and 17, and so on. Thus the 8th decile group contains the numbers **27 and 28**.

Now try the following exercise

Exercise 115 Further problems on quartiles, deciles and percentiles (Answers on page 283)

1. The number of working days lost due to accidents for each of 12 one-monthly periods are as shown. Determine the median and first and third quartile values for this data.

   27 37 40 28 23 30 35 24 30 32 31 28

2. The number of faults occurring on a production line in a nine-week period are as shown below. Determine the median and quartile values for the data.

   30 27 25 24 27 37 31 27 35

3. Determine the quartile values and semi-interquartile range for the frequency distribution given in Problem 2 of Exercise 113, page 236.

4. Determine the numbers contained in the 5th decile group and in the 61st to 70th percentile groups for the set of numbers:

40	46	28	32	37	42	50	31	48	45
32	38	27	33	40	35	25	42	38	41

5. Determine the numbers in the 6th decile group and in the 81st to 90th percentile group for the set of numbers:

   43 47 30 25 15 51 17 21 37 33 44 56 40 49 22

   36 44 33 17 35 58 51 35 44 40 31 41 55 50 16

32

Probability

32.1 Introduction to probability

Probability

The **probability** of something happening is the likelihood or chance of it happening. Values of probability lie between 0 and 1, where 0 represents an absolute impossibility and 1 represents an absolute certainty. The probability of an event happening usually lies somewhere between these two extreme values and is expressed either as a proper or decimal fraction. Examples of probability are:

that a length of copper wire has zero resistance at 100°C	0
that a fair, six-sided dice will stop with a 3 upwards	$\frac{1}{6}$ or 0.1667
that a fair coin will land with a head upwards	$\frac{1}{2}$ or 0.5
that a length of copper wire has some resistance at 100°C	1

If p is the probability of an event happening and q is the probability of the same event not happening, then the total probability is $p+q$ and is equal to unity, since it is an absolute certainty that the event either does or does not occur, i.e. $\boldsymbol{p+q=1}$

Expectation

The **expectation**, E, of an event happening is defined in general terms as the product of the probability p of an event happening and the number of attempts made, n, i.e. $\boldsymbol{E=pn}$.

Thus, since the probability of obtaining a 3 upwards when rolling a fair dice is $\frac{1}{6}$, the expectation of getting a 3 upwards on four throws of the dice is $\frac{1}{6}\times 4$, i.e. $\frac{2}{3}$

Thus expectation is the average occurrence of an event.

Dependent event

A **dependent event** is one in which the probability of an event happening affects the probability of another ever happening. Let 5 transistors be taken at random from a batch of 100 transistors for test purposes, and the probability of there being a defective transistor, p_1, be determined. At some later time, let another 5 transistors be taken at random from the 95 remaining transistors in the batch and the probability of there being a defective transistor, p_2, be determined. The value of p_2 is different from p_1 since batch size has effectively altered from 100 to 95, i.e. probability p_2 is dependent on probability p_1. Since transistors are drawn, and then another 5 transistors drawn without replacing the first 5, the second random selection is said to be **without replacement**.

Independent event

An independent event is one in which the probability of an event happening does not affect the probability of another event happening. If 5 transistors are taken at random from a batch of transistors and the probability of a defective transistor p_1 is determined and the process is repeated after the original 5 have been replaced in the batch to give p_2, then p_1 is equal to p_2. Since the 5 transistors are replaced between draws, the second selection is said to be **with replacement**.

32.2 Laws of probability

The addition law of probability

The addition law of probability is recognised by the word **'or'** joining the probabilities. If p_A is the probability of event A happening and p_B is the probability of event B happening, the probability of **event *A* or event *B*** happening is given by p_A+p_B. Similarly, the probability of events ***A* or *B* or *C* or** ... ***N*** happening is given by

$$\boldsymbol{p_A+p_B+p_C+\cdots+p_N}$$

The multiplication law of probability

The multiplication law of probability is recognised by the word **'and'** joining the probabilities. If p_A is the probability of event A happening and p_B is the probability of event B happening, the probability of **event** $\boldsymbol{A}$ **and event** $\boldsymbol{B}$ happening is given by $p_A \times p_B$. Similarly, the probability of events $\boldsymbol{A}$ **and** $\boldsymbol{B}$ **and** $\boldsymbol{C}$ **and ...** $\boldsymbol{N}$ happening is given by

$$\boldsymbol{p_A \times p_B \times p_C \times \cdots \times p_N}$$

32.3 Worked problems on probability

Problem 1. Determine the probabilities of selecting at random (a) a man, and (b) a woman from a crowd containing 20 men and 33 women.

(a) The probability of selecting at random a man, p, is given by the ratio

$$\frac{\text{number of men}}{\text{number in crowd}} \quad \text{i.e.} \quad p = \frac{20}{20+33} = \boldsymbol{\frac{20}{53}} \quad \text{or} \quad \mathbf{0.3774}$$

(b) The probability of selecting at random a women, q, is given by the ratio

$$\frac{\text{number of women}}{\text{number in crowd}} \quad \text{i.e.} \quad q = \frac{33}{20+33} = \boldsymbol{\frac{33}{53}} \quad \text{or} \quad \mathbf{0.6226}$$

(Check: the total probability should be equal to 1;

$$p = \frac{20}{53} \text{ and } q = \frac{33}{53}$$

thus the total probability,

$$p + q = \frac{20}{53} + \frac{33}{53} = 1$$

hence no obvious error has been made).

Problem 2. Find the expectation of obtaining a 4 upwards with 3 throws of a fair dice.

Expectation is the average occurrence of an event and is defined as the probability times the number of attempts. The probability, p, of obtaining a 4 upwards for one throw of the dice is $\frac{1}{6}$.

Also, 3 attempts are made, hence $n = 3$ and the expectation, E, is pn, i.e. $\boldsymbol{E} = \frac{1}{6} \times 3 = \frac{1}{2}$ or **0.50**.

Problem 3. Calculate the probabilities of selecting at random:

(a) the winning horse in a race in which 10 horses are running,

(b) the winning horses in both the first and second races if there are 10 horses in each race.

(a) Since only one of the ten horses can win, the probability of selecting at random the winning horse is $\frac{\text{number of winners}}{\text{number of horses}}$, i.e. $\boldsymbol{\frac{1}{10}}$ or **0.10**

(b) The probability of selecting the winning horse in the first race is $\frac{1}{10}$. The probability of selecting the winning horse in the second race is $\frac{1}{10}$. The probability of selecting the winning horses in the first **and** second race is given by the multiplication law of probability,

$$\text{i.e.} \quad \textbf{probability} = \frac{1}{10} \times \frac{1}{10} = \boldsymbol{\frac{1}{100}} \quad \text{or} \quad \mathbf{0.01}$$

Problem 4. The probability of a component failing in one year due to excessive temperature is $\frac{1}{20}$, due to excessive vibration is $\frac{1}{25}$ and due to excessive humidity is $\frac{1}{50}$. Determine the probabilities that during a one-year period a component: (a) fails due to excessive temperature and excessive vibration, (b) fails due to excessive vibration or excessive humidity, and (c) will not fail because of both excessive temperature and excessive humidity.

Let p_A be the probability of failure due to excessive temperature, then

$$p_A = \frac{1}{20} \quad \text{and} \quad \overline{p_A} = \frac{19}{20}$$

(where $\overline{p_A}$ is the probability of not failing.)

Let p_B be the probability of failure due to excessive vibration, then

$$p_B = \frac{1}{25} \quad \text{and} \quad \overline{p_B} = \frac{24}{25}$$

Let p_C be the probability of failure due to excessive humidity, then

$$p_C = \frac{1}{50} \quad \text{and} \quad \overline{p_C} = \frac{49}{50}$$

(a) The probability of a component failing due to excessive temperature **and** excessive vibration is given by:

$$p_A \times p_B = \frac{1}{20} \times \frac{1}{25} = \mathbf{\frac{1}{500}} \quad \text{or} \quad \mathbf{0.002}$$

(b) The probability of a component failing due to excessive vibration **or** excessive humidity is:

$$p_B + p_C = \frac{1}{25} + \frac{1}{50} = \mathbf{\frac{3}{50}} \quad \text{or} \quad \mathbf{0.06}$$

(c) The probability that a component will not fail due excessive temperature **and** will not fail due to excess humidity is:

$$\overline{p_A} \times \overline{p_C} = \frac{19}{20} \times \frac{49}{50} = \mathbf{\frac{931}{1000}} \quad \text{or} \quad \mathbf{0.931}$$

Problem 5. A batch of 100 capacitors contains 73 which are within the required tolerance values, 17 which are below the required tolerance values, and the remainder are above the required tolerance values. Determine the probabilities that when randomly selecting a capacitor and then a second capacitor: (a) both are within the required tolerance values when selecting with replacement, and (b) the first one drawn is below and the second one drawn is above the required tolerance value, when selection is without replacement.

(a) The probability of selecting a capacitor within the required tolerance values is $\frac{73}{100}$. The first capacitor drawn is now replaced and a second one is drawn from the batch of 100. The probability of this capacitor being within the required tolerance values is also $\frac{73}{100}$.

Thus, the probability of selecting a capacitor within the required tolerance values for both the first **and** the second draw is

$$\frac{73}{100} \times \frac{73}{100} = \mathbf{\frac{5329}{10\,000}} \quad \text{or} \quad \mathbf{0.5329}$$

(b) The probability of obtaining a capacitor below the required tolerance values on the first draw is $\frac{17}{100}$. There are now only 99 capacitors left in the batch, since the first capacitor is not replaced. The probability of drawing a capacitor above the required tolerance values on the second draw is $\frac{10}{99}$, since there are $(100 - 73 - 17)$, i.e. 10 capacitors above the required tolerance value. Thus, the probability of randomly selecting a capacitor below the required tolerance values and followed by randomly selecting a capacitor above the tolerance values is

$$\frac{17}{100} \times \frac{10}{99} = \frac{170}{9900} = \mathbf{\frac{17}{990}} \quad \text{or} \quad \mathbf{0.0172}$$

Now try the following exercise

Exercise 116 Further problems on probability (Answers on page 283)

1. In a batch of 45 lamps there are 10 faulty lamps. If one lamp is drawn at random, find the probability of it being (a) faulty and (b) satisfactory.
2. A box of fuses are all of the same shape and size and comprises 23 2 A fuses, 47 5 A fuses and 69 13 A fuses. Determine the probability of selecting at random (a) a 2 A fuse, (b) a 5 A fuse and (c) a 13 A fuse.
3. (a) Find the probability of having a 2 upwards when throwing a fair 6-sided dice. (b) Find the probability of having a 5 upwards when throwing a fair 6-sided dice. (c) Determine the probability of having a 2 and then a 5 on two successive throws of a fair 6-sided dice.
4. The probability of event A happening is $\frac{3}{5}$ and the probability of event B happening is $\frac{2}{3}$. Calculate the probabilities of (a) both A and B happening, (b) only event A happening, i.e. event A happening and event B not happening, (c) only event B happening, and (d) either A, or B, or A and B happening.
5. When testing 1000 soldered joints, 4 failed during a vibration test and 5 failed due to having a high resistance. Determine the probability of a joint failing due to (a) vibration, (b) high resistance, (c) vibration or high resistance and (d) vibration and high resistance.

32.4 Further worked problems on probability

Problem 6. A batch of 40 components contains 5 which are defective. A component is drawn at random from the batch and tested and then a second component is drawn. Determine the probability that neither of the components is defective when drawn (a) with replacement, and (b) without replacement.

(a) With replacement

The probability that the component selected on the first draw is satisfactory is $\frac{35}{40}$, i.e. $\frac{7}{8}$. The component is now replaced and a second draw is made. The probability that this component is also satisfactory is $\frac{7}{8}$. Hence, the probability that both the first component drawn **and** the second component drawn are satisfactory is:

$$\frac{7}{8} \times \frac{7}{8} = \mathbf{\frac{49}{64}} \quad \text{or} \quad \mathbf{0.7656}$$

(b) Without replacement

The probability that the first component drawn is satisfactory is $\frac{7}{8}$. There are now only 34 satisfactory components left in the batch and the batch number is 39. Hence, the probability of drawing a satisfactory component on the second draw is $\frac{34}{39}$. Thus the probability that the first component drawn **and** the second component drawn are satisfactory, i.e. neither is defective, is:

$$\frac{7}{8}\times\frac{34}{39}=\mathbf{\frac{238}{312}}\quad\text{or}\quad\mathbf{0.7628}$$

Problem 7. A batch of 40 components contains 5 which are defective. If a component is drawn at random from the batch and tested and then a second component is drawn at random, calculate the probability of having one defective component, both with and without replacement.

The probability of having one defective component can be achieved in two ways. If p is the probability of drawing a defective component and q is the probability of drawing a satisfactory component, then the probability of having one defective component is given by drawing a satisfactory component and then a defective component **or** by drawing a defective component and then a satisfactory one, i.e. by $q\times p+p\times q$

With replacement:

$$p=\frac{5}{40}=\frac{1}{8}\quad\text{and}\quad q=\frac{35}{40}=\frac{7}{8}$$

Hence, probability of having one defective component is:

$$\frac{1}{8}\times\frac{7}{8}+\frac{7}{8}\times\frac{1}{8}$$

$$\text{i.e.}\qquad \frac{7}{64}+\frac{7}{64}=\mathbf{\frac{7}{32}}\quad\text{or}\quad\mathbf{0.2188}$$

Without replacement:

$p_1=\frac{1}{8}$ and $q_1=\frac{7}{8}$ on the first of the two draws. The batch number is now 39 for the second draw, thus,

$$p_2=\frac{5}{39}\quad\text{and}\quad q_2=\frac{35}{39}$$

$$p_1q_2+q_1p_2=\frac{1}{8}\times\frac{35}{39}+\frac{7}{8}\times\frac{5}{39}=\frac{35+35}{312}$$

$$=\mathbf{\frac{70}{312}}\quad\text{or}\quad\mathbf{0.2244}$$

Problem 8. A box contains 74 brass washers, 86 steel washers and 40 aluminium washers. Three washers are drawn at random from the box without replacement. Determine the probability that all three are steel washers.

Assume, for clarity of explanation, that a washer is drawn at random, then a second, then a third (although this assumption does not affect the results obtained). The total number of washers is $74+86+40$, i.e. 200.

The probability of randomly selecting a steel washer on the first draw is $\frac{86}{200}$. There are now 85 steel washers in a batch of 199. The probability of randomly selecting a steel washer on the second draw is $\frac{85}{199}$. There are now 84 steel washers in a batch of 198. The probability of randomly selecting a steel washer on the third draw is $\frac{84}{198}$. Hence the probability of selecting a steel washer on the first draw **and** the second draw **and** the third draw is:

$$\frac{86}{200}\times\frac{85}{199}\times\frac{84}{198}=\frac{614\,040}{7\,880\,400}=\mathbf{0.0779}$$

Problem 9. For the box of washers given in Problem 8 above, determine the probability that there are no aluminium washers drawn, when three washers are drawn at random from the box without replacement.

The probability of not drawing an aluminium washer on the first draw is $1-\left(\frac{40}{200}\right)$, i.e. $\frac{160}{200}$. There are now 199 washers in the batch of which 159 are not aluminium washers. Hence, the probability of not drawing an aluminium washer on the second draw is $\frac{159}{199}$. Similarly, the probability of not drawing an aluminium washer on the third draw is $\frac{158}{198}$. Hence the probability of not drawing an aluminium washer on the first **and** second **and** third draws is

$$\frac{160}{200}\times\frac{159}{199}\times\frac{158}{198}=\frac{4\,019\,520}{7\,880\,400}=\mathbf{0.5101}$$

Problem 10. For the box of washers in Problem 8 above, find the probability that there are two brass washers and either a steel or an aluminium washer when three are drawn at random, without replacement.

Two brass washers (A) and one steel washer (B) can be obtained in any of the following ways:

1st draw	2nd draw	3rd draw
A	A	B
A	B	A
B	A	A

Two brass washers and one aluminium washer (C) can also be obtained in any of the following ways:

1st draw	2nd draw	3rd draw
A	A	C
A	C	A
C	A	A

Thus there are six possible ways of achieving the combinations specified. If A represents a brass washer, B a steel washer and C an aluminium washer, then the combinations and their probabilities are as shown:

DRAW			PROBABILITY
First	Second	Third	
A	A	B	$\frac{74}{200} \times \frac{73}{199} \times \frac{86}{198} = 0.0590$
A	B	A	$\frac{74}{200} \times \frac{86}{199} \times \frac{73}{198} = 0.0590$
B	A	A	$\frac{86}{200} \times \frac{74}{199} \times \frac{73}{198} = 0.0590$
A	A	C	$\frac{74}{200} \times \frac{73}{199} \times \frac{40}{198} = 0.0274$
A	C	A	$\frac{74}{200} \times \frac{40}{199} \times \frac{73}{198} = 0.0274$
C	A	A	$\frac{40}{200} \times \frac{74}{199} \times \frac{73}{198} = 0.0274$

The probability of having the first combination **or** the second, **or** the third, and so on, is given by the sum of the probabilities,

i.e. by $3 \times 0.0590 + 3 \times 0.0274$, that is, **0.2592**

Now try the following exercise

Exercise 117 Further problems on probability (Answers on page 283)

1. The probability that component A will operate satisfactorily for 5 years is 0.8 and that B will operate satisfactorily over that same period of time is 0.75. Find the probabilities that in a 5 year period: (a) both components operate satisfactorily, (b) only component A will operate satisfactorily, and (c) only component B will operate satisfactorily.
2. In a particular street, 80% of the houses have telephones. If two houses selected at random are visited, calculate the probabilities that (a) they both have a telephone and (b) one has a telephone but the other does not have telephone.
3. Veroboard pins are packed in packets of 20 by a machine. In a thousand packets, 40 have less than 20 pins. Find the probability that if 2 packets are chosen at random, one will contain less than 20 pins and the other will contain 20 pins or more.
4. A batch of 1 kW fire elements contains 16 which are within a power tolerance and 4 which are not. If 3 elements are selected at random from the batch, calculate the probabilities that (a) all three are within the power tolerance and (b) two are within but one is not within the power tolerance.
5. An amplifier is made up of three transistors, A, B and C. The probabilities of A, B or C being defective are $\frac{1}{20}$, $\frac{1}{25}$ and $\frac{1}{50}$, respectively. Calculate the percentage of amplifiers produced (a) which work satisfactorily and (b) which have just one defective transistor.
6. A box contains 14 40 W lamps, 28 60 W lamps and 58 25 W lamps, all the lamps being of the same shape and size. Three lamps are drawn at random from the box, first one, then a second, then a third. Determine the probabilities of: (a) getting one 25 W, one 40 W and one 60 W lamp, with replacement, (b) getting one 25 W, one 40 W and one 60 W lamp without replacement, and (c) getting either one 25 W and two 40 W or one 60 W and two 40 W lamps with replacement.

Assignment 14

This assignment covers the material in Chapters 30 to 32. The marks for each question are shown in brackets at the end of each question.

1. A company produces five products in the following proportions:

Product A	24	Product B	6	Product C	15
Product D	9	Product E	18		

   Draw (a) a horizontal bar chart, and (b) a pie diagram to represent these data visually. (7)

2. State whether the data obtained on the following topics are likely to be discrete or continuous:

   (a) the number of books in a library
   (b) the speed of a car
   (c) the time to failure of a light bulb (3)

3. Draw a histogram, frequency polygon and ogive for the data given below which refers to the diameter of 50 components produced by a machine.

Class intervals	Frequency
1.30–1.32 mm	4
1.33–1.35 mm	7
1.36–1.38 mm	10
1.39–1.41 mm	12
1.42–1.44 mm	8
1.45–1.47 mm	5
1.48–1.50 mm	4

   (10)

4. Determine the mean, median and modal values for the lengths given in metres:

   28, 20, 44, 30, 32, 30, 28, 34, 26, 28 (3)

5. The length in millimetres of 100 bolts is as shown below:

50–56	6	57–63	16	64–70	22
71–77	30	78–84	19	85–91	7

   Determine for the sample (a) the mean value, and (b) the standard deviation, correct to 4 significant figures. (9)

6. The number of faulty components in a factory in a 12 week period is:

   14 12 16 15 10 13 15 11 16 19 17 19

   Determine the median and the first and third quartile values. (3)

7. Determine the probability of winning a prize in a lottery by buying 10 tickets when there are 10 prizes and a total of 5000 tickets sold. (2)

8. A sample of 50 resistors contains 44 which are within the required tolerance value, 4 which are below and the remainder which are above. Determine the probability of selecting from the sample a resistor which is (a) below the required tolerance, and (b) above the required tolerance. Now two resistors are selected at random from the sample. Determine the probability, correct to 3 decimal places, that neither resistor is defective when drawn (c) with replacement, and (d) without replacement. (e) If a resistor is drawn at random from the batch and tested, and then a second resistor is drawn from those left, calculate the probability of having one defective component when selection is without replacement. (13)

33

Introduction to differentiation

33.1 Introduction to calculus

Calculus is a branch of mathematics involving or leading to calculations dealing with continuously varying functions.

Calculus is a subject that falls into two parts:

(i) **differential calculus** (or **differentiation**) and

(ii) **integral calculus** (or **integration**).

Differentiation is used in calculations involving rates of change (see section 33.10), velocity and acceleration, and maximum and minimum values of curves (see 'Engineering Mathematics').

33.2 Functional notation

In an equation such as $y = 3x^2 + 2x - 5$, y is said to be a function of x and may be written as $y = f(x)$.

An equation written in the form $f(x) = 3x^2 + 2x - 5$ is termed **functional notation**.

The value of $f(x)$ when $x = 0$ is denoted by $f(0)$, and the value of $f(x)$ when $x = 2$ is denoted by $f(2)$ and so on.

Thus when $f(x) = 3x^2 + 2x - 5$, then

$f(0) = 3(0)^2 + 2(0) - 5 = -5$

and $f(2) = 3(2)^2 + 2(2) - 5 = 11$ and so on.

Problem 1. If $f(x) = 4x^2 - 3x + 2$ find: $f(0)$, $f(3)$, $f(-1)$ and $f(3) - f(-1)$

$$f(x) = 4x^2 - 3x + 2$$
$$f(0) = 4(0)^2 - 3(0) + 2 = \mathbf{2}$$
$$f(3) = 4(3)^2 - 3(3) + 2 = 36 - 9 + 2 = \mathbf{29}$$
$$f(-1) = 4(-1)^2 - 3(-1) + 2 = 4 + 3 + 2 = \mathbf{9}$$
$$f(3) - f(-1) = 29 - 9 = \mathbf{20}$$

Problem 2. Given that $f(x) = 5x^2 + x - 7$ determine:

(i) $f(2) \div f(1)$ (ii) $f(3 + a)$

(iii) $f(3 + a) - f(3)$ (iv) $\dfrac{f(3 + a) - f(3)}{a}$

$$f(x) = 5x^2 + x - 7$$

(i) $f(2) = 5(2)^2 + 2 - 7 = 15$

$f(1) = 5(1)^2 + 1 - 7 = -1$

$f(2) \div f(1) = \dfrac{15}{-1} = \mathbf{-15}$

(ii) $f(3 + a) = 5(3 + a)^2 + (3 + a) - 7$

$= 5(9 + 6a + a^2) + (3 + a) - 7$

$= 45 + 30a + 5a^2 + 3 + a - 7 = \mathbf{41 + 31a + 5a^2}$

(iii) $f(3) = 5(3)^2 + 3 - 7 = 41$

$f(3 + a) - f(3) = (41 + 31a + 5a^2) - (41) = \mathbf{31a + 5a^2}$

(iv) $\dfrac{f(3 + a) - f(3)}{a} = \dfrac{31a + 5a^2}{a} = \mathbf{31 + 5a}$

Now try the following exercise

Exercise 118 Further problems on functional notation (Answers on page 283)

1. If $f(x) = 6x^2 - 2x + 1$ find $f(0)$, $f(1)$, $f(2)$, $f(-1)$ and $f(-3)$
2. If $f(x) = 2x^2 + 5x - 7$ find $f(1)$, $f(2)$, $f(-1)$, $f(2) - f(-1)$