11 PROBABILITY AND STATISTICS

WHAT YOU NEED TO KNOW

- Standard deviation, σ, is a measure of how spread out the data is relative to the mean, μ. The square of the standard deviation is called the variance, and it has the formula:

  $\sigma^2 = \dfrac{\sum_{i=1}^{k} f_i x_i^2}{n} - \mu^2$ where $\mu = \dfrac{\sum_{i=1}^{k} f_i x_i}{n}$, $n = \sum_{i=1}^{k} f_i$ and f_i is the frequency of the ith data value.

- The probability of an event can be found by listing or counting all possible outcomes. The probabilities of combined events are often best found using Venn diagrams or tree diagrams.
  - Venn diagrams are often useful when the question involves a union of events.
    - The union of events A and B can be found using the formula:

      $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$
    - For mutually exclusive events, the formula becomes $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$.
  - Tree diagrams can be used when the question involves a sequence of events.
- The probability of event A happening given that an event B has already happened is known as conditional probability and is given by:

  $$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}$$
  - Events A and B are independent if $\mathrm{P}(A \mid B) = \mathrm{P}(A)$ or $\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B)$.
- Bayes' theorem is a formula for relating conditional probabilities:
  - When there are only two outcomes: $\mathrm{P}(B \mid A) = \dfrac{\mathrm{P}(B)\,\mathrm{P}(A \mid B)}{\mathrm{P}(B)\,\mathrm{P}(A \mid B) + \mathrm{P}(B')\,\mathrm{P}(A \mid B')}$
  - When there are more than two outcomes for event B:

    $$\mathrm{P}(B_i \mid A) = \frac{\mathrm{P}(A \mid B_i)\mathrm{P}(B_i)}{\mathrm{P}(A \mid B_1)\mathrm{P}(B_1) + \mathrm{P}(A \mid B_2)\mathrm{P}(B_2) + \cdots + \mathrm{P}(A \mid B_n)\mathrm{P}(B_n)}$$
- A discrete random variable can be described by its probability distribution, which is the list of all possible values and their probabilities.
  - The total of all the probabilities must always equal 1.
  - The expected value is $\mathrm{E}(X) = \mu = \sum_x x\,\mathrm{P}(X = x)$.
  - The variance is $\mathrm{Var}(X) = \mathrm{E}(X - \mu)^2 = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$.
- If there is a fixed number of trials (n) with constant and independent probability of success (p) in each trial, the number of successes follows a binomial distribution: $X \sim \mathrm{B}(n, p)$.
  - $\mathrm{E}(X) = np$
  - $\mathrm{Var}(X) = np(1 - p)$

- If successes occur independently and at a constant average rate (m), the number of successes in a given period follows a Poisson distribution: $X \sim \text{Po}(m)$.
  - $\text{E}(X) = m$
  - $\text{Var}(X) = m$
- A continuous random variable X can be described by its probability density function, $f(x)$.
  - Its probability is found by integration: $\text{P}(a < X < b) = \int_a^b f(x)\,\text{d}x$.
  - The total probability over all cases must equal 1: $\int_{-\infty}^{\infty} f(x)\,\text{d}x = 1$.
  - The expected value is $\text{E}(X) = \mu = \int_{-\infty}^{\infty} x f(x)\,\text{d}x$.
  - The variance is $\text{Var}(X) = \text{E}(X^2) - [\text{E}(X)]^2$ where $\text{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,\text{d}x$.
- The normal distribution is determined by its mean (μ) and variance (σ^2): $X \sim \text{N}(\mu, \sigma^2)$.
  - The inverse normal distribution is used to find the value of x which corresponds to a given cumulative probability: $p = \text{P}(X \le x)$.
  - If μ or σ is unknown, use the standard normal distribution to replace the values of X with their Z-scores, $z = \frac{x - \mu}{\sigma}$, which satisfy $Z \sim \text{N}(0, 1)$.

EXAM TIPS AND COMMON ERRORS

- Make sure that you do not confuse standard deviation and variance, especially when working with the normal distribution.
- When interpreting probability questions, pay particular attention to whether the required probability is conditional or not.
- When you use your GDC to find probabilities, you must write the results using the correct mathematical notation, **not** calculator notation.
- If a question mentions average rate of success, or events occurring at a constant rate, you should use the Poisson distribution. If you can identify a fixed number of trials, then the binomial distribution is appropriate.
- With the Poisson distribution, make sure that you are using the correct mean for the time (or spatial) interval.

11.1 CALCULATING THE MEAN AND STANDARD DEVIATION FROM SUMMARY STATISTICS

WORKED EXAMPLE 11.1

A teacher records the time, t minutes, it takes her to drive to work every morning. The times for 12 days are summarised by: $\sum t_i = 256$ and $\sum t_i^2 = 5963$. She then adds the time for the 13th day, which was 24 minutes. Calculate the mean and standard deviation of all 13 times.

The new sums are:

$$\sum t_i = 256 + 24 = 280$$

$$\sum t_i^2 = 5963 + 24^2 = 6539$$

Find the new totals including the extra data value of 24 minutes.

$$\bar{t} = \frac{280}{13} = 21.5 \text{ (3 SF)}$$

Find the new mean.

$$s_n^2 = \frac{6539}{13} - \left(\frac{280}{13}\right)^2 = 39.0946...$$

$$\therefore s_n = \sqrt{39.0946...} = 6.25 \text{ (3 SF)}$$

Find the new variance and take its square root to find the standard deviation.

For a question involving several steps, do not use a rounded value before reaching the final answer.

Practice questions 11.1

1. The ages, y years, of 42 children are summarised by $\sum y_i = 483$ and $\sum y_i^2 = 6015$. Find the mean and the standard deviation of the children's ages.

2. The mean height of a group of 15 basketball players is 207.2 cm. When another player joins the group, the mean height decreases to 206.5 cm. Find the height of the new player.

3. In her first five attempts at long jump, Greta's mean jump length was 4.80 m and the standard deviation of the lengths was 0.2 m. After her sixth jump the mean increased to 4.85 m. Find the standard deviation of all six jumps.

11.2 FREQUENCY TABLES AND GROUPED DATA

WORKED EXAMPLE 11.2

The heights of 50 trees (measured to the nearest metre) are summarised below.

Height (m)	2–5	6–10	11–15	16–22	23–30
Frequency	5	11	17	14	3

Estimate the mean and variance of the heights.

h_i	3.5	8	13	19	26.5
f_i	5	11	17	14	3

Use the mid-interval value for each group.

From GDC:

$\bar{h} = 13.4$ cm (3 SF)

$s_n = 5.9453\ldots$

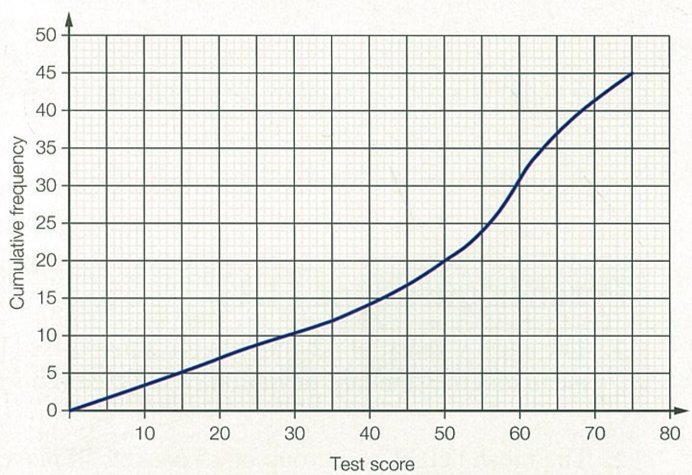
$\Rightarrow s_n^2 = 35.3$ cm^2 (3 SF)

When using a calculator in statistics questions, make sure you show the numbers being entered into your GDC.

Practice questions 11.2

4. The cumulative frequency diagram shows the test scores of a group of students.

Test score, S	Frequency
$0 \le S \le 20$	7
$20 < S \le 35$	q
$35 < S \le 50$	r
$50 < S \le p$	17
$p < S \le 75$	8



(a) Estimate the median and the interquartile range of the scores.

(b) Find the values of p, q and r to complete the frequency table.

(c) Hence estimate the mean and standard deviation of the scores.

5. The results of a group of students on a mathematics test are summarised below.

Score	20–30	31–40	41–55	56–70	71–82	83–100
Frequency	6	13	k	25	11	9

(a) Given that the mean score is 59 (rounded to the nearest integer), find the value of k.

(b) Find the standard deviation of the results.

11.3 CALCULATING PROBABILITIES BY CONSIDERING POSSIBLE OUTCOMES

WORKED EXAMPLE 11.3

A cubical die has the numbers 1, 1, 2, 3, 3, 3 written on its faces. The die is rolled twice. Find the probability that the sum of the two scores is greater than 4.

The possible sums are $S=5$ and $S=6$.

$$P(S=6)=P(3\cap 3)$$
$$=P(3)\times P(3)$$
$$=\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$$

The scores on the two rolls of the die are independent, so $P(3\cap 3)=P(3)\times P(3)$.

There are two ways of getting a sum of 5: $2+3$ and $3+2$.

$$P(S=5)=P(2 \text{ then } 3) \text{ OR } (3 \text{ then } 2)$$
$$=(P(2)\times P(3))+(P(3)\times P(2))$$
$$=\frac{1}{6}\times\frac{1}{2}+\frac{1}{2}\times\frac{1}{6}=\frac{1}{6}$$

The event 'a 2 followed by a 3' has to be counted as a separate event from 'a 3 followed by a 2'.

The total probability is $\frac{1}{4}+\frac{1}{6}=\frac{5}{12}$.

The two events $S=5$ and $S=6$ are mutually exclusive, so we use $P(A\cup B)=P(A)+P(B)$.

Practice questions 11.3

6. The random variable X has probability distribution as shown in the table below.

x	1	2	3
$P(X=x)$	0.4	0.3	0.3

Find the probability that the sum of two independent observations of X is 4.

7. Seven students are randomly arranged in a line. Find the probability that Anne and Beth are standing next to each other.

> Counting principles can be used to calculate probabilities. Counting principles are covered in Chapter 1.

8. A bag contains seven caramels and one chocolate. Three children, Peng, Quinn and Raul, take turns to pick a sweet out of the bag at random. If the sweet is a chocolate, they take it; if it is a caramel, they put it back and pass the bag around.

(a) Find the probability that Raul gets the chocolate on his second turn.

(b) Find the probability that the chocolate is still in the bag when it gets to Quinn for the fifth time.

11.4 VENN DIAGRAMS

WORKED EXAMPLE 11.4

The events A and B are such that $P(A \cup B) = \frac{3}{5}$, $P(B) = \frac{2}{5}$ and $P(B \mid A) = \frac{3}{7}$.

(a) State, with a reason, whether A and B are independent.

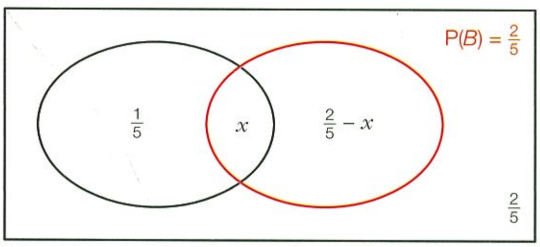
(b) Find $P(A)$.

(a) The events A and B are not independent because $P(B|A) \neq P(B)$.

For independent events, knowing that A has occurred has no impact on the probability of B occurring, i.e. $P(B \mid A) = P(B)$.

(b) Let $P(A \cap B) = x$. Then $P(B \cap A') = \frac{2}{5} - x$ and

$$P(A \cap B') = \frac{3}{5} - \left(\frac{2}{5} - x\right) - x = \frac{1}{5}$$



Venn diagrams are a useful way of representing information when you are given information about the union. Label the intersection x and work outwards.

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

$$\therefore \frac{3}{7} = \frac{x}{\frac{1}{5} + x}$$

$$\Rightarrow x = \frac{3}{20}$$

Hence $P(A) = \frac{1}{5} + \frac{3}{20} = \frac{7}{20}$.

For conditional probability we can use the formula $P(B \mid A) = \frac{P(B \cap A)}{P(A)}$.

Practice questions 11.4

9. Given that $P(B) = 0.5$, $P(A' \cap B') = 0.2$ and $P(B \mid A) = 0.4$, find $P(A)$.

10. All of the 100 students at a college take part in at least one of three activities: chess, basketball and singing. 10 play both chess and basketball, 12 play chess and sing, and 7 take part in both singing and basketball. 40 students play basketball, 62 play chess and 22 sing.

(a) How many students take part in all three activities?

(b) A student is chosen at random. Given that this student plays basketball and sings, what is the probability that she also plays chess?

11.5 BAYES' THEOREM AND TREE DIAGRAMS

WORKED EXAMPLE 11.5

The events A and B are such that $P(A)=\frac{1}{3}$, $P(B \mid A)=\frac{1}{3}$ and $P(B \mid A')=\frac{1}{2}$. Find $P(A \mid B)$.

This type of question can be done using Bayes' theorem or by drawing a tree diagram.

$$P(A|B)=\frac{P(A)P(B|A)}{P(A)P(B|A)+P(A')P(B|A')}$$

$$=\frac{\frac{1}{3}\times\frac{1}{3}}{\frac{1}{3}\times\frac{1}{3}+\frac{2}{3}\times\frac{1}{2}}=\frac{1}{4}$$

Write out Bayes' theorem in terms of the events mentioned.

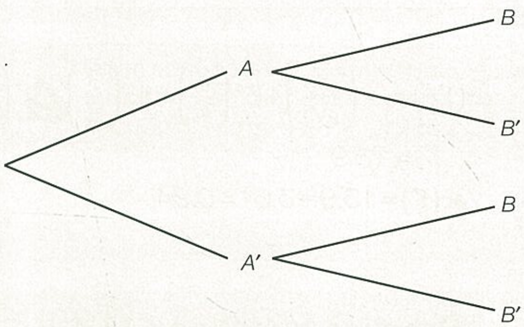
Practice questions 11.5

11. The events A and B are such that $P(B \mid A)=0.2$, $P(B' \mid A')=0.3$, $P(A \mid B)=0.4$ and $P(A)=x$.

(a) Complete the following tree diagram.

(b) Find the value of x.

(c) State, with a reason, whether the events A and B are independent.



12. Given that $P(B)=0.3$, $P(A \mid B)=0.6$ and $P(A \mid B')=0.4$, find $P(B \mid A)$.

13. Every morning I either walk or cycle to school, with equal probability. If I walk, the probability that I am late is 0.2. If I cycle, the probability that I am late is 0.4. Given that I was late for school yesterday, what is the probability that I walked?

14. A large box contains three different types of toys. One third of the toys are cars, one quarter are yo-yos and the rest are balloons. 20% of the cars, 30% of the yo-yos and 40% of the balloons are pink. A toy is selected at random from the box. Given that the toy is pink, find the probability that it is a balloon.

11.6 EXPECTATION AND VARIANCE OF DISCRETE RANDOM VARIABLES

WORKED EXAMPLE 11.6

A discrete random variable Y has probability distribution as shown in the table below.

y	2	3	4	5
$P(Y=y)$	k	$\frac{2}{5}$	$\frac{1}{4}$	$2k$

(a) Find the exact value of k.

(b) Find the variance of Y.

(a) $k+\frac{2}{5}+\frac{1}{4}+2k=1 \Rightarrow k=\frac{7}{60}$

We use the fact that the probabilities must add up to 1.

(b) $E(Y)=2\left(\frac{7}{60}\right)+3\left(\frac{2}{5}\right)+4\left(\frac{1}{4}\right)+5\left(\frac{14}{60}\right)=3.6$

First, find the expectation using the formula $E(Y)=\sum yP(Y=y)$.

$$E(Y^2)=2^2\left(\frac{7}{60}\right)+3^2\left(\frac{2}{5}\right)+4^2\left(\frac{1}{4}\right)+5^2\left(\frac{14}{60}\right)$$

$$=13.9$$

$$\text{Var}(Y)=13.9-3.6^2=0.94$$

To find the variance, use the formula $\text{Var}(Y)=E(Y^2)-[E(Y)]^2$.

Practice questions 11.6

15. Find the expected value and the variance of the following discrete random variable.

z	11	13	17	19	23
$P(Z=z)$	0.2	0.1	0.1	0.2	0.4

16. The random variable X has probability distribution as shown in the table below.

x	1	2	3	4
$P(X=x)$	c	p	0.4	0.2

Given that $E(X)=2.6$, find the values of c and p.

11.7 THE BINOMIAL AND POISSON DISTRIBUTIONS

WORKED EXAMPLE 11.7

Sandra receives 8 emails per day on average. It is assumed that the number of emails follows a Poisson distribution.

(a) Find the probability that Sandra receives more than 10 emails in one day.

(b) Find the probability that Sandra receives more than 10 emails on two out of seven days.

(a) Let X = number of emails received in a day.
Then $X \sim \text{Po}(8)$.

We start by defining the random variable and stating the probability distribution.

$$\begin{aligned} P(X>10) &= 1-P(X\leq 10) \\ &= 1-0.81588\ldots \text{ (from GDC)} \\ &= 0.184 \text{ (3 SF)} \end{aligned}$$

Write down the probability required. To use the calculator we must relate it to $P(X \leq k)$.

(b) Let Y = number of days out of 7 with more than 10 emails. Then $Y \sim B(7, 0.184)$.
$P(Y=2) = 0.257$ (3 SF) from GDC

There is a fixed number of days, so we need to use the binomial distribution this time.

When using your GDC, always state the distribution used and the probability calculated, and give the answer to 3 SF.

Practice questions 11.7

17. A fair six-sided die is rolled 16 times. Find the probability it lands on a '4' more than five times.

18. During the winter months, snowstorms occur at a constant rate of 1.2 per week, independently of each other.

(a) Find the probability that no snowstorms occur in a given week.

(b) Find the probability that in seven consecutive weeks there is at least one week with no snowstorms.

19. The random variable X has distribution $B(n, p)$. The mean of X is equal to three times its variance, and $P(X=2) = 0.0384$ (to three significant figures). Find the values of n and p.

20. Rebekah recorded the number of cars in the school car park over a period of 14 days. The results are summarised by $\sum x_i = 245$ and $\sum x_i^2 = 4462$.

(a) Find the mean and variance of the number of cars in the car park.

(b) State, with a reason, whether a Poisson distribution would be an appropriate model for the number of cars in the car park.

11.8 EXPECTATION AND VARIANCE OF CONTINUOUS RANDOM VARIABLES

WORKED EXAMPLE 11.8

A continuous random variable X has probability density function

$$f(x)=\begin{cases}3x^2+\frac{7}{4} & \text{for } 0\le x\le\frac{1}{2}\\ 0 & \text{otherwise}\end{cases}$$

(a) Find the value of q such that $P(X<q)=0.25$.

(b) Given that $E(X)=\frac{17}{64}$, calculate $Var(X)$.

(a) $\int_0^q 3x^2+\frac{7}{4}\,dx=0.25$

$P(X<q)$, or equivalently the probability that X lies in $[0, q]$, is given by $\int_0^q f(x)\,dx$.

$\left[x^3+\frac{7}{4}x\right]_0^q=0.25$

Make sure you know how to use your GDC to solve polynomial equations.

$\Rightarrow q^3+\frac{7}{4}q-0.25=0$

$\Rightarrow q=0.141$ (from GDC)

(b) $Var(X)=\int_0^{\frac{1}{2}} 3x^4+\frac{7x^2}{4}\,dx-\left(\frac{17}{64}\right)^2$

$=0.0211$ (from GDC)

$Var(X)=E(X^2)-[E(X)]^2$ where $E(X^2)=\int_{-\infty}^{\infty} x^2 f(x)\,dx$.

We can use the GDC to evaluate the integral.

Practice questions 11.8

21. The continuous random variable X is defined for $-2\le x\le 2$ and has the probability density function $f(x)=ke^{2x}$.

(a) Find the exact value of k.

(b) Find the median of X.

The median, m, of a continuous random variable satisfies $\int_{-\infty}^{m} f(x)\,dx=\frac{1}{2}$. The mode is the value of x at the maximum of $f(x)$.

22. A continuous random variable X has probability density function $f(x)=\frac{1}{9}(4-x^2)$ for $x\in[-1,2]$.

(a) Show that $E(X)=\frac{1}{4}$.

(b) Find the probability that X takes on a value between the mean and the mode.

11.9 THE NORMAL AND INVERSE NORMAL DISTRIBUTIONS

WORKED EXAMPLE 11.9

Test scores are normally distributed with mean 62 and standard deviation 12.

(a) Find the percentage of candidates who scored below 50.

(b) It is known that 44% of candidates scored between 62 and p. Find the value of p.

(a) Let X = test score, so $X \sim N(62, 12^2)$.

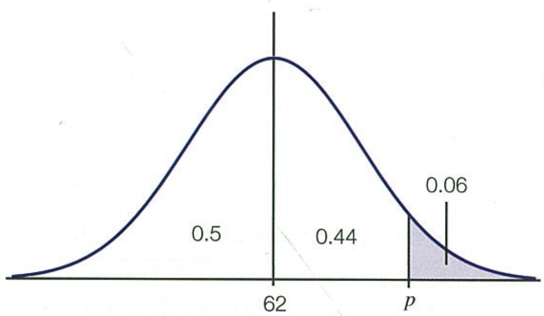
We start by defining the random variable and stating the distribution used.

$P(X < 50) = 0.159$ (3 SF) from GDC

So 15.9% of candidates scored below 50.

For a normal distribution you must be able to use a GDC to find a probability.

(b) $P(62 < X < p) = 0.44$



Sketching a diagram allows us to see more clearly exactly what we need to find from the GDC.

As the probability is known, we need to use the inverse normal distribution.

On some calculators it is necessary to change problems like this into the form $P(X \le x) = p$ (i.e. a left-tail calculation).

From the diagram, $P(X > p) = 0.06$

From GDC, $p = 80.7$ (3 SF)

Practice questions 11.9

23. The random variable Y follows a normal distribution with mean 7.5 and variance 1.44. Find:

(a) $P(6 < Y < 7)$

(b) $P(Y \ge 8.5)$

(c) the value of k such that $P(Y \le k) = 0.35$.

24. The weights, W kg, of babies born at a certain hospital satisfy $W \sim N(3.2, 0.7^2)$. Find the value of m such that 35% of the babies weigh between m kg and 3.2 kg.

25. The time a laptop battery can last before needing to be recharged is assumed to be normally distributed with mean 4 hours and standard deviation 20 minutes.

(a) Find the probability that a laptop battery will last more than 4.5 hours.

(b) A manufacturer wants to ensure that 95% of batteries will last for $4 \pm x$ hours. Find x.

11.10 USING THE STANDARD NORMAL DISTRIBUTION WHEN μ OR σ ARE UNKNOWN

WORKED EXAMPLE 11.10

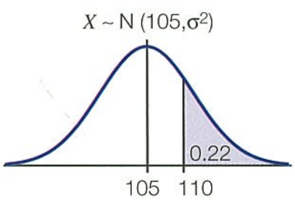
It is known that the average height of six-year-old boys is 105 cm and that 22% of the boys are taller than 110 cm. Find the standard deviation of the heights.

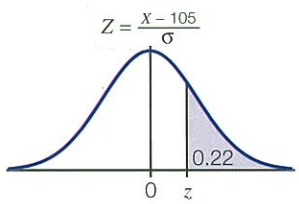
Let X = height of a six-year-old boy.
Then $X \sim N(105, \sigma^2)$.

> We start by defining the random variable and stating the distribution.

The standardised variable is $Z = \dfrac{X-105}{\sigma}$ and $Z \sim N(0,1)$.

> As σ is unknown, we need to use the standard normal distribution, $Z \sim N(0, 1)$ where $Z = \dfrac{X-\mu}{\sigma}$.





> It is always a good idea to sketch a normal distribution diagram to help you visualise the solution.

$P(Z > z) = 0.22$
$\Rightarrow z = 0.7722$ (from GDC)

> As the probability is known (22% = 0.22), we need to use the inverse normal distribution.

$z = \dfrac{x-105}{\sigma}$

$\therefore 0.7722 = \dfrac{110-105}{\sigma}$

$\Rightarrow \sigma = 6.48$ cm

> We can now find σ from z.

Practice questions 11.10

26. The weights of apples sold at a market are normally distributed with mean weight 125 g. It is found that 26% of the apples weigh less than 116 g. Find the standard deviation of the weights.

27. A machine dispenses cups of coffee. The volume of coffee in a cup is normally distributed with standard deviation 5.6 ml. If 10% of cups contain more than 160 ml, find to one decimal place the mean volume of coffee in a cup.

28. It is known that the scores on a test follow a normal distribution $N(\mu, \sigma^2)$. 20% of the scores are above 82 and 10% are below 47.

(a) Show that $\mu + 0.8416\sigma = 82$.

(b) By writing a similar equation, find the mean and standard deviation of the scores.

Mixed practice 11

1. The heights of trees in a forest are normally distributed with mean height 26.2 m and standard deviation 5.6 m.

(a) Find the probability that a tree is more than 30 m tall.

(b) What is the probability that among 16 randomly selected trees at least 2 are more than 30 m tall?

2. A discrete random variable X is given by $P(X = n) = kn^2$ for $n = 1, 2, 3, 4$.

Find the expected value and variance of X.

3. If $P(A) = 0.3$, $P(B \mid A') = 0.5$ and $P(A \mid B) = \frac{7}{16}$, find $P(B \mid A)$.

4. The discrete random variable Y has probability distribution as shown in the table below.

y	1	2	3	4
$P(Y = y)$	0.1	0.2	p	q

Given that $E(Y) = 3.1$, find the values of p and q.

5. A continuous random variable X has probability density function $f(x) = k(4 - x^2)$ for $0 \le x \le 2$.

(a) Find the value of k.

(b) Show that the median satisfies $m^3 - 12m + 8 = 0$.

6. A cat visits my garden at random, at the constant average rate of four times a day.

(a) What is the probability that the cat visits my garden at least once on a given day?

(b) What is the probability that the cat visits my garden at least once every day in a seven-day week?

(c) Given that the cat has already visited my garden once today, what is the probability that it will visit the garden at least five times?

7. A continuous random variable X has probability density function

$$f(x) = \begin{cases} ke^{-x^2} & \text{for } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find:

(i) the value of k

(ii) the lower quartile

(iii) the mean of X.

(b) Two independent observations of X are recorded.

(i) Find the probability that one of them is above and the other is below the mean.

(ii) Given that exactly one of the observations is above the mean, find the probability that it is the first one.

8. A company hires out vans on a daily basis. It has three vans it can hire out. The number of requests it gets for hiring a van can be modelled by a Poisson distribution with a mean of 1.8 requests per day.

(a) Find the probability that in one day some requests have to be turned down.

(b) Given that some requests have to be turned down, find the probability that there were exactly four requests.

(c) Find the probability that there are more than six requests in two days.

(d) Find the probability that in a seven-day week there are at least two days in which requests are rejected.

(e) Find the probability distribution of the number of vans which are hired out each day.

(f) The price of hiring a van is $120. Find the expected daily takings of the company.

(g) The number of kilometres travelled by each van can be modelled by a normal distribution with mean 150 km. 10% of vans travel more than 200 km. Find the standard deviation of the normal distribution.

(h) If two vans are hired, find the probability that each travels less than 100 km.

Going for the top 11

1. The continuous random variable X has probability density function

$f(x) = kx\sin x$ for $x \in [0, \pi]$.

(a) Show that $k = \frac{1}{\pi}$.

(b) Find the interquartile range of X.

2. The random variable X has a Poisson distribution with mean $\lambda > 1$.
Given that $P(X = 0) + P(X = 2) = 3P(X = 1)$, find the exact value of λ.

3. Three basketball players, Annie, Brent and Carlos, try to shoot a free throw. Annie shoots first, then Brent, then Carlos. The probability that Annie scores is 0.6, the probability that Brent scores is 0.5, and the probability that Carlos scores is 0.8. The shots are independent of each other and the first player to score wins.

(a) Find the probability that Annie wins with her second shot.

(b) What is the probability that Carlos gets a second shot?

(c) (i) Show that the probability of Brent winning with his kth shot is $0.2 \times 0.04^{k-1}$.

(ii) Hence find the probability that Brent wins.

6. (a) (i) $$\sqrt{\frac{1-3x}{1+3x}} = \sqrt{\frac{1-3x}{1+3x}} \times \sqrt{\frac{1-3x}{1-3x}}$$
$$= \frac{\left(\sqrt{1-3x}\right)^2}{\sqrt{(1+3x)(1-3x)}}$$
$$= \frac{1-3x}{\sqrt{1-9x^2}}$$

(ii) $$\int \sqrt{\frac{1-3x}{1+3x}}\, dx = \int \frac{1-3x}{\sqrt{1-9x^2}}\, dx$$
$$= \int \frac{1}{\sqrt{1-9x^2}}\, dx - \int \frac{3x}{\sqrt{1-9x^2}}\, dx$$
$$= \frac{1}{3}\int \frac{1}{\sqrt{\left(\frac{1}{3}\right)^2 - x^2}}\, dx - \left(\frac{1}{-6}\right)\int \frac{-18x}{\sqrt{1-9x^2}}\, dx$$
$$= \frac{1}{3}\arcsin\left(\frac{x}{\frac{1}{3}}\right) + \frac{2}{6}\sqrt{1-9x^2} + C$$
$$= \frac{1}{3}\left(\arcsin(3x) + \sqrt{1-9x^2}\right) + C$$

(b) (i) $$\int \sec^3 x\, dx = \int \sec x \sec^2 x\, dx$$

Let $u = \sec x$ and $\frac{dv}{dx} = \sec^2 x$

Then $\frac{du}{dx} = \sec x \tan x$ and $v = \tan x$

$$\int \sec x \sec^2 x\, dx = \sec x \tan x - \int \sec x \tan x \tan x\, dx$$
$$= \sec x \tan x - \int \sec x \tan^2 x\, dx$$
$$= \sec x \tan x - \int \sec x\left(\sec^2 x - 1\right) dx$$
$$= \sec x \tan x - \int \sec^3 x - \sec x\, dx$$
$$= \sec x \tan x - \int \sec^3 x\, dx + \int \sec x\, dx$$
$$= \sec x \tan x - \int \sec^3 x\, dx + \ln|\sec x + \tan x| + C$$

$$\therefore 2\int \sec^3 x\, dx = \sec x \tan x + \ln|\sec x + \tan x| + C$$

Hence $$\int \sec^3 x\, dx = \frac{1}{2}\left(\sec x \tan x + \ln|\sec x + \tan x|\right) + C$$

(ii) Let $x = \sqrt{3}\tan\theta$

Then $\frac{dx}{d\theta} = \sqrt{3}\sec^2\theta \Rightarrow dx = \sqrt{3}\sec^2\theta\, d\theta$

To change the limits:

At $x = 0$, $0 = \sqrt{3}\tan\theta \Rightarrow \theta = 0$

At $x = 1$, $1 = \sqrt{3}\tan\theta \Rightarrow \tan\theta = \frac{1}{\sqrt{3}}$

$$\Rightarrow \theta = \frac{\pi}{6}$$

So $$\int_0^1 \sqrt{x^2+3}\, dx = \int_0^{\pi/6} \sqrt{(\sqrt{3}\tan\theta)^2 + 3}\left(\sqrt{3}\sec^2\theta\, d\theta\right)$$
$$= \int_0^{\pi/6} \sqrt{3\tan^2\theta + 3}\left(\sqrt{3}\sec^2\theta\, d\theta\right)$$
$$= \int_0^{\pi/6} \sqrt{3}\sqrt{\tan^2\theta + 1}\left(\sqrt{3}\sec^2\theta\, d\theta\right)$$
$$= 3\int_0^{\pi/6} \sec^3\theta\, d\theta$$
$$= 3\left[\frac{1}{2}\left(\sec\theta\tan\theta + \ln|\sec\theta + \tan\theta|\right)\right]_0^{\frac{\pi}{6}}$$
$$= \frac{3}{2}\left[\left(\sec\frac{\pi}{6}\tan\frac{\pi}{6} + \ln\left|\sec\frac{\pi}{6} + \tan\frac{\pi}{6}\right|\right) - \left(\sec 0 \tan 0 + \ln|\sec 0 + \tan 0|\right)\right]$$
$$= \frac{3}{2}\left(\frac{2}{3} + \ln\sqrt{3}\right)$$
$$= \frac{3}{2}\left(\frac{2}{3} + \frac{1}{2}\ln 3\right)$$
$$= 1 + \frac{3}{4}\ln 3$$

11 PROBABILITY AND STATISTICS

Mixed practice 11

1. (a) Let X = height of a tree. Then $X \sim N(26.2, 5.6^2)$.

$$P(X > 30) = 1 - P(X < 30)$$
$$= 1 - 0.75129\ldots \quad \text{(from GDC)}$$
$$= 0.249$$

(b) Let Y = the number of trees out of the 16 that are more than 30 m tall.

Then $Y \sim B(16, 0.249)$.

$$P(Y \geq 2) = 1 - P(Y \leq 1)$$
$$= 1 - 0.06487\ldots \quad \text{(from GDC)}$$
$$= 0.935$$

2. Probability distribution of X:

x	1	2	3	4
$P(X = x)$	k	$4k$	$9k$	$16k$

$$k + 4k + 9k + 16k = 1 \Leftrightarrow k = \frac{1}{30}$$

$$E(X) = 1(k) + 2(4k) + 3(9k) + 4(16k)$$
$$= 100k = \frac{100}{30} = 3.33$$

$$\text{Var}(X) = [1(k)+4(4k)+9(9k)+16(16k)] - \left(\frac{100}{30}\right)^2$$

$$= \frac{354}{30} - \left(\frac{100}{30}\right)^2$$

$$= 0.689$$

3. By Bayes' theorem:

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(A)P(B \mid A) + P(A')P(B \mid A')}$$

$$\therefore \frac{7}{16} = \frac{0.3P(B \mid A)}{0.3P(B \mid A) + 0.7 \times 0.5}$$

$$\Leftrightarrow 2.1P(B \mid A) + 2.45 = 4.8P(B \mid A)$$

$$\Leftrightarrow 2.7P(B \mid A) = 2.45$$

$$\Leftrightarrow P(B \mid A) = \frac{49}{54}$$

4. $\sum p_i = 1$

$$\Rightarrow 0.3 + p + q = 1$$

$$\Rightarrow p + q = 0.7 \quad \cdots (1)$$

$$E(Y) = 3.1$$

$$\Rightarrow 0.1 + 0.4 + 3p + 4q = 3.1$$

$$\Rightarrow 3p + 4q = 2.6 \quad \cdots (2)$$

Solving equations (1) and (2) simultaneously gives $p = 0.2$, $q = 0.5$.

5. (a) $\int_0^2 k(4 - x^2)\,dx = 1$

$$\Rightarrow k\left[4x - \frac{x^3}{3}\right]_0^2 = 1$$

$$\Rightarrow k\left(\frac{16}{3}\right) = 1$$

$$\Rightarrow k = \frac{3}{16}$$

(b) $\int_0^m \frac{3}{16}(4 - x^2)\,dx = \frac{1}{2}$

$$\Rightarrow \frac{3}{16}\left(4m - \frac{m^3}{3}\right) = \frac{1}{2}$$

$$\Rightarrow 12m - m^3 = 8$$

$$\Rightarrow m^3 - 12m + 8 = 0$$

6. Let X = the number of times the cat visits my garden in one day.
Then $X \sim \text{Po}(4)$.

(a) $P(X \geq 1) = 1 - P(X = 0)$

$$= 1 - 0.018315...$$

$$= 0.98168... \approx 0.982$$

(b) $(0.98168\ldots)^7 = 0.879$

(c) $P(X \geq 5 \mid X \geq 1) = \frac{P(X \geq 5)}{P(X \geq 1)}$

$$= \frac{1 - P(X \leq 4)}{0.982}$$

$$= \frac{1 - 0.629}{0.982}$$

$$= 0.378$$

7. (a) (i) $\int_0^2 k e^{-x^2}\,dx = 1$

From GDC, $\int_0^2 e^{-x^2}\,dx = 0.882$

$$\therefore k = \frac{1}{0.882} = 1.13$$

(ii) Let q_1 be the lower quartile; then

$$\int_0^{q_1} k e^{-x^2}\,dx = 0.25$$

From GDC, $q_1 = 0.224$.

(iii) $\mu = E(X) = \int_0^2 kxe^{-x^2}\,dx = 0.556$

(3 SF, from GDC)

(b) (i) $P(X > \mu) = \int_{0.556}^2 k e^{-x^2}\,dx = 0.429$

$$P(X < \mu) = \int_0^{0.556} k e^{-x^2}\,dx = 0.571$$

P(one above and one below the mean)

$$= P\Big(\big((X_1 > \mu) \cap (X_2 < \mu)\big) \cup \big((X_1 < \mu) \cap (X_2 > \mu)\big)\Big)$$

$$= (0.429 \times 0.571) + (0.571 \times 0.429)$$

$$= 0.490$$

(ii) $P(X_1 > \mu \mid \text{one above and one below the mean})$

$$= \frac{P\big((X_1 > \mu) \cap (X_2 < \mu)\big)}{P(\text{one above and one below the mean})}$$

$$= \frac{0.429 \times 0.571}{(0.429 \times 0.571) + (0.571 \times 0.429)} = 0.5$$

8. (a) Let X = number of requests in one day.
Then $X \sim \text{Po}(1.8)$.

Some requests have to be turned down if there are more than 3 requests in a day.

$$P(X > 3) = 1 - P(X \leq 3)$$

$$= 1 - 0.891$$

$$= 0.109$$

(b) $P(X = 4 \mid X > 3) = \frac{P\big((X = 4) \cap (X > 3)\big)}{P(X > 3)}$

$$= \frac{P(X = 4)}{P(X > 3)}$$

$$= \frac{0.0723}{0.109} = 0.665$$

(c) Let Y = number in requests in two days. Then $Y \sim \text{Po}(3.6)$.

$$\text{P}(Y > 6) = 1 - \text{P}(Y \leq 6)$$
$$= 1 - 0.9267 = 0.0733$$

(d) Let Z = number of days in a 7-day week with more than 3 requests.

Then $Z \sim \text{B}(7, 0.109)$.

$$\text{P}(Z \geq 2) = 1 - \text{P}(Z \leq 1)$$
$$= 1 - 0.828 = 0.172$$

(e) Let N = number of vans hired out in one day.

When the number of requests is 3 or fewer, $N = X$; when the number of requests is more than 3, $N = 3$. Therefore:

$\text{P}(N = 0) = \text{P}(X = 0) = 0.165$

$\text{P}(N = 1) = \text{P}(X = 1) = 0.298$

$\text{P}(N = 2) = \text{P}(X = 2) = 0.268$

$\text{P}(N = 3) = \text{P}(X = 3) + \text{P}(X > 3)$
$= 0.161 + 0.109 = 0.269$

So the probability distribution of N is:

N	0	1	2	3
$\text{P}(N = n)$	0.165	0.298	0.268	0.269

(f) Let I = the income in one day; then $I = 120N$, so the probability distribution of I is:

I	0	120	240	360
$\text{P}(I = i)$	0.165	0.298	0.268	0.269

$\text{E}(I) = 0 \times 0.165 + 120 \times 0.298 + 240 \times 0.268 + 360 \times 0.269 = \$197$

(g) Let D = the distance travelled by a van.

Then $D \sim \text{N}(150, \sigma^2)$ and $\text{P}(D > 200) = 0.1$.

$$\frac{D-150}{\sigma} \sim \text{N}(0,1)$$

$$\text{P}\left(\frac{D-150}{\sigma} > \frac{200-150}{\sigma}\right) = 0.1$$

$$\Rightarrow \frac{200-150}{\sigma} = 1.28 \quad \text{(from GDC)}$$

$$\Rightarrow \sigma = 39.0 \text{ km}$$

(h) $D \sim \text{N}(150, 39.0^2)$, so $\text{P}(D < 100) = 0.100$

P(two vans each travel less than 100 km) = $0.100^2 = 0.0100$

Going for the top 11

1. (a) $\int_0^{\pi} kx \sin x \, dx = 1$

Using integration by parts:

$$k\left(\left[-x\cos x\right]_0^{\pi} - \int_0^{\pi} -\cos x \, dx\right) = 1$$

$$\Rightarrow k\left(-\pi(-1) + \left[\sin x\right]_0^{\pi}\right) = 1$$

$$\Rightarrow k\pi = 1$$

$$\Rightarrow k = \frac{1}{\pi}$$

(b) Let q_1 = first quartile, q_3 = third quartile. Then:

$$\int_0^{q_1} \frac{1}{\pi} x \sin x \, dx = 0.25$$

$$\Rightarrow \frac{1}{\pi}\left[-x\cos x + \sin x\right]_0^{q_1} = 0.25$$

$$\Rightarrow -q_1 \cos q_1 + \sin q_1 = 0.25\pi$$

$$\Rightarrow q_1 = 1.43 \quad \text{(from GDC)}$$

$$\int_0^{q_3} \frac{1}{\pi} x \sin x \, dx = 0.75$$

$$\Rightarrow \frac{1}{\pi}\left[-x\cos x + \sin x\right]_0^{q_3} = 0.75$$

$$\Rightarrow -q_3 \cos q_3 + \sin q_3 = 0.75\pi$$

$$\Rightarrow q_3 = 2.35 \quad \text{(from GDC)}$$

$\text{IQR} = q_3 - q_1 = 0.919$

2. $X \sim \text{Po}(\lambda)$, $\text{P}(X = k) = \dfrac{e^{-\lambda}\lambda^k}{k!}$

$$\text{P}(X = 0) + \text{P}(X = 2) = 3\text{P}(X = 1)$$

$$\Leftrightarrow \frac{e^{-\lambda}\lambda^0}{0!} + \frac{e^{-\lambda}\lambda^2}{2!} = 3 \times \frac{e^{-\lambda}\lambda^1}{1!}$$

$$\Leftrightarrow 1 + \frac{\lambda^2}{2} = 3\lambda$$

$$\Leftrightarrow \lambda^2 - 6\lambda + 2 = 0$$

$$\Leftrightarrow \lambda = \frac{6 \pm \sqrt{36-8}}{2}$$

$$= \frac{6 \pm \sqrt{28}}{2}$$

$$= \frac{6 \pm 2\sqrt{7}}{2}$$

$$= 3 \pm \sqrt{7}$$

$$\therefore \lambda = 3 + \sqrt{7} \quad (\text{as } \lambda > 1)$$

3. (a) For Annie to win with her second shot, Annie, Brent and Carlos have to miss once and then Annie scores:

$$0.4 \times 0.5 \times 0.2 \times 0.6 = 0.024$$

(b) For Carlos to get a second shot, he has to miss once and Annie and Brent have to miss twice:

$0.4^2 \times 0.5^2 \times 0.2 = 0.008$

(c) (i) For Brent to win with his kth shot, he has to miss $(k-1)$ times and score once, Annie has to miss k times, and Carlos has to miss $(k-1)$ times:

$$0.5^{k-1} \times 0.5 \times 0.4^k \times 0.2^{k-1} = (0.5 \times 0.4 \times 0.2)^{k-1} \times (0.5 \times 0.4) = 0.04^{k-1} \times 0.2$$

(ii) $\mathrm{P}(\text{Brent wins})$

$$= \sum_{k=1}^{\infty} \mathrm{P}(\text{Brent wins on } k\text{th shot})$$

$$= \sum_{k=1}^{\infty} 0.2 \times 0.04^{k-1}$$

Using the formula for the sum of the geometric series with $u_1 = 0.2$ and $r = 0.04$:

$$\mathrm{P}(\text{Brent wins}) = \frac{0.2}{1-0.04} = 0.208$$

12 INDUCTION

Mixed practice 12

1. Writing odd numbers as $2r-1$ for $r \geq 1$, we need to prove that $\sum_{r=1}^{n} 2r-1 = n^2$.

Let $n = 1$:

LHS $= 2 \times 1 - 1 = 1$

RHS $= 1^2 = 1$

So the statement is true for $n = 1$.

Assume it is true for $n = k$:

$$\sum_{r=1}^{k} 2r-1 = k^2 \quad \cdots (*)$$

Let $n = k+1$:

$$\sum_{r=1}^{k+1} 2r-1 = \left(\sum_{r=1}^{k} 2r-1\right) + 2(k+1) - 1$$
$$= \left(\sum_{r=1}^{k} 2r-1\right) + 2k+1$$
$$= (k^2) + 2k + 1 \qquad \text{(by (*))}$$
$$= (k+1)^2$$

So the statement is true for $n = k+1$.

The statement is true for $n = 1$, and if true for $n = k$ it is also true for $n = k+1$. Hence, the statement is true for all $n \geq 1$ by the principle of mathematical induction.

2. Let $n = 0$:

$3^{4n+2} + 2^{6n+3} = 3^2 + 2^3 = 17 = 17 \times 1$

So the statement is true for $n = 0$.

Assume it is true for $n = k$:

$3^{4k+2} + 2^{6k+3} = 17A$ for some $A \in \mathbb{Z}$ $\quad \cdots (*)$

Let $n = k+1$:

$$3^{4(k+1)+2} + 2^{6(k+1)+3} = 3^{4k+6} + 2^{6k+9}$$
$$= 3^4(3^{4k+2}) + 2^{6k+9}$$
$$= 3^4(17A - 2^{6k+3}) + 2^{6k+9} \qquad \text{(by (*))}$$
$$= 81 \times 17A - 81 \times 2^{6k+3} + 2^{6k+9}$$
$$= 81 \times 17A - 2^{6k+3}(81 - 2^6)$$
$$= 81 \times 17A - 2^{6k+3}(17)$$
$$= 17(81A - 2^{6k+3})$$

So the statement is true for $n = k+1$.

The statement is true for $n = 0$, and if true for $n = k$ it is also true for $n = k+1$. Hence, the statement is true for all $n \geq 0$ by the principle of mathematical induction.

3. (a) $3^1 = 3 < 1 + 5$

$3^2 = 9 > 2 + 5$

So the smallest positive integer M is $M = 2$.

(b) The statement is true for $n = 2$ by part (a).

Assume it is true for $n = k$:

$3^k > k + 5$ where $k \geq 2$ $\quad \cdots (*)$

Let $n = k+1$:

$$3^{k+1} = 3(3^k)$$
$$> 3(k+5) \qquad \text{(by (*))}$$
$$= 3k + 15$$
$$= (k+1) + 5 + 2k + 9$$
$$> (k+1) + 5$$

So the statement is true for $n = k+1$.

The statement is true for $n = 2$, and if true for $n = k$ it is also true for $n = k+1$. Hence, the statement is true for all $n \geq 2$ by the principle of mathematical induction.

4. (a) Differentiating $f(x) = x\cos x$:

$$f'(x) = 1 \times \cos x + x(-\sin x)$$
$$= \cos x - x\sin x$$

$$f''(x) = -\sin x - (1 \times \sin x + x\cos x)$$
$$= -2\sin x - x\cos x$$

(b) Let $n = 1$:

LHS $= f^{(2\times 1)}(x) = f^{(2)}(x) = f''(x)$

RHS $= (-1)^1(x\cos x + 2 \times 1 \times \sin x) = -x\cos x - 2\sin x$

15. $k = -25$

16. (a) $v = 10 - 2e^{-t}$ (b) 10.2 s

11 PROBABILITY AND STATISTICS

1. Mean 11.5 years; standard deviation 3.31 years

2. 196 cm

3. 0.214 m

4. (a) Median ≈ 53.5; IQR ≈ 29

(b) $p = 65, q = 5, r = 8$

(c) Mean 46.3; standard deviation 19.9

5. (a) 7 (b) 19.9

6. 0.33

7. $\frac{2}{7}$

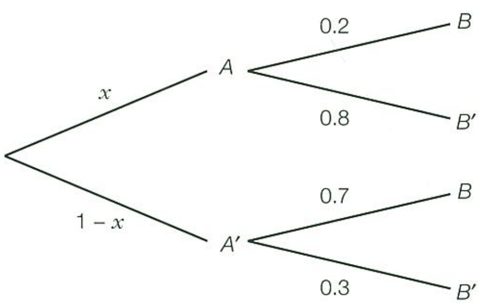
8. (a) $\left(\frac{7}{8}\right)^5\left(\frac{1}{8}\right) = \frac{16\,807}{262\,144} \approx 0.0641$

(b) $\left(\frac{7}{8}\right)^{13} \approx 0.176$

9. 0.5

10. (a) 5 (b) $\frac{5}{7}$

11. (a)



(b) 0.7

(c) No; $P(A) \neq P(A \mid B)$

12. $\frac{9}{23} \approx 0.391$

13. $\frac{1}{3}$

14. $\frac{20}{37}$

15. $E(Z) = 18.2$, $Var(Z) = 22.56$

16. $c = 0.2, p = 0.2$

17. 0.0378

18. (a) 0.301 (b) 0.919

19. $p = \frac{2}{3}, n = 7$

20. (a) Mean 17.5; variance 12.5

(b) Not appropriate; mean and variance should be equal for a Poisson distribution.

21. (a) $\frac{2}{e^4 - e^{-4}}$ (b) 1.654

22. (b) 0.111

23. (a) 0.233 (b) 0.202

(c) 7.04

24. 2.47 or 3.93

25. (a) 0.0668 (b) 0.653

26. 14.0 g

27. 152.8 ml

28. (b) $\mu - 1.28155\sigma = 47$; $\mu = 68.12, \sigma = 16.48$

13 EXAMINATION SUPPORT

Spot the common errors

1. (i) Tried to integrate a product factor by factor

(ii) Integrated e^{2x} to $2e^{2x}$ rather than $\frac{1}{2}e^{2x}$

(iii) Missed out '+ c'

2. (i) Expanded $\ln(10 - x)$ into two logarithms

(ii) Sign error in performing the 'expansion' led to $-\ln x$, which then cancelled with the first $\ln x$

(iii) Tried to undo ln without first getting everything on each side of the equation (including minus signs) inside a ln

(iv) Obtained an answer which cannot go into the original expression (log of a negative number)

3. 'Cancelled' an expression which is not a factor of the entire denominator