PROBABILITY DISTRIBUTIONS
ANSWERS

Interpreting Data

Discrete vs. Continuous
STOP AND CHECK (PAGE 6)

- Examples of continuous data could be height, weight, time, or other measurements. For example, the height of students or the time taken to run 100 metres.
- The Poisson and Binomial distributions use discrete data.

Expectation and Variance
STOP AND CHECK (PAGE 10)

- The expected value is the average (mean) value you'd get if you ran a lot of trials of an experiment and recorded all the answers.
- Variance measures how spread out all the values are in a data set. A large variance means all the values are very different.
- The standard deviation is the square root of the variance, so it also measures how spread out the data is.

Interpreting Data
QUICK QUESTIONS (PAGE 15)

- $E(X) = \Sigma x \times P(X = x)$
- The probabilities are given by $\frac{frequency}{total}$, so we can add those to the table:

Number of Coffees	Frequency	Proportion
1	5	0.1
2	6	0.12
3	14	0.28
4	3	0.06
5	20	0.4
6	2	0.04

$E(X) = (1 \times 0.1) + (2 \times 0.12) + (3 \times 0.28) + (4 \times 0.06) + (5 \times 0.4) + (6 \times 0.04)$

$E(X) = 3.66$

The expected value of the number of coffees is 3.66.

- $Var(X) = E(X^2) - [E(X)]^2$

  $E(X^2) = 1^2 \times 0.1 + 2^2 \times 0.12 + 3^2 \times 0.28 + 4^2 \times 0.06 + 5^2 \times 0.4 + 6^2 \times 0.04$

  $E(X^2) = 15.5$

  $Var(X) = 15.5 - 3.66^2$

  $Var(X) = 2.1044$

  The variance of the number of coffees is 2.1044.

- $\sigma = (Var(X))^2$

  $\sigma = \sqrt{2.1044}$

  $\sigma = 1.451$ (3 d.p.)

  The standard deviation of the number of coffees is 1.451.

- Table:

Outcome	Red	White	Black
Probability	$\frac{4}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Profit	\$1 (4-3=1)	-\$3 (they lose the \$3 they paid)	\$2 (5-3 = 2)

To get the expected profit we multiply each profit by its probability and then add those together:

$$(1 \times \frac{4}{8}) + (-3 \times \frac{3}{8}) + (2 \times \frac{1}{8}) = -0.375$$

Since this number is negative, it represents a loss rather than a profit, so the player's expected loss is \$0.375.

- $Var(X) = E(X^2) - [E(X)]^2$
  $E(X^2) = 1^2 \times \frac{4}{8} + (-3)^2 \times \frac{3}{8} + 2^2 \times \frac{1}{8} = 4.375$

  (Remember to square -3 in brackets so that it comes out positive!)

  $Var(X) = 4.375 - 0.375^2$
  $Var(X) = 4.234$ (3 d.p.)
  The variance of the player's profit or loss is 4.234.

- $E[4X - Y] = 4 \times E[X] + (-1) \times E[Y]$, using the rule:

$$E[aX + bY] = aE[X] + bE[Y]$$

  Here a = 4 and b = -1
  $E[4X - Y] = 4 \times E[X] - E[Y]$
  $E[4X - Y] = 4 \times 6.7 - 5.2$
  $E[4X - Y] = 21.6$
  The expected value is 21.6.

- $Var(4X - Y) = 4^2 \times Var(X) + (-1)^2 \times Var(Y)$, using the rule:

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$$

  $Var(4X - Y) = 4^2 \times 2 + (-1)^2 \times 3.5$

Var(4X - Y) = 35.5

The variance is 35.5.

- $\sigma = (\text{Var}(4X - Y))^2$

  $\sigma = \sqrt{35.5}$

  σ = 5.96 (2 d.p).

  The standard deviation is 5.96.

The Distributions

Throwback to the Normal Distribution

STOP AND CHECK (PAGE 19)

- We standardise so that we can use the same formula to describe any normal distribution, even if they have really different means and standard deviations.
- The standard normal curve has a mean of 0 and a standard deviation of 1.
- The four features of the normal distribution are:
  - It measures a continuous variable with no upper or lower limits on what value it can take.
  - It has a mean which is also the most common value (or peak).
  - It has a bell-curve, symmetrical shape.
  - Values far from the mean are very unlikely.

The Binomial Distribution

STOP AND CHECK (PAGE 21)

- To use the Binomial distribution, we need to have:
  - A discrete variable.
  - Two outcomes for each trial: success or failure.
  - A fixed number of trials.
  - Independent trials (i.e., they don't affect each other).
  - A constant probability of success.

The Poisson Distribution

STOP AND CHECK (PAGE 22)

- To use the Poisson distribution, you need these things to be true:
  - Events happen randomly.
  - Events happen independently (don't affect each other).
  - Events can't happen at the same time or in the exact same place
  - Events happen at a constant rate, that is, the smaller the interval, the less events are likely to occur.

The Uniform Distribution

STOP AND CHECK (PAGE 24)

- You can tell a question is a uniform distribution question when it has:
  - A continuous variable.
  - Maximum and minimum values for the variable.
  - All values are equally likely.
- The formula for the uniform distribution is the area of a rectangle formula, base × height. To find the height, we use:

$$\text{Height} = \frac{1}{maximum - minimum}$$

The Triangular Distribution

STOP AND CHECK (PAGE 27)

- To use the triangular distribution, we need:
  - A continuous variable.
  - Not all values are equally likely (unlike the uniform distribution).
  - A most likely value.
  - Maximum and minimum values for the variable.
- The mode is the most likely value to occur, and it is where the peak of the triangle will be.

The Distributions

QUICK QUESTIONS (PAGE 27)

- This is a normal distribution because we have continuous data with no upper or lower bounds. We also know this because we've been given a mean and standard deviation. The parameters for this distribution are:
  - $\mu = 320$
  - $\sigma = 30$

- This is a uniform distribution because we have continuous data with upper and lower limits but we *don't* have a most likely value. Its parameters are:
  - Minimum = 8
  - Maximum = 20

- This is a triangular distribution because this time we have continuous data with upper and lower limits *and* we have a most likely value. The parameters are:
  - Minimum = 2
  - Maximum = 18
  - Mode = 6

- This is a binomial distribution because we have discrete data, with a set number of trials (staff members who are given cakes), independent trials (random decision), two outcomes (over 50 or under 50), and a fixed probability of success (being over 50). The parameters for this distribution are:
  - $p = 10$
  - $p = 0.45$ or 45%

- This is a Poisson distribution because we have discrete data, with a set rate of receiving texts, and we can *assume* the texts are random and independent. The parameter is:
  - $\lambda = 4$ (per hour)

Solving Distribution Problems

Solving the Normal Distribution

STOP AND CHECK (PAGE 31)

- To find the probability that x is less than a value, go into Ncd in your calculator, and input:
  - Lower as -999999999
  - Upper as the value you're looking at.
  - The mean (μ) and standard deviation (σ) you've been given.
- If you need to find two separate probabilities, do exactly what you did above, but set the upper as +99999999 when you're finding the 'x is more than a value' part. Then just add the two answers together!

Continuity Corrections

STOP AND CHECK (PAGE 32)

- You use a continuity correction when the data has been rounded.

Binomial by Calculator

STOP AND CHECK (PAGE 34)

- Bpd is what you use when you're trying to find the probability that x **equals** a value, and Bcd is what you use when you're finding the probability that x is **less than** or **equal to** a value.

Poisson by Calculator

STOP AND CHECK (PAGE 35)

- Exactly like in the Binomial, Ppd is what you use when you're trying to find the probability that x **equals** a value, and Pcd is what you use when you're finding the probability that x is **less than** or **equal to** a value.

Solving Distribution Problems

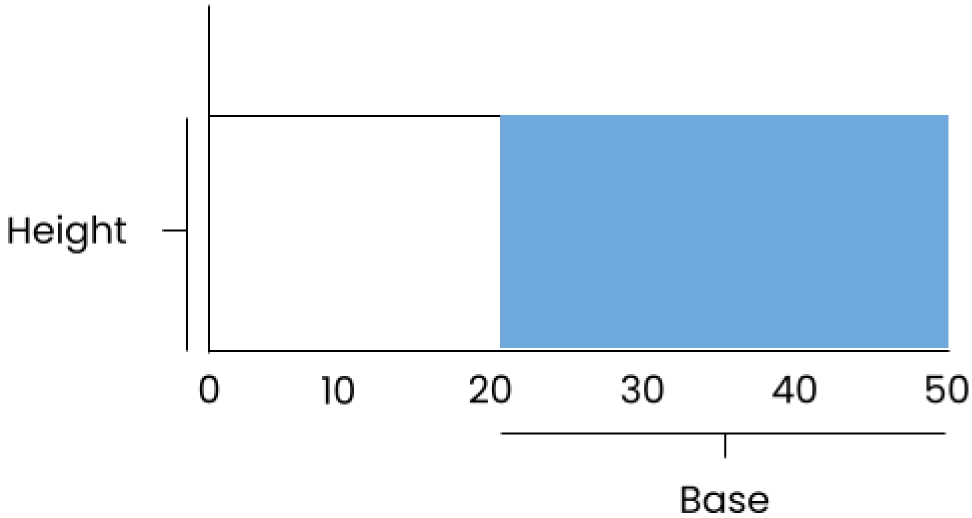
QUICK QUESTIONS (PAGE 40)

- This is a Poisson distribution (as it is discrete data without trials and with a 'rate' style parameter), where $\lambda = 2$.
  We want to find $P(X = 0)$, so we go to Ppd in the calculator and set $x = 0$ and $\mu = 2$.
  The answer comes out as $P(X = 0) = 0.1353$, so there is only a 0.1353 chance that a page will not have to be reprinted.

- This is a Binomial distribution (since it's discrete data with a set number of trials, 3 jobs). We want to find $P(X \geq 2)$, so we need to find $1 - P(X \leq 2)$, since the calculator can only do the calculation this way around.
  In Bcd, we input:
  - x = 2
  - Numtrial = 3
  - p = 0.05

  $P(X \leq 2) = 0.999875$
  $1 - P(X \leq 2) = 0.000125$
  So the probability that Bobby has to write a report is 0.0001 (to 4d.p.).

- This is a uniform distribution because it is continuous data with maximum and minimum values. The shaded area represents the probability we are looking for:



P = base x height (area of a rectangle formula)

Height = $\frac{1}{maximum - minimum}$

Height = $\frac{1}{50-0}$

Height = 0.02

P(X > 20) = (50 - 20) x 0.02

P(X > 20)= 0.6

The probability that he will be at least 20cm off is 0.6.

- This is a normal distribution because it is continuous data with a mean and no upper or lower limits.

  We need to first find P(X > 1500), so we go to Ncd and set:
    - Lower = 1500
    - Upper = 9999999
    - μ = 1000
    - σ = 200

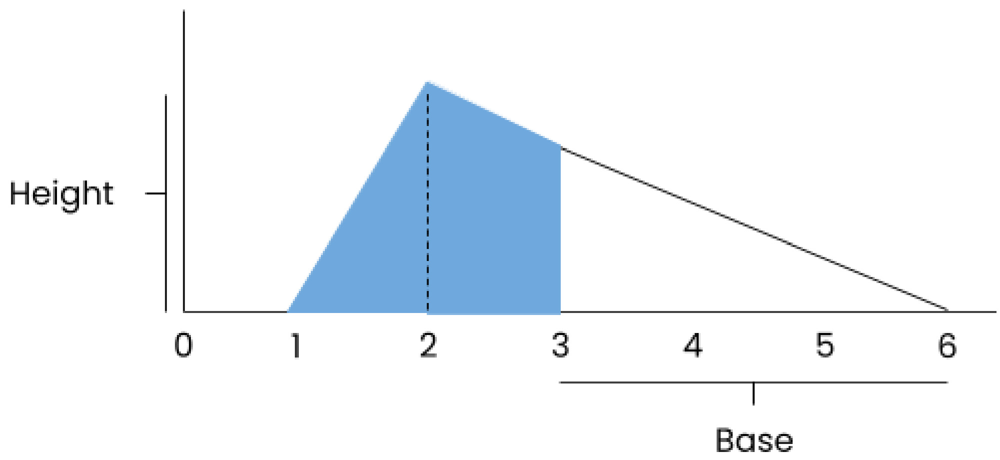
  We get P(X > 1500) =0.0062

  To find how many out of 20 booths would fall, we multiply this probability by 20:
    - 0.0062 x 20 = 0.124, which we'll round to 0 booths.

  So out of 20 booths, we wouldn't really expect any to fall.

- This is a triangular distribution because it uses continuous data with an upper limit, lower limit and most likely value.

The area we're looking for to find the probability loading the webpage will take less than 3 seconds is:



To find the area, we first need to find the area of the unshaded triangle. Our x value is between the mode and the maximum, so we use the formula:

$$\text{Height} = \frac{2(b-x)}{(b-a)(b-c)}$$

Our values are:

- a = 1
- b = 6
- c = 2
- x = 3

Height = $\frac{2(6-3)}{(6-1)(6-2)}$

Height = 0.3

Then, we find the probability by the area of a triangle formula:

$$p = \frac{1}{2} \times \text{base} \times \text{height}$$

p = $\frac{1}{2}$ x (6-3) x 0.3

p = 0.45

Finally, we take this away from 1 to find the shaded area:

1- 0.45 = 0.55

The probability the webpage loads in less than three seconds is 0.55.

Inverse Distributions

Inverse Normal

STOP AND CHECK (PAGE 43)

- To solve these questions, you need to:
  - Go to InvN on your calculator.
  - Put in the correct tail side and the area (which is your probability).
  - Set μ = 0 and σ =1, for the standard normal distribution
  - Solve to find Xinv, which will be your Z value
  - Substitute all the values you know, including the Z value, into this equation:

$$Z = \frac{x-\mu}{\sigma}$$

  Rearrange to find the missing value!
- You'll know a question uses the inverse normal because you'll be given a probability, which corresponds to an area on the graph!

Inverse Binomial

STOP AND CHECK (PAGE 45)

- To find P(X = 0), just calculate 1 – P(X ≥ 1)

Inverse Poisson

STOP AND CHECK (PAGE 47)

- You'll know a question uses the inverse Poisson if you're told a **probability** in the question and need to find the **parameter**.

Inverse Distributions

QUICK QUESTIONS (PAGE 47)

- We can tell that this is an inverse Poisson question because:
  - It uses discrete data.
  - It involves a time period and random events during that time period.
  - We're told a probability and need to find the parameter.

  Remember that our golden rule to solving these is to use the probability that x = 0, which we need to find by 1-P(X ≥ 1)

  P(X = 0) = 1 – 0.92

  P(X = 0) = 0.08

  Next, we'll substitute that into the formula:

  $P(X=x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$

  $0.08 = \frac{e^{-\lambda}\lambda^{0}}{0!}$

  This collapses down to:

  $0.08 = e^{-\lambda}$

  Logging both sides:

  ln0.08 = -λ

  -2.526 = -λ

  λ = 2.526

  But remember this is for **three** months, so the last thing we need to do is divide it by 3 to find the mean for *one* month.

  $\lambda = \frac{2.526}{3}$

  λ = 0.8419 seals caught per month on average.

- This is also an inverse Poisson question, so we need to follow the same steps as for the last question: first, we find P(X = 0) by subtracting P(X ≥ 1) from 1.

  P(X = 0) = 1−0.27
  P(X = 0) =0.73

  Using the formula, we get:

  $P(X=x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$
  $0.72 = \frac{e^{-\lambda}\lambda^{0}}{0!}$
  $0.72 = e^{-\lambda}$
  ln0.72 = -λ
  -0.3285 = -λ
  λ= 0.3285

  Finally, to find the mean for 50 square metres, we multiply this mean by 50:

  λ= 0.3285 x 50
  λ=16.425

  The mean number of chocolate coins in an area of $50m^2$ is 16.425.

