Statistics:
Distributions & Hypothesis Testing

Material Covered

Binomial Distribution

Normal Distribution

Hypothesis Testing

Binomial Distribution

$$P(X = x) = C_x^n p^x (1 - p)^{n-x}$$

Specification Points - AQA

	Content
N1	Understand and use simple, discrete probability distributions (calculation of mean and variance of discrete random variables is excluded), including the binomial distribution, as a model; calculate probabilities using the binomial distribution.

Specification Points – OCR A

2.04 Statistical Distributions			
2.04a	Discrete probability distributions	a) Understand and be able to use simple, finite, discrete probability distributions, defined in the form of a table or a formula such as: $P(X=x)=0.05x(x+1)$ for $x=1,2,3$. [*Calculation of mean and variance of discrete random variables is excluded.*]	
2.04b 2.04d		b) Understand and be able to use the binomial distribution as a model.	d) Know and be able to use the formulae $\mu = np$ and $\sigma^2 = npq$ when choosing a particular normal model to use as an approximation to a binomial model.
2.04c		c) Be able to calculate probabilities using the binomial distribution, using appropriate calculator functions. *Includes understanding and being able to use the formula* $P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}$ *and the notation* $X \sim B(n,p)$. *Learners should understand the conditions for a random variable to have a binomial distribution, be able to identify which of the modelling conditions (assumptions) is/are relevant to a given scenario and be able to explain them in context. They should understand the distinction between conditions and assumptions.*	

Specification Points – OCR MEI

Situations leading to a binomial distribution	MR1	Recognise situations which give rise to a binomial distribution.		
	R2	Be able to identify the probability of success, p, for the binomial distribution.	The binomial distribution as a model for observed data.	$B(n, p)$, $q = 1 - p$ ~ means 'has the distribution'.
Calculations relating to binomial distribution	R3	Be able to calculate probabilities using the binomial distribution.	Including use of calculator functions.	
Mean and expected frequencies for binomial distribution	R4	Understand and use mean $= np$.		
	R5	Be able to calculate expected frequencies associated with the binomial distribution.		
Discrete probability distributions	R6	Be able to use probability functions, given algebraically or in tables. Know the term discrete random variable.	Restricted to simple finite distributions.	X for the random variable. x or r for a value of the random variable.
	R7	Be able to calculate the numerical probabilities for a simple distribution. Understand the term discrete uniform distribution.	Restricted to simple finite distributions.	$P(X = x)$ $P(X \leq x)$

Specification Points - Edexcel

4.1	**Understand and use simple, discrete probability distributions (calculation of mean and variance of discrete random variables is excluded), including the binomial distribution, as a model; calculate probabilities using the binomial distribution.**	**Students will be expected to use distributions to model a real-world situation and to comment critically on the appropriateness.** **Students should know and be able to identify the discrete uniform distribution.** **The notation $X \sim B(n, p)$ may be used.** **Use of a calculator to find individual or cumulative binomial probabilities.**

Binomial Probability

- A **random variable** X can **take** any **one** of a **range** of **particular values** x in a **sample space**.

- A **probability distribution function models** the **probability** of all **outcomes** in the **sample space**.

The **binomial distribution** can be used to **model** the **probabilities** of **discrete random variables**.

Binomial Probability

The **binomial distribution** can be used to **estimate** the **probability** of the **number** of **successful outcomes** from **multiple trials**.

- 2 possible **outcomes** for each **trial** (**success** or **failure**).
- **Fixed number** of **trials** (n).
- **Trials** are **independent**.
- **Trials** are **identical** (p is **constant**).

Binomial Probability

The **binomial formula** is used to **determine** the **probability** P that the **discrete random variable** X **equals** a given **number of successes** x.

$$P(X = x) = C_x^n p^x (1 - p)^{n-x}$$

$$C_x^n = \frac{n!}{x!\,(n - x)!}$$

Why **4** and **not 5**?

Exemplar Exam Question

Pull out **key information** from **context**

1) Greg and Tim are playing Yahtzee. The game lasts for 13 turns. On each turn Greg and Tim both throw five 6-sided dice, and try to match numbers they roll.

(i) On his first turn, Greg manages to get a Yahtzee (all five dice have the same value) in a single throw. Explain why the probability of this happening is $\left(\frac{1}{6}\right)^4$ **[2 Marks]**

(ii) Over the course of the game, Tim manages to get a Yahtzee in a single throw 3 times. Calculate the probability of this happening. Your answer should be given to 3 significant figures. **[2 Marks]**

Remember to **round**

4 marks over **2 parts**

Exemplar Exam Question Answer

(i) Interpret probability.

Number on each die is independent

Number on first die can take any value

Next 4 must then match this value

Probability of each die doing so is $\frac{1}{6}$

So probability of all 4 doing so is $\frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \left(\frac{1}{6}\right)^4$

Alternatively $6 \times \left(\frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6}\right) = \left(\frac{1}{6}\right)^4$

[2 Marks]

Exemplar Exam Question Answer

(ii) Determine which distribution to use.

Results of die rolls are discrete outcomes.

Check that conditions for binomial distribution apply:

- 2 possible outcomes – Yahtzee or not a Yahtzee.
- Fixed number of trials – Set number of dice.
- Trials are independent – One roll does not affect another.
- Trials are identical – Probability of a Yahtzee is always the same.

Exemplar Exam Question Answer

Determine variables for binomial distribution.

Need to determine probability throwing three Yahtzees in a single throw.

There are 13 turns and so 13 chances to for success.

This means number of trials $n = 13$.

Chance of throwing a Yahtzee was calculated in previous question.

This means probability of success $p = \frac{1}{1296}$

Looking for exactly 3 Yahtzees thrown

This means number of successes $x = 3$

[1 Mark]

Exemplar Exam Question Answer

$n = 13$

$x = 3$

$p = \frac{1}{1296}$

Calculate probability

Probability that discrete random variable $X = x$ is

$$P(X = x) = C_x^n p^x (1 - p)^{n-x}, \qquad C_x^n = \frac{n!}{x!(n-x)!}$$

Plug in values

$$P(X = 3) = C_3^{13} \left(\frac{1}{1296}\right)^3 \left(1 - \left(\frac{1}{1296}\right)\right)^{13-3}, \qquad C_3^{13} = \frac{13!}{3!(13-3)!}$$

$$= 286 \times 4.95 \times 10^{-10} \times 0.992$$

$$= 1.30 \times 10^{-7}$$

[1 Mark]

Cumulative Binomial Probabilities

A **cumulative probability density function models** the **probability** that the **random variable** X takes a **value** which is **less than or equal** to x.

- **Calculating** the **sum** of each **individual probability** can be quite **time consuming** in an **exam**.
- We can look up **cumulative probability values** in a **statistical table** or use the **binomial cumulative probability function** on a **calculator**.

Pull **relevant information** from **question**

Exemplar Exam Question

Question on **calculating probabilities**

1) A carnival game involves drawing a token from a box. There are 20 tokens in the box total, with 1 gold, 2 silver and 4 bronze tokens which win a respective prize.
Jess plays the game 8 times to try and win a prize. The token she picks is replaced after each game. Calculate the probability that she wins

Constant probabilities means apply **binomial distribution**

(i) At least 2 gold prizes **[2 Marks]**

(ii) Between 3 and 6 (inclusive) prizes that aren't gold. **[2 Marks]**

Read each part **carefully** and consider which **calculation** to do.

4 marks over **2 parts**.

Exemplar Exam Question Answer

(i) Determine variables for binomial distribution.

Need to determine probability of more than 2 of the 8 tokens being gold

8 tokens are being taken out of the box.

This means number of trials $n = 8$.

20 possible tokens to choose, only 1 is gold.

This means probability of success $p = \frac{1}{20} = 0.05$

Need at least 2 tokens to be gold

This means number of successes is in range $X \geq 2$

[1 Mark]

Exemplar Exam Question Answer

$n = 8$
$X \geq 2$
$p = 0.05$

Calculate correct value

Use calculator to find $\mathrm{P}(X \geq 2)$ for given n and p:

$$\Rightarrow \mathrm{P}(X \geq 2) = 0.0572$$

[1 Mark]

Exemplar Exam Question Answer

(ii) Determine variables for binomial distribution.

Need to determine probability of between 3 and 6 tokens being bronze or silver.

8 tokens are being taken out of the box.

This means number of trials $n = 8$.

20 possible tokens to choose, 2 silver and 4 bronze.

This means probability of success $p = \frac{2+4}{20} = 0.30$

Need between 3 and 6 inclusive tokens to be silver or bronze

This means number of successes is in range $3 \leq X \leq 6$

[1 Mark]

Exemplar Exam Question Answer

$n = 8$
$3 \leq X \leq 6$
$p = 0.3$

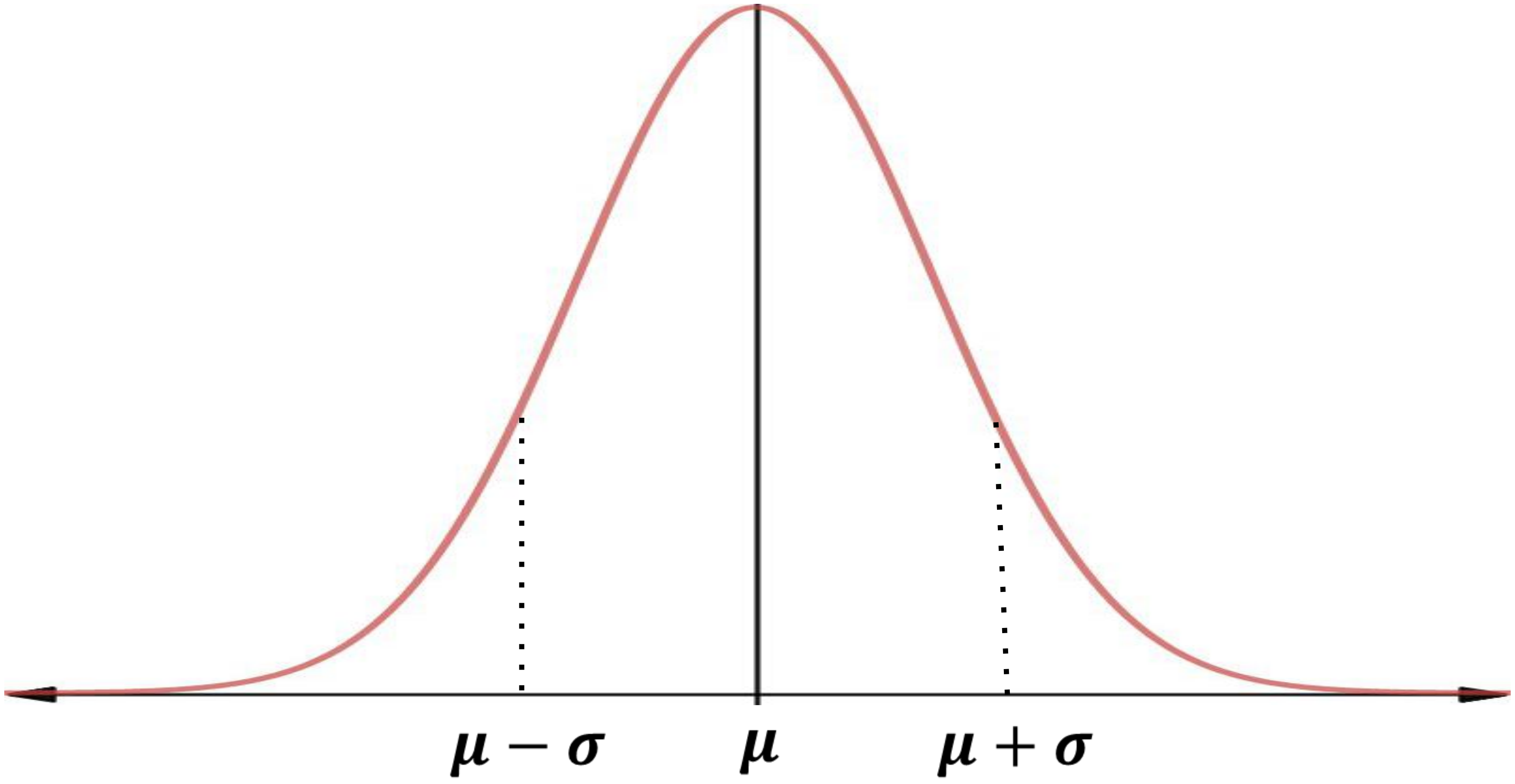
Calculate correct value

Use calculator to find $\mathrm{P}(3 \leq X \leq 6)$ for given n and p:

$$\Rightarrow \mathrm{P}(3 \leq X \leq 6) = 0.4469$$

[1 Mark]

Normal Distribution



Specification Points - AQA

	Content
N2	Understand and use the Normal distribution as a model; find probabilities using the Normal distribution. Link to histograms, mean, standard deviation, points of inflection and the binomial distribution.

	Content
N3	Select an appropriate probability distribution for a context, with appropriate reasoning, including recognising when the binomial or Normal model may not be appropriate.

Specification Points – OCR A

e) Understand and be able to use the normal distribution as a model.

Includes understanding and being able to use the notation $X \sim \mathrm{N}(\mu, \sigma^2)$.

f) Be able to find probabilities using the normal distribution, using appropriate calculator functions.

This includes finding x*, for a given normal variable, when* $\mathrm{P}(X < x)$ *is known.*

Learners should understand the standard normal distribution, Z*, and the transformation* $Z = \dfrac{X - \mu}{\sigma}$.

g) Understand links to histograms, mean and standard deviation.

Learners should know and be able to use the facts that in a normal distribution,

1. about two-thirds of values lie in the range $\mu \pm \sigma$,
2. about 95% of values lie in the range $\mu \pm 2\sigma$,
3. almost all values lie in the range $\mu \pm 3\sigma$ *and*
4. the points of inflection in a normal curve occur at $x = \mu \pm \sigma$.

[*The equation of the normal curve is excluded.*]

h) Be able to select an appropriate probability distribution for a context, with appropriate reasoning, including recognising when the binomial or normal model may not be appropriate.

Includes understanding that a given binomial distribution with large n can be approximated by a normal distribution.

[*Questions explicitly requiring calculations using the normal approximation to the binomial distribution are excluded.*]

Specification Points – OCR MEI

MR8	Be able to use the Normal distribution as a model.	Includes recognising when a Normal distribution may not be appropriate. Understand how and why a continuity correction is used when using a Normal distribution as a model for a distribution of discrete data. Recognise from the shape of the distribution when a binomial distribution can be approximated by a Normal distribution.	$X \sim \mathrm{N}(\mu, \sigma^2)$
R9	Know the shape of the Normal curve and understand that histograms from increasingly large samples from a Normal distribution tend to the Normal curve.	Includes understanding that the area under the Normal curve represents probability.	
R10	Know that linear transformation of a Normal variable gives another Normal variable and know how the mean and standard deviation are affected. Be able to standardise a Normal variable.	$y_i = a + bx_i \Rightarrow \overline{y} = a + b\overline{x},\ s_y^2 = b^2 s_x^2$	Standard Normal $Z \sim \mathrm{N}(0, 1)$ $Z = \frac{X - \mu}{\sigma}$
R11	Know that the line of symmetry of the Normal curve is located at the mean and the points of inflection are located one standard deviation away from the mean.		
R12	Be able to calculate and use probabilities from a Normal distribution.	Including use of calculator functions.	

Specification Points - Edexcel

4.2	Understand and use the Normal distribution as a model; find probabilities using the Normal distribution	The notation $X \sim N(\mu, \sigma^2)$ may be used. Knowledge of the shape and the symmetry of the distribution is required. Knowledge of the probability density function is not required. Derivation of the mean, variance and cumulative distribution function is not required. Questions may involve the solution of simultaneous equations. Students will be expected to use their calculator to find probabilities connected with the normal distribution.
	Link to histograms, mean, standard deviation, points of inflection	Students should know that the points of inflection on the normal curve are at $x = \mu \pm \sigma$. The derivation of this result is not expected.
	and the binomial distribution.	Students should know that when n is large and p is close to 0.5 the distribution $B(n, p)$ can be approximated by $N(np, np[1 - p])$ The application of a continuity correction is expected.

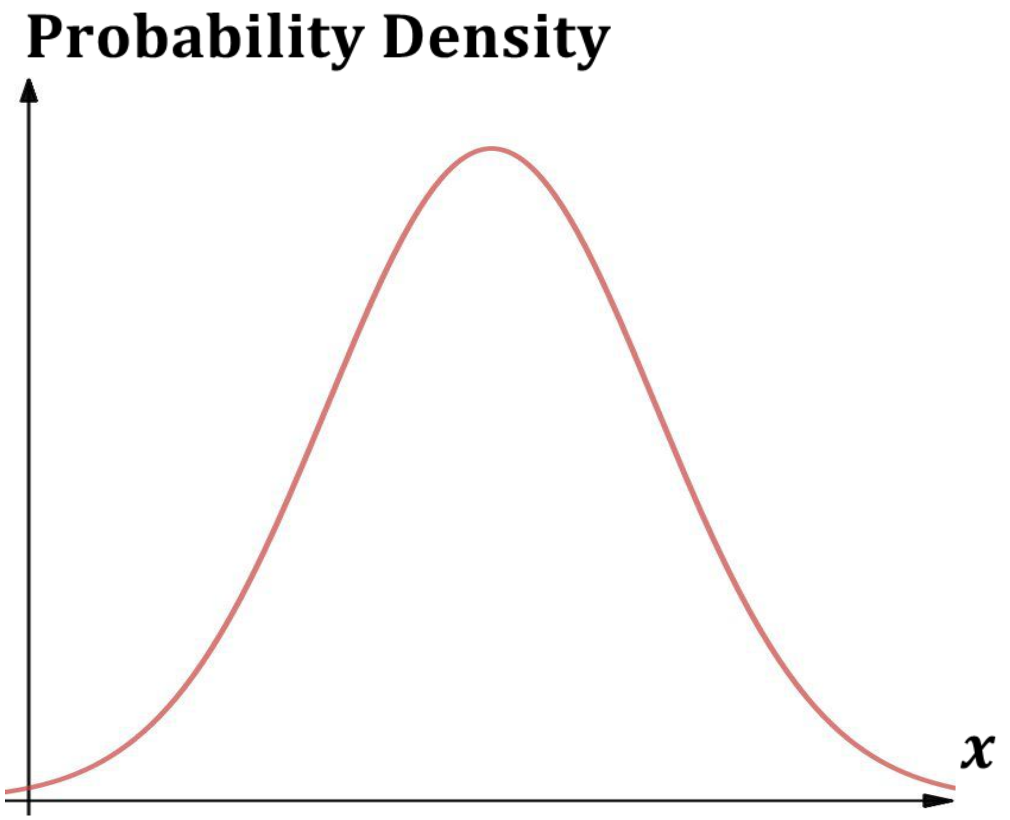
4.3	Select an appropriate probability distribution for a context, with appropriate reasoning, including recognising when the binomial or Normal model may not be appropriate.	Students should know under what conditions a binomial distribution or a Normal distribution might be a suitable model.

Properties of the Normal Distribution

The **normal distribution** is a **probability distribution** used to **model continuous random variables**.

- **Continuous random variables** can take **one** of an **infinite number** of **values**.
- **Continuous random variables** have a **continuous probability distribution**.

The **normal distribution** is a **probability density** function used to **model** many **naturally occurring distributions**.



Properties of the Normal Distribution

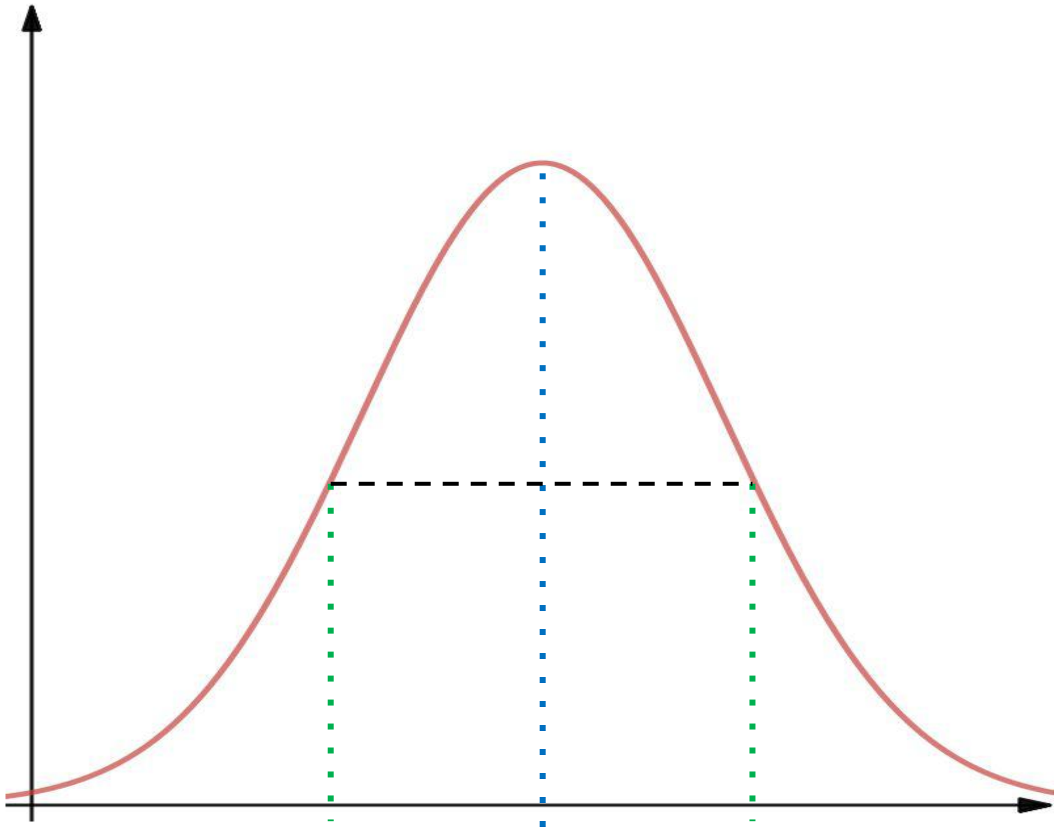
The **normal distribution** is **defined** by the **parameters**:

$\boldsymbol{\mu}$

$\boldsymbol{\sigma}$

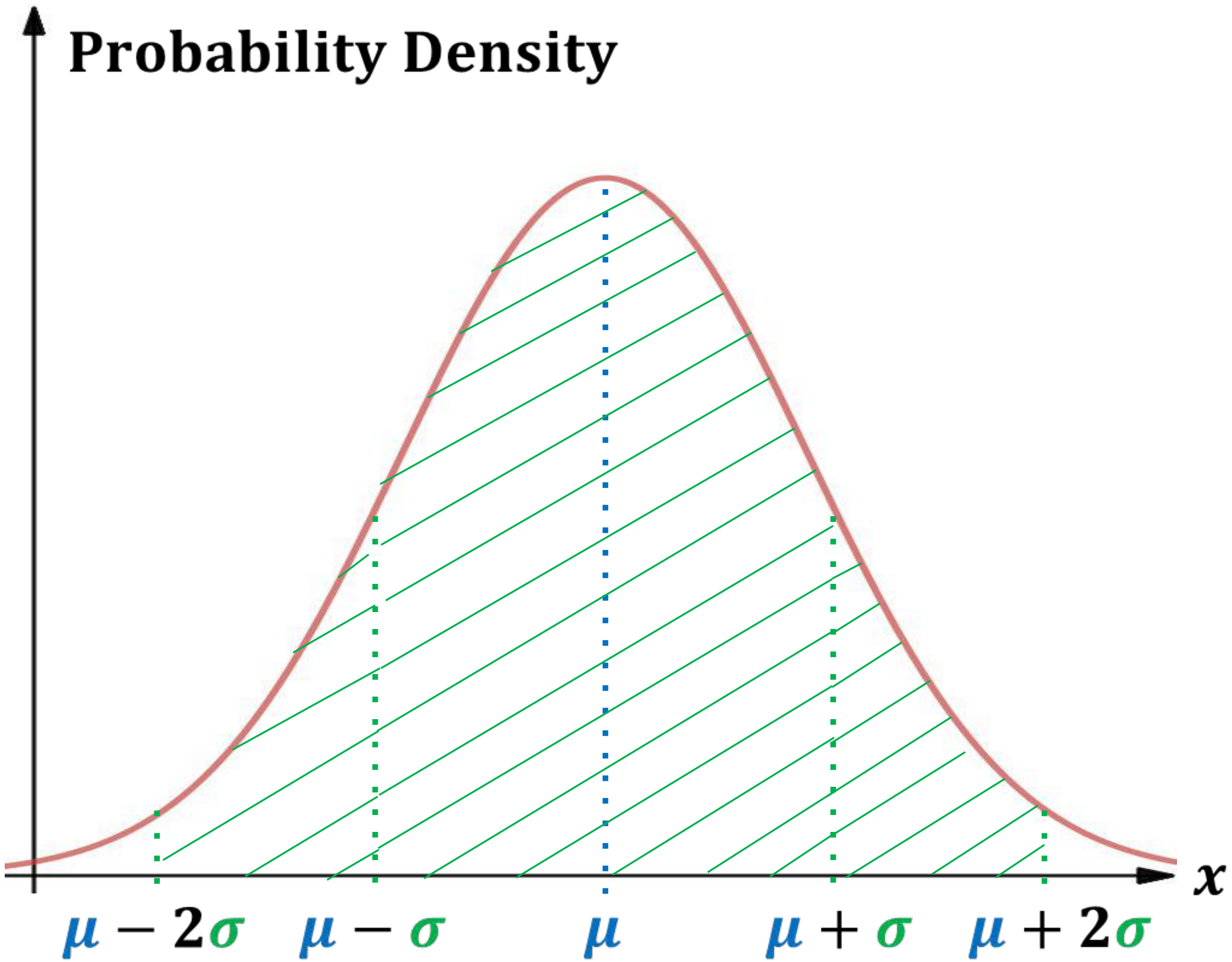
The **normal distribution**:

- Is **symmetric**.
- Has **points of inflection** at $\mu \pm \sigma$.
- Has **total area** under the **curve** $= 1$.



Properties of the Normal Distribution

- A **random value** has a **68%** chance of being within **1**σ of the **mean**.
- A **random value** has a **95%** chance of being within **2**σ of the **mean**.
- A **distribution** with a **smaller** σ is **narrower** with a **sharper peak**.
- A **distribution** with a **higher** σ is **flatter** with a **shallower peak**.



Remember to **justify answers** for **full marks**

Exemplar Exam Question

Use all **information** given in **graphs**

1) On the next page are $\mathbf{4}$ graphs, with the x values of their turning points and points of inflection given and the y axis showing a probability density. Clearly explaining your reasoning, determine which one of the graphs

(i) Does not describe a normal distribution. **[1 Mark]**

(ii) Is the normal distribution with the lowest variance. **[1 Mark]**

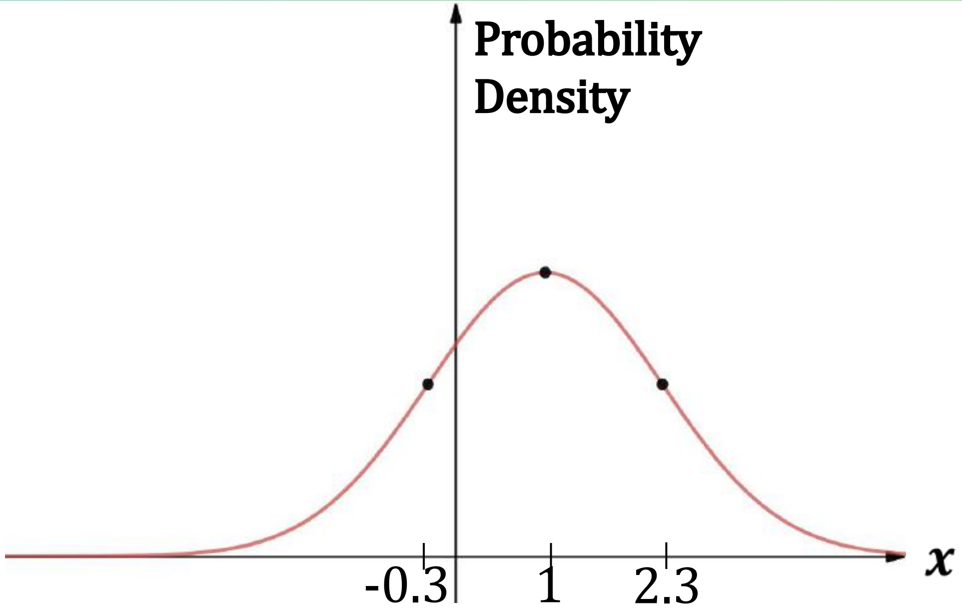
(iii) Has an area under its curve for $-1.6 \leq x \leq 3.6$ equal to 0.95. **[2 Marks]**

You may use each graph as an answer more than once.

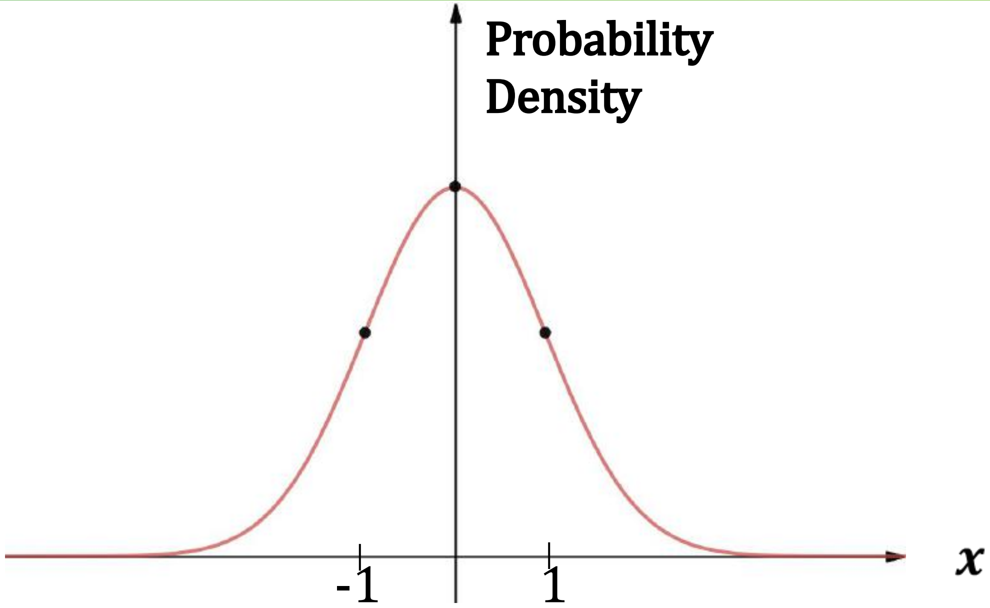
Remember **key features** of **normal distribution curves**.

2 quick parts, but last needs an **extra step**

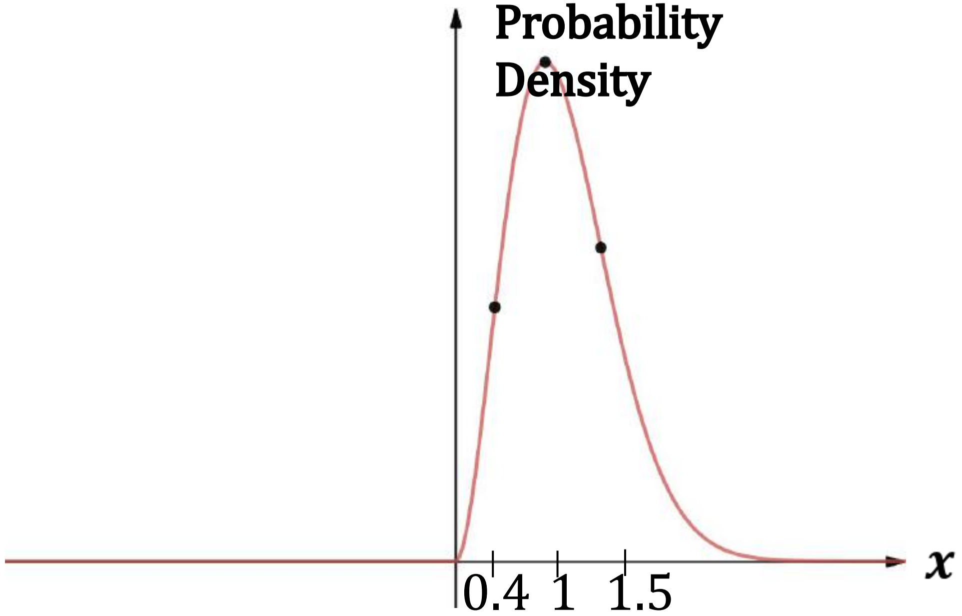
a)


Probability
Density
x
-0.3
1
2.3

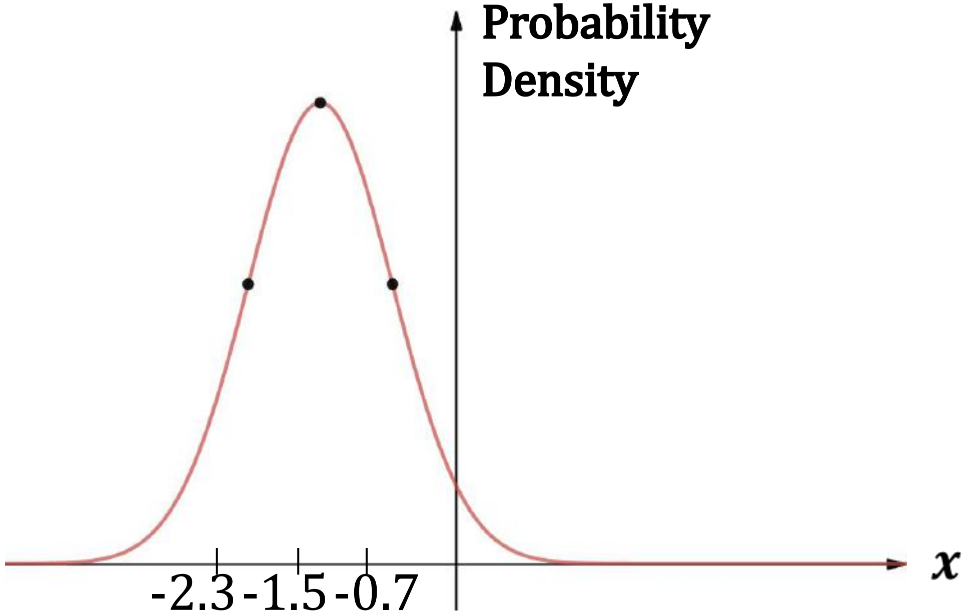
b)


Probability
Density
x
-1
1

c)


Probability
Density
x
0.4
1
1.5

d)


Probability
Density
x
-2.3
-1.5
-0.7

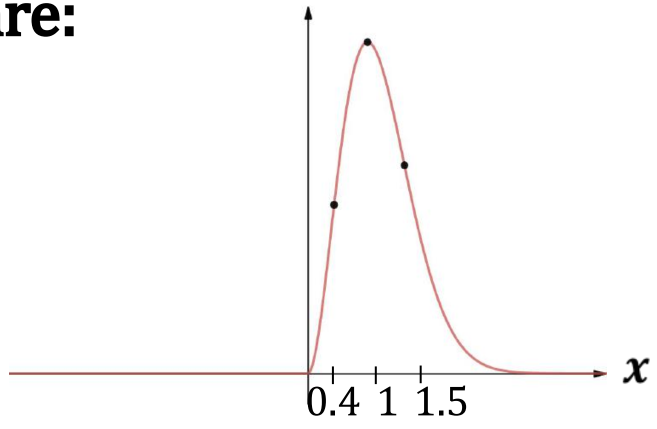
Exemplar Exam Question Answer

(i) Check shape of graphs

Two key properties of normal distribution graphs are:

- symmetry in x-axis
- Points of inflection at at $\mu \pm \sigma$

Looking at graph c) we notice that:



- there is no symmetry in the x-axis
- points of inflection are unevenly spaced from μ, so does not follow formula

So c) is not a normal distribution graph

[1 Mark]

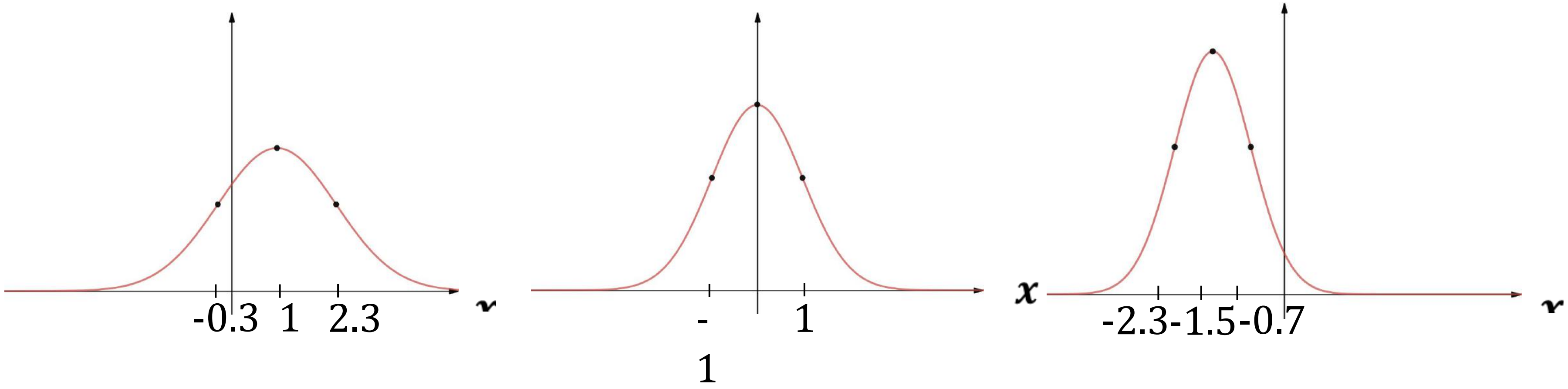
Exemplar Exam Question Answer

(ii) Check shape of graphs

Normal Distribution graph with lowest variance will have the sharpest peak

Inspecting three normal distribution graphs shows d) must have highest variance

[1 Mark]



Exemplar Exam Question Answer

(iii) Interpret region

$$-1.6 \leq x \leq 3.6$$

Area under normal distribution graph would be 0.95 for region:

$$\mu - 2\sigma \leq x \leq \mu + 2\sigma$$

Compare to given region to deduce

$$\mu + 2\sigma = 3.6$$

$$\mu - 2\sigma = -1.6$$

Exemplar Exam Question Answer

Solve simultaneous equation

$$\mu + 2\sigma = 3.6$$

$$\mu - 2\sigma = -1.6$$

Add equations to give

$$2\mu = 2$$

$$\mu = 1$$

Subtract equations to give

$$4\sigma = 5.2$$

$$\sigma = 1.3$$

[1 Marks]

Exemplar Exam Question Answer

Check key points of graphs

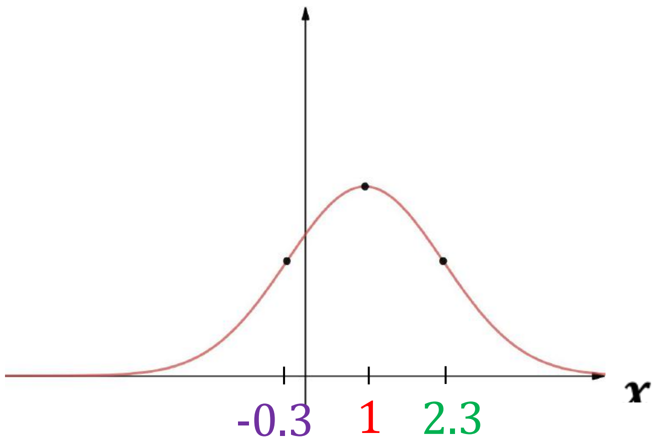
Now know that $\mu = 1$ and $\sigma = 1.3$

So graph would have peak at $x = 1$ and points of inflection at

$$x = \mu - \sigma = 1 - 1.3 = -0.3$$

$$x = \mu + \sigma = 1 + 1.3 = 2.3$$

Check graphs to see this matches with graph a)

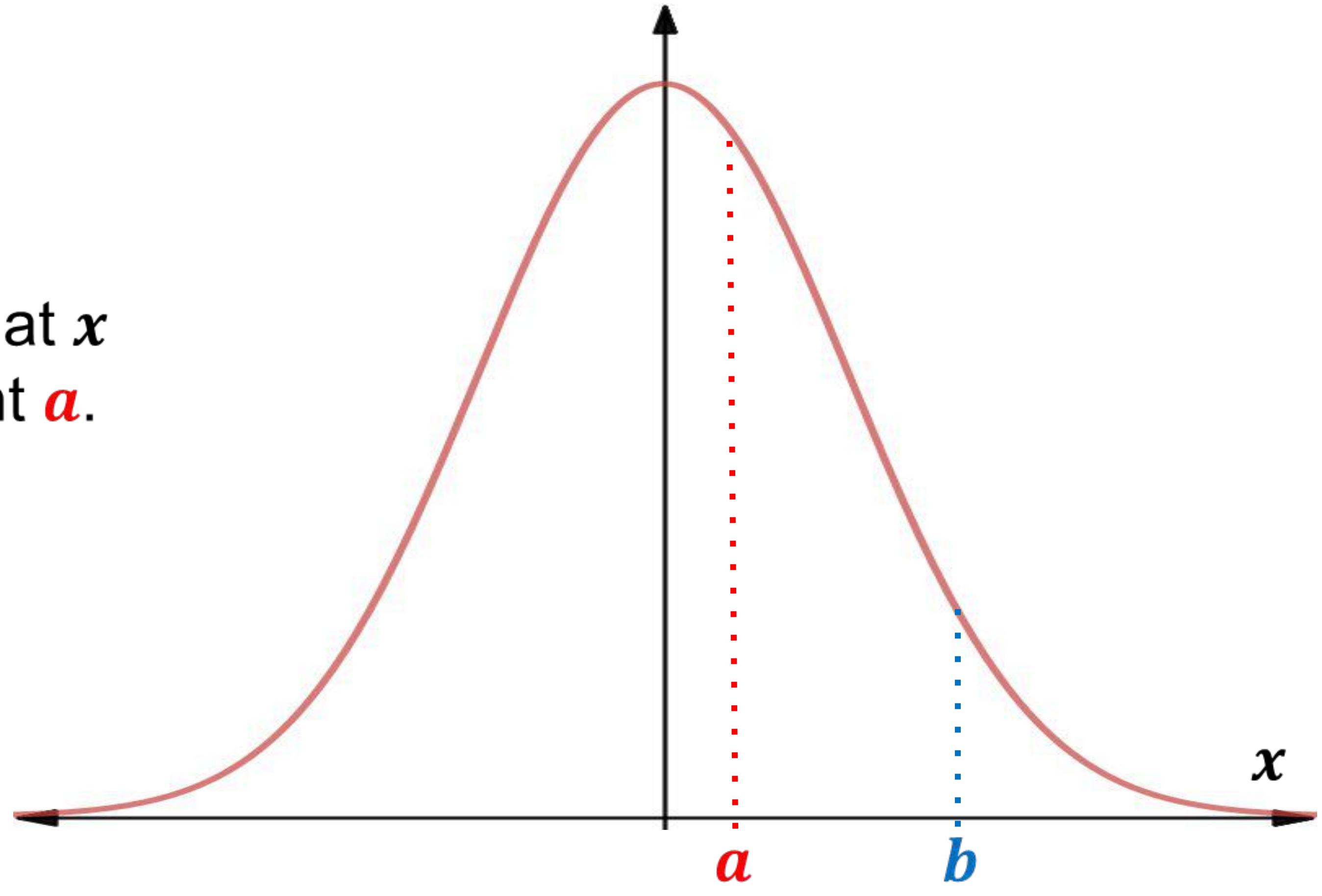


[1 Mark]

Calculating Probabilities

The **probability** that x takes a value between **2 positive constants** a and b can be **determined** using a **calculator**.

We can also calculate the **probability** that x takes a value **above** or **below** a constant a.



Standard Normal Distribution

Any **normal distribution** $X \sim N(\mu, \sigma^2)$ can be **transformed** to the **standard normal distribution** $Z \sim N(0, 1)$.

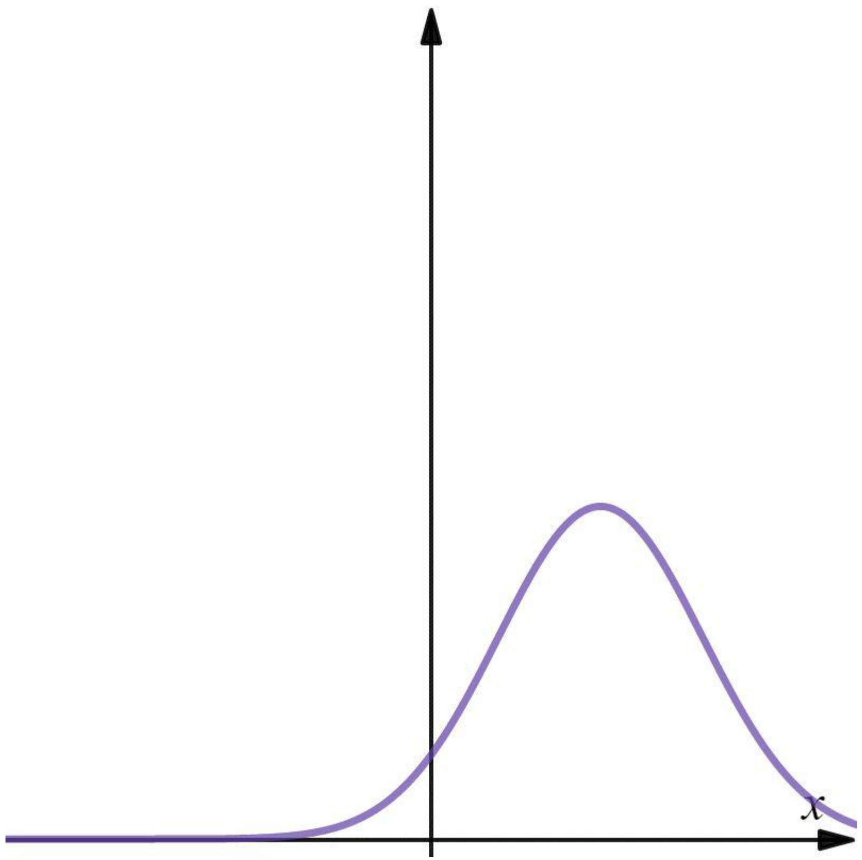
- This is **useful** as it allows us to use a **given probability** for an x**-value** to find the **equivalent** z**-value**, and calculate an unknown **mean** or **standard deviation**.

x-value in $X \sim N(\mu, \sigma^2)$ ⟶ z-value in $Z \sim N(0, 1)$

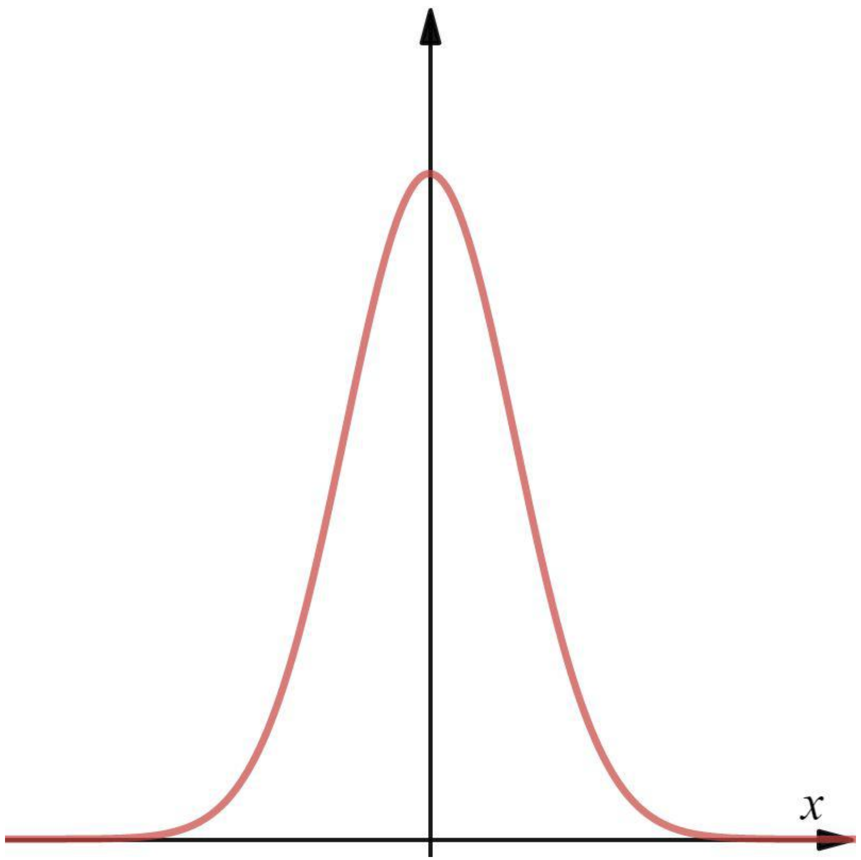
$$P(X \leq x) = P\left(Z \leq \frac{x - \mu}{\sigma}\right)$$

Standard Normal Distribution

$X \sim N(\mu, \sigma^2)$



$Z \sim N(0, 1)$



Careful! Not **standard deviation**.

Exemplar Exam Question

Pull **key information** out of **question**.

1) "*Engineering Monthly*" is planning a segment called "*World's Greatest Skyscrapers*" and shortlists 850 potential candidates for the segment. Their heights follow a normal distribution with mean $\mathbf{425\ m}$.

(i) The judges select a building at random. Given that there is a $\mathbf{70\%}$ chance of them selecting a building shorter than Taipei 101 ($\mathbf{508\ m}$), calculate the variance of the heights of the shortlisted buildings to 3 significant figures.

[3 Marks]

(ii) The judges decide to instead include any buildings shorter than $\mathbf{300\ m}$ in a separate segment. Calculate the probability of a randomly selected building from the shortlist being in this segment.

[1 Mark]

Can't fully describe distribution, need to relate to **standard distribution**.

4 marks in total, fair amount of work for **first part** but then **simple calculation** after.

Percentage Points of the Normal Distribution

p	z	p	z
0.5000	0.0000	0.0500	1.6449
0.4000	0.2533	0.0250	1.9600
0.3000	0.5244	0.0100	2.3263
0.2000	0.8416	0.0050	2.5758
0.1500	1.0364	0.0010	3.0902
0.1000	1.2816	0.0005	3.2905

$$p = P(Z > z)$$

Exemplar Exam Question Answer

(i) List key information

Have mean $\mu = 425$ and unknown standard distribution σ

Given an x-value of 508

Probability for given x-value $P(X < x) = 0.700$

Find z-value for given probability

From statistical table or calculator with $\mu = 0$ and $\sigma = 1$:

$$P(Z < z) = 0.700 \Rightarrow z = 0.5244$$

[1 Mark]

Exemplar Exam Question Answer

$z = 0.5244$ $\quad x = 508$ $\quad \mu = 425$

Calculate standard deviation

z-value is related to x-value through formula

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

$$\Rightarrow \sigma = \frac{x - \mu}{z} = \frac{508 - 425}{0.5244}$$

$$= 158.276 \ldots$$

[1 Mark]

Exemplar Exam Question Answer

Calculate variance

$$\sigma^2 = 158.276\ldots^2 = 25000 \text{ m}^2$$

[1 Mark]

Exemplar Exam Question Answer

$x = 300$ $\mu = 425$ $\sigma = 158.276\ldots$

(ii) Calculate probability

From calculator with given μ and σ

$$P(X < 300) = 0.2148$$

[1 Mark]

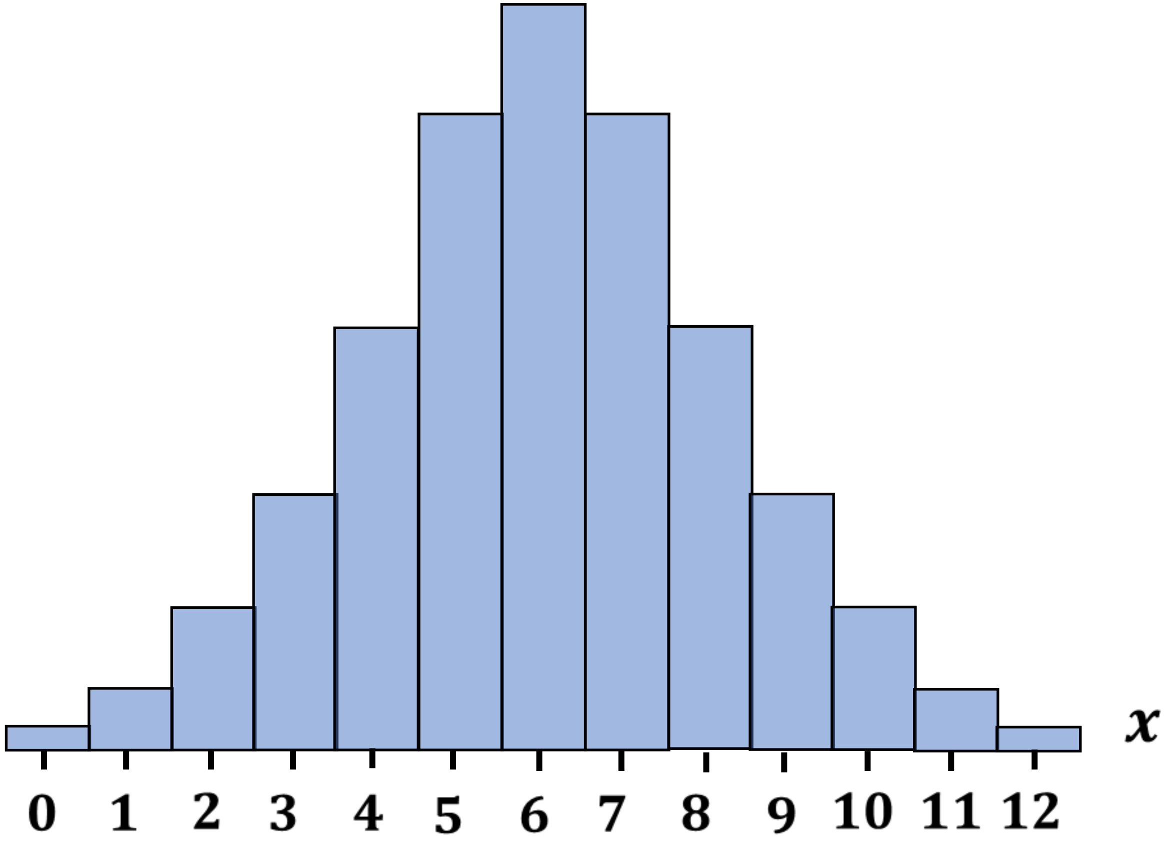
Binomial Approximation

At **large values** of n and $p \approx 0.5$, the **binomial distribution** can be **approximated** by the **normal distribution**.

- The **approximating normal distribution** has **mean** and **standard deviation**:

$$\mu = np$$

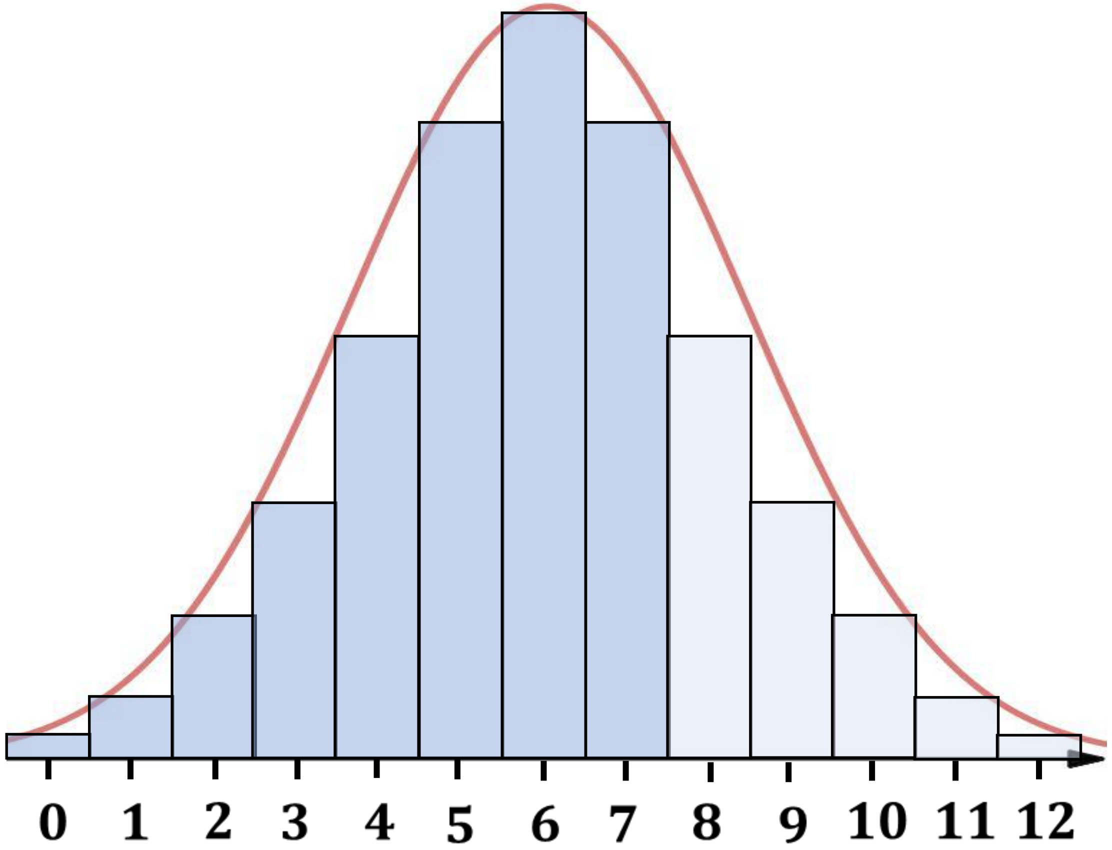
$$\sigma = \sqrt{np(1-p)}$$



Binomial Approximation

The **binomial distribution** is **discrete** while the **normal distribution** is **continuous**. This means we must **apply** a **continuity correction** when **calculating probabilities**.

$X \sim B(n, p)$	$X \sim N\left(np, \sqrt{np(1-p)}\right)$



Make sure all **information** you use is **relevant** to the **current question**.

Exemplar Exam Question

What **approximation** is **suitable?**

1) Seahorses can give birth to up to $\mathbf{2000}$ babies at once, with each baby having a $\mathbf{0.48}$ chance of being male.

Consider a seahorse that gives birth to $\mathbf{1500}$ babies. Calculate the approximate probability of there being more male than female babies.

[3 Marks]

3 marks, fair amount of **calculation**.

Exemplar Exam Question Answer

(i) Determine variables for binomial distribution.

1500 babies are being born.

This means number of trials $n = 1500$.

0.48 chance of a baby being male.

This means probability of success $p = 0.48$

Need more than 50% of 1500 babies to be male

This means number of successes is in range $X > 0.5 \times 1500 = 750$

n is large and p is roughly 0.5, so can approximate by normal distribution

Exemplar Exam Question Answer

$$n = 1500 \qquad p = 0.48$$

Determine variables μ and σ for normal distribution.

Mean and standard deviation for normal distribution approximation are given by

$$\mu = np$$

$$\sigma = \sqrt{np(1-p)}$$

Substitute in values

$$\mu = 1500 \times 0.48 = 720$$

$$\sigma = \sqrt{1500 \times 0.48 \times (1 - 0.48)} = 19.35$$

[1 Mark]

Exemplar Exam Question Answer

$\mu = 720$ $\sigma = 19.35$

Apply Continuity Correction

Looking for x' such that $P(X \geq x')$ is equivalent to $P(X > x = 750)$

$$x = 750 \rightarrow x' = 750.5$$

[1 Mark]

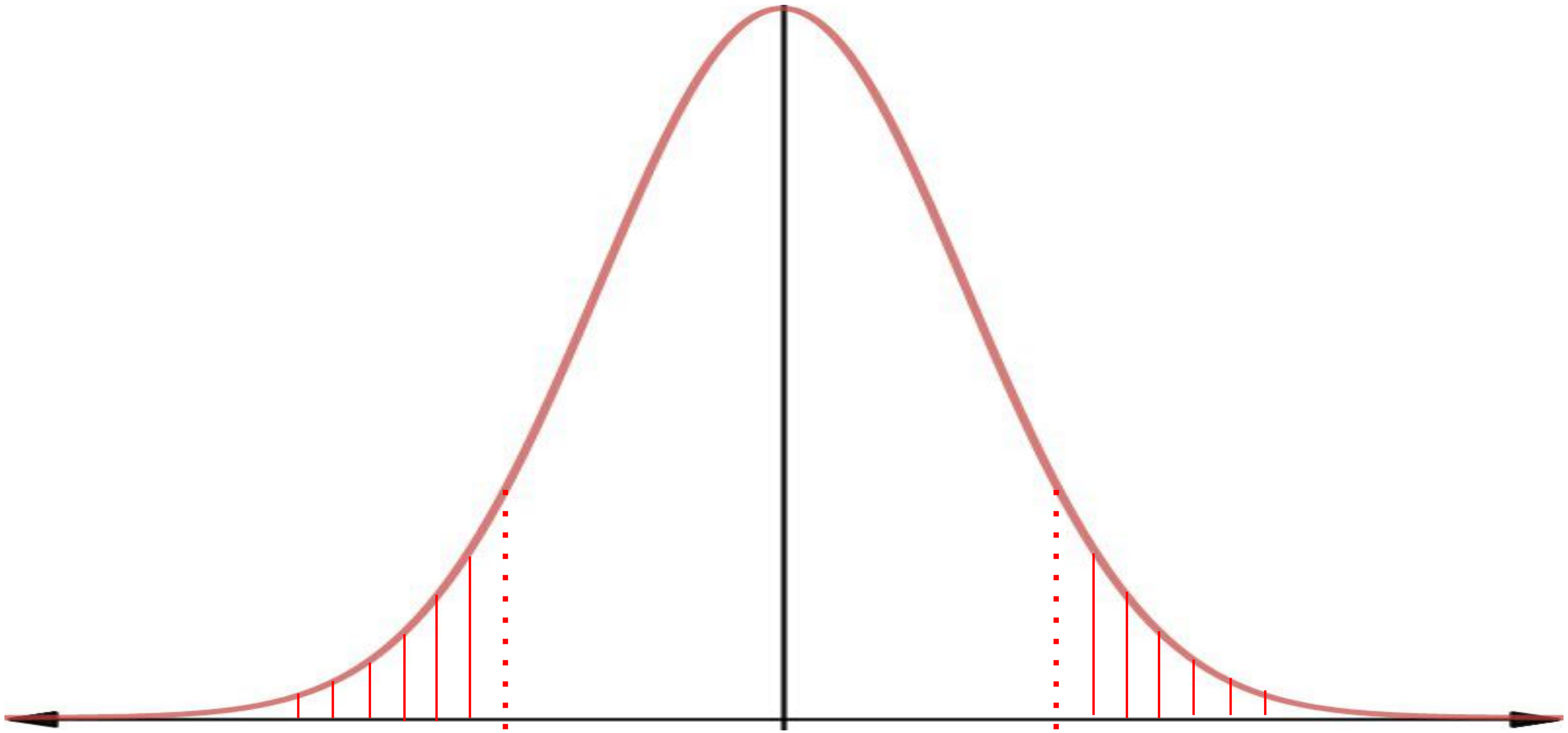
Find Probability

From calculator for given μ and σ

$$P(X \geq 750.5) = 0.0575$$

[1 Mark]

Hypothesis Testing



Specification Points - AQA

	Content
O1	Understand and apply the language of statistical hypothesis testing, developed through a binomial model: null hypothesis, alternative hypothesis, significance level, test statistic, 1-tail test, 2-tail test, critical value, critical region, acceptance region, p -value; extend to correlation coefficients as measures of how close data points lie to a straight line and be able to interpret a given correlation coefficient using a given p -value or critical value (calculation of correlation coefficients is excluded).

	Content
O2	Conduct a statistical hypothesis test for the proportion in the binomial distribution and interpret the results in context. Understand that a sample is being used to make an inference about the population and appreciate that the significance level is the probability of incorrectly rejecting the null hypothesis.

	Content
O3	Conduct a statistical hypothesis test for the mean of a Normal distribution with known, given or assumed variance and interpret the results in context.

Specification Points – OCR A

a) Understand and be able to use the language of statistical hypothesis testing, developed through a binomial model: null hypothesis, alternative hypothesis, significance level, test statistic, 1-tail test, 2-tail test, critical value, critical region, acceptance region, *p*-value.

Hypotheses should be stated in terms of parameter values (where relevant) and the meanings of symbols should be stated. For example,
"$H_0 : p = 0.7$, $H_1 : p \neq 0.7$, where p is the population proportion in favour of the resolution".

Conclusions should be stated in such a way as to reflect the fact that they are not certain. For example,
"There is evidence at the 5% level to reject H_0. It is likely that the mean mass is less than 500 g."
"There is no evidence at the 2% level to reject H_0. There is no reason to suppose that the mean journey time has changed."

Some examples of incorrect conclusion are as follows:
"H_0 is rejected. Waiting times have increased."
"Accept H_0. Plants in this area have the same height as plants in other areas."

b) Be able to conduct a statistical hypothesis test for the proportion in the binomial distribution and interpret the results in context.

c) Understand that a sample is being used to make an inference about the population and appreciate that the significance level is the probability of incorrectly rejecting the null hypothesis.

Learners should be able to use a calculator to find critical values.

Includes understanding that, where the significance level of a test is specified, the probability of the test statistic being in the rejection region will always be less than or equal to this level.

[The use of normal approximation is excluded.]

d) Recognise that a sample mean, $\bar{X}$, can be regarded as a random variable.

Learners should know and be able to use the result that if $X \sim N(\mu, \sigma^2)$ then $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

[The proof is excluded.]

e) Be able to conduct a statistical hypothesis test for the mean of a normal distribution with known, given or assumed variance and interpret the results in context.

Learners should be able to use a calculator to find critical values, but standard tables of the percentage points will be provided in the assessment.

[Test for the mean of a non-normal distribution is excluded.]

[Estimation of population parameters from a sample is excluded]

Specification Points – OCR MEI

MH1	Understand the process of hypothesis testing and the associated language.	Null hypothesis, alternative hypothesis. Significance level, test statistic, 1-tail test, 2-tail test. Critical value, critical region (rejection region), acceptance region, p-value.
H2	Understand when to apply 1-tail and 2-tail tests.	
H3	Understand that a sample is being used to make an inference about the population and appreciate that the significance level is the probability of incorrectly rejecting the null hypothesis.	For a binomial hypothesis test, the probability of the test statistic being in the rejection region will always be less than or equal to the intended significance level of the test, and will usually be less than the significance level of the test. Learners will not be tested on this distinction. If asked to give the probability of incorrectly rejecting the null hypothesis for a particular binomial test, either the intended significance level or the probability of the test statistic being in the rejection region will be acceptable.

H4	Be able to identify null and alternative hypotheses (H_0 and H_1) when setting up a hypothesis test based on a binomial probability model.	H_0 of form $p =$ a particular value, with p a probability for the whole population.
H5	Be able to conduct a hypothesis test at a given level of significance. Be able to draw a correct conclusion from the results of a hypothesis test based on a binomial probability model and interpret the results in context.	
H6	Be able to identify the critical and acceptance regions.	
MH7	Know that random samples of size n from $X \sim N(\mu, \sigma^2)$ have the sample mean Normally distributed with mean μ and variance $\frac{\sigma^2}{n}$.	
H8	Be able to carry out a hypothesis test for a single mean using the Normal distribution and be able to interpret the results in context.	In situations where either (a) the population variance is known or (b) the population variance is unknown but the sample size is large Learners may be asked to use a p-value or a critical region. H_0 of form $\mu =$ a particular value, where μ is the population mean. Significance level will be given.
H9	Be able to identify the critical and acceptance regions.	

Specification Points - Edexcel

5.1	**Understand and apply the language of statistical hypothesis testing, developed through a binomial model: null hypothesis, alternative hypothesis, significance level, test statistic, 1-tail test, 2-tail test, critical value, critical region, acceptance region, *p*-value;**	**An informal appreciation that the expected value of a binomial distribution is given by np may be required for a 2-tail test.**
	extend to correlation coefficients as measures of how close data points lie to a straight line.	Students should know that the product moment correlation coefficient r satisfies $\|r\| \le 1$ and that a value of $r = \pm 1$ means the data points all lie on a straight line.
	and	
	be able to interpret a given correlation coefficient using a given *p*-value or critical value (calculation of correlation coefficients is excluded).	Students will be expected to calculate a value of r using their calculator but use of the formula is not required. Hypotheses should be stated in terms of ρ with a null hypothesis of $\rho = 0$ where ρ represents the population correlation coefficient. Tables of critical values or a *p*-value will be given.

5.2	**Conduct a statistical hypothesis test for the proportion in the binomial distribution and interpret the results in context.**	
	Understand that a sample is being used to make an inference about the population.	**Hypotheses should be expressed in terms of the population parameter *p***
	and	
	appreciate that the significance level is the probability of incorrectly rejecting the null hypothesis.	**A formal understanding of Type I errors is not expected.**
5.3	Conduct a statistical hypothesis test for the mean of a Normal distribution with known, given or assumed variance and interpret the results in context.	Students should know that: If $X \sim \mathrm{N}(\mu, \sigma^2)$ then $\overline{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$ and that a test for μ can be carried out using: $\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim \mathrm{N}(0, 1^2)$. No proofs required. Hypotheses should be stated in terms of the population mean μ. Knowledge of the Central Limit Theorem or other large sample approximations is not required.

Language of Hypothesis Testing

We **perform** a **hypothesis test** to **determine** if a **sample** provides **significant evidence** that a **population parameter** has changed from a **previously known** or **assumed value**.

A **hypothesis** is a **statement** made about the **value** of a **population parameter**.

Language of Hypothesis Testing

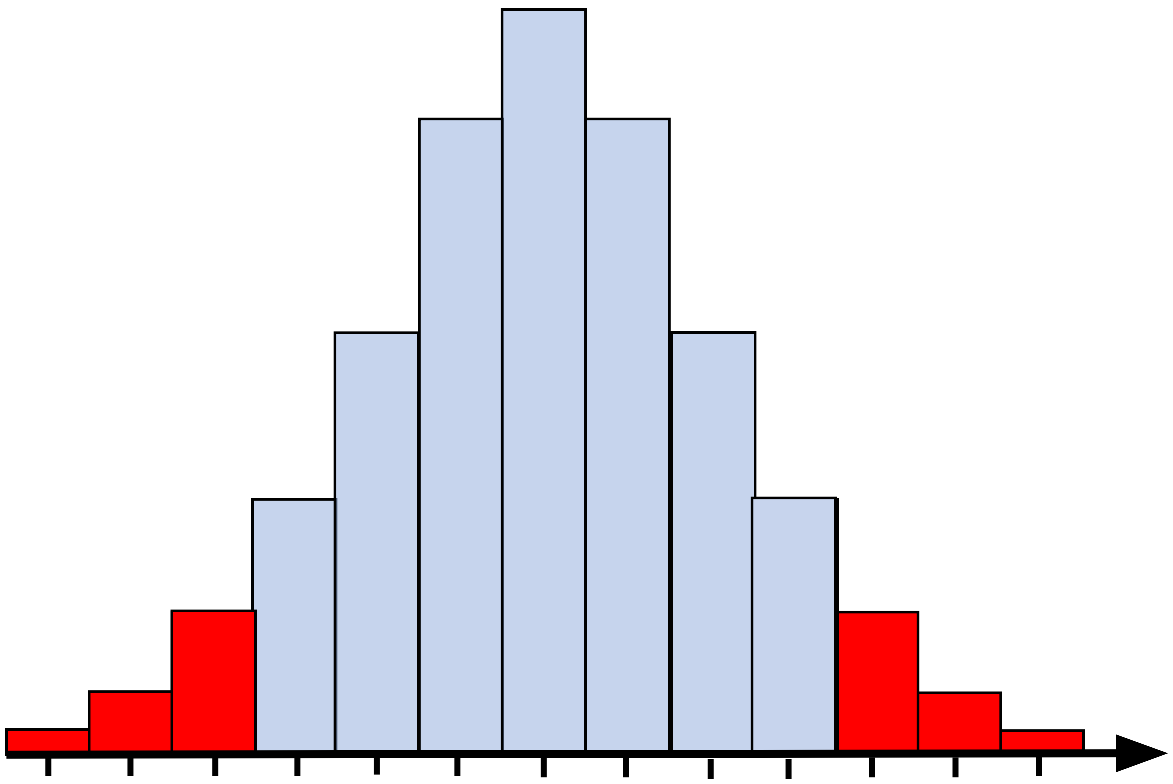
Hypothesis tests are carried out to a certain **significance level**.

- A **test statistic** is a **random variable calculated** from **sample data**.

Language of Hypothesis Testing

- The **critical region** for a **hypothesis test** is the **set of values** of the **test statistic** that **provide significant evidence** to **reject** the **null hypothesis**.
- The **critical value** is the **first value** to fall **inside** the **critical region**.

Language of Hypothesis Testing



Language of Hypothesis Testing

Question to **test understanding of definitions**

Exemplar Exam Question

Try to show **reasoning** if possible and make **final answer clear**

1) For each of the following hypotheses, state whether it will require a one-tail or two tail test.

(i) “Martin’s average score during a weekly game of bowling has increased.”
(ii) “A die is biased towards sixes.”
(iii) “There is no correlation between the number of awards won by actors in a film and the box office revenue of the film.”

[3 marks]

Can we link these to **types of hypothesis test?**

3 marks total, one mark for each

Exemplar Exam Question Answer

(i) Determine null and alternate hypothesis

Null hypothesis is that the average score hasn't changed

Alternative hypothesis is that average score has increased

Alternate hypothesis involves a > sign.

So this is a one-tailed test

[1 Mark]

Exemplar Exam Question Answer

(ii) Determine null and alternate hypothesis

Testing to see if the probability of rolling a 6 is greater than other numbers

Null hypothesis is all numbers have the same probability of being rolled

Alternative hypothesis is that there is a greater probability of rolling a 6

Alternate hypothesis involves a > sign.

So this is a one-tailed test

[1 Mark]

Exemplar Exam Question Answer

(iii) Determine null and alternate hypothesis

Testing to see if the product moment correlation is zero

Null hypothesis is that the product moment correlation is zero

Alternative hypothesis is that the product moment correlation is not zero

So this is a two-tailed test

[1 Mark]

Hypothesis Testing: Binomial Distribution

We test the **smaller cumulative probability** for the **successive trials from the test statistic** assuming the **null hypothesis is correct**

Hypothesis Testing: Binomial Distribution

Exemplar Exam Question

Pull relevant information, what is the **probability we're testing**?

1) David gets pains in his back 3 days a week on average. His friend Brian suggests replacing his mattress to see if it helps. David does so and doesn't get any pains over the next week.

   Test at 1% significance level if the new mattress has reduced the frequency of David's back pains. You should clearly state your null and alternative hypothesises. You may assume that the occurrence of David's back pains can be modelled by a binomial distribution.

   [4 marks]

Hypothesis testing, what **probabilities to calculate**?

Marks for **clearly listing variables and working**.

4 mark question for a **hypothesis test**, implies **calculating the probability itself isn't difficult**.

Exemplar Exam Question Answer

List given information:

Testing to find out if the new mattress reduces the frequency of David's back pains

Null hypothesis is that David's back pains occur 3 days a week

Alternative hypothesis is that back pains occur less than 3 days a week

Test statistic is that back pains have not occurred once all week

[1 Mark]

Exemplar Exam Question Answer

This is a binomial distribution problem with:

Hence

[1 Mark]

[1 Mark]

Exemplar Exam Question Answer

Compare to significance level and form conclusion.

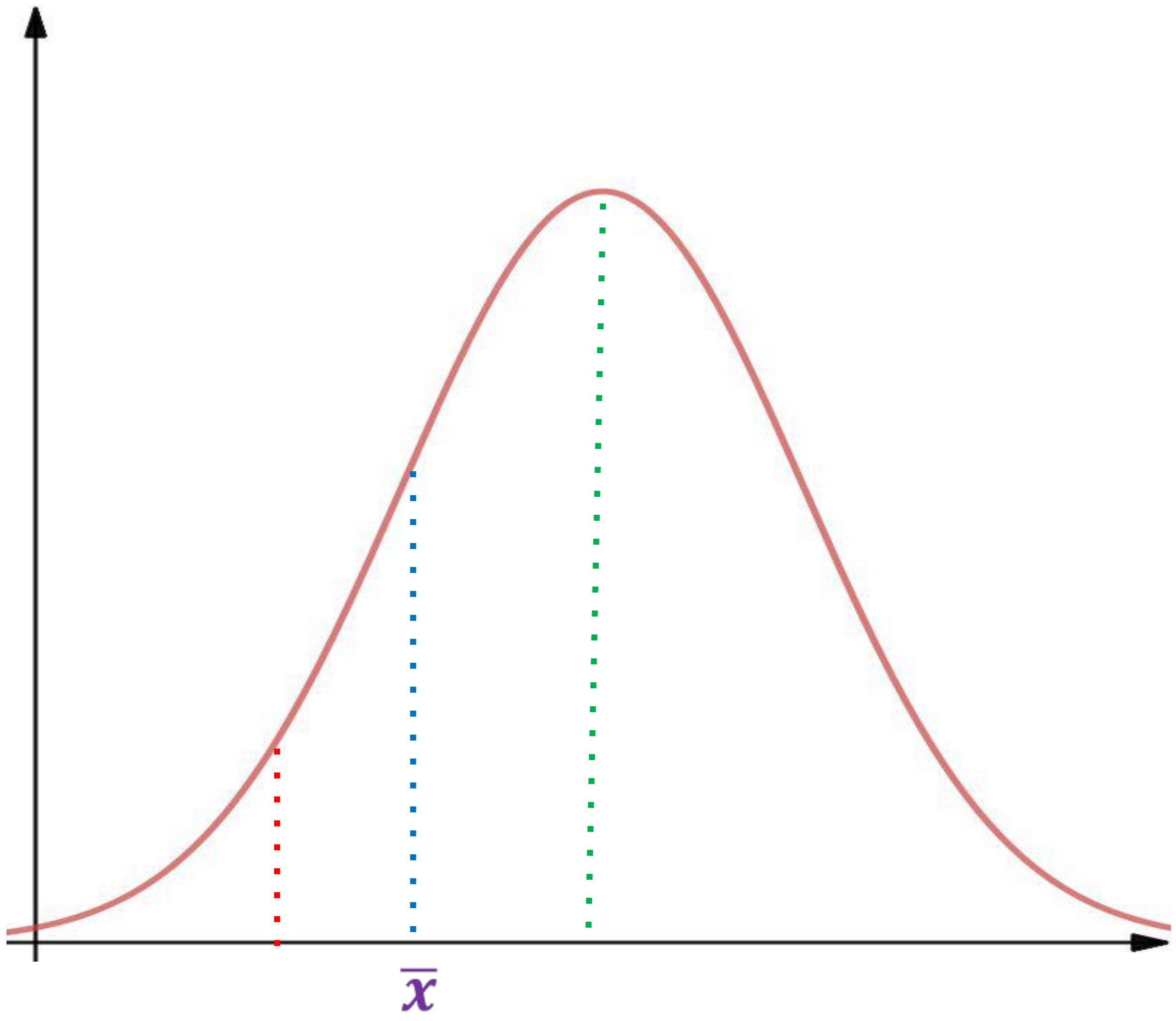
[1 Mark]

Hypothesis Testing: Normal Distribution

Hypothesis Testing: Normal Distribution

To **test** the **null hypothesis**:

- **Compare** to the **significance level**.



Exemplar Exam Question

Make sure to note **number of tests** for **calculated mean**.

Read all **information,** do we need a **one tailed or two tailed test**?

1) Cassie thinks her dad needs to fix or replace the internet router but her dad disagrees. Their internet speed varies, following a normal distribution with standard deviation 11Mbps. Cassie performs 3 speed tests and gets a mean download speed of 37.3Mbps. Last month their mean download speed was 46.2 Mbps. Test with 5% significance if their internet speed is, on average, slower than before.

[4 marks]

Testing **mean** for **normal distribution**.

4 marks, fair amount of work.

Exemplar Exam Question Answer

List given information

Testing to find out if Cassie's average internet speed is lower than before.

Null hypothesis is that internet speed matches previous average.

Alternative hypothesis is that internet speed is slower than previous average.

Test statistic is that Cassie has an average internet speed of 37.3 after 3 tests

[1 Mark]

Exemplar Exam Question Answer

Translate to Sample Mean Distribution

Use given information: Previous average is 46.2 Mbps with standard deviation 11 Mbps. Cassie tested her internet 3 times

[1 Mark]

Exemplar Exam Question Answer

From calculator

[1 Mark]

Draw conclusion from result.

Probability found is greater than 5% significance level

[1 Mark]

MINI MOCK PAPER

Exam Question

1. Casinos keep track of player win rates to catch potential cheaters. Staff have flagged a particular player for security to review

   (i) The player plays roulette 16 times, always betting on a single number out of 37 possible numbers, and wins 7 times. Calculate the probability of this happening to 3 significant figures

   [2 marks]

Exam Question

1. Casinos keep track of player win rates to catch potential cheaters. Staff have flagged a particular player for security to review

   (ii) They then sit at a slot machine. Each time they spin the machine it displays 25 symbols. Each symbol has a 10% chance of being a "Lucky 7", and if at least 5 are a "Lucky 7" he receives a bronze payout. Calculate the probability to 3 significant figures of the player receiving a bronze payout 3 times in a row.

   [2 marks]

Exam Question

1. Casinos keep track of player win rates to catch potential cheaters. Staff have flagged a particular player for security to review

(iii) 4 other casinos in the area have previously flagged the individual as a potential cheater. They found the player made an average profit of $£50$ during each trip to the 5 casinos. It is assumed that the profits of their customers follow a normal distribution with a mean of $-£20$ and a standard deviation of $£80$. Perform a hypothesis test to a 1% significance level that the individual has been cheating at the casinos

[3 marks]

Exam Question Answer

(i) Deduce required distribution

16 independent trials of fixed probability $\frac{1}{37} \Rightarrow X \sim \mathrm{B}\left(16, \frac{1}{37}\right)$

[1 Mark]

Calculate probability

$$P(X = 7) = {}^{16}C_7 \times \left(\frac{1}{37}\right)^7 \times \left(\frac{36}{37}\right)^{16-7}$$

$$= 9.42 \times 10^{-8}$$

[1 Mark]

Exam Question Answer

(ii) Calculate probability of a single bronze payout

25 independent symbols of fixed probability $0.1 \Rightarrow X \sim \mathrm{B}(25, 0.1)$

From calculator, $P(X \geq 5) = 0.0980$

[1 Mark]

Calculate probability of 3 consecutive bronze payouts

$$0.0980^3 = 9.41 \times 10^{-4}$$

[1 Mark]

Exam Question Answer

(iii) List given information

Testing to find out if induvial losses match average of population.

Null hypothesis is that they lose an average of £20 per casino trip. $H_0: \mu = \mu_0 = -20$

Alternative hypothesis is that they lose less than this (or makes a profit). $H_1: \mu > \mu_0 = -20$

Test statistic is that individual made an average profit of £50 in 5 trips

p-value is probability of getting this test result if individual's average profit is equal to that of population (H_0 is true).

[1 Mark]

Exam Question Answer

Translate to Sample Mean Distribution

Find p-value by calculating $\mathrm{P}(\bar{X} \geq 50)$ on Sample mean distribution

$$\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$$

Use given information: Population average is -20 with standard deviation 80. Individual's average is based off 5 trips

$$\bar{X} \sim N\left(-20, \frac{80^2}{5}\right)$$

Exam Question Answer

$\mu = -20$ $\sigma = \frac{80}{\sqrt{5}}$

Calculate value of $P(\bar{X} \geq 50)$

From calculator

$$P(\bar{X} \geq 50) = 0.0252$$

[1 Mark]

Draw conclusion from result.

Probability found is greater than 1% significance level

Therefore do not reject H_0, there is insufficient evidence to suggest individual makes greater profit/smaller losses than population

[1 Mark]