Edexcel International AS Maths: Statistics 1

Your notes

Normal Distribution

Contents

Your notes

Properties of a normal distribution

The binomial distribution is an example of a discrete probability distribution. The normal distribution is an example of a **continuous** probability distribution.

What is a continuous random variable?

- A continuous random variable (often abbreviated to CRV) is a random variable that can take **any value** within a range of infinite values
  - Continuous random variables **usually measure** something
  - For example, height, weight, time, etc

What is a continuous probability distribution?

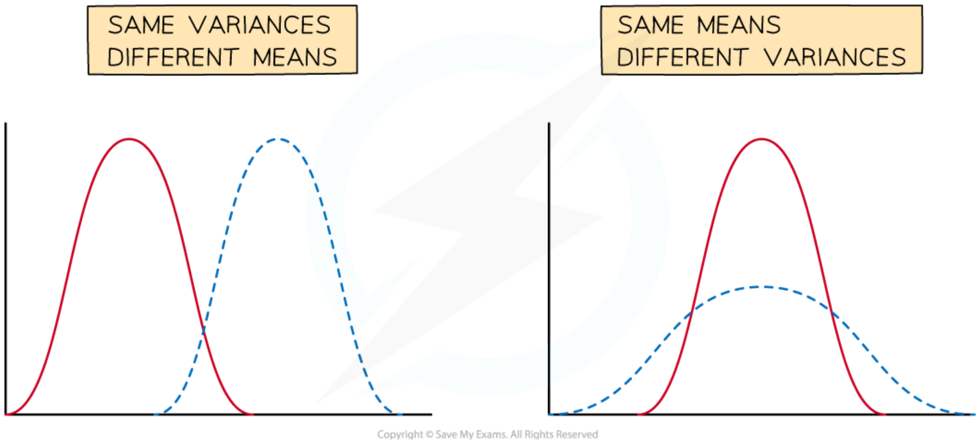
- A continuous probability distribution is a probability distribution in which the random variable *X* is continuous
- The probability of *X* being a **particular value is always zero**
  - **P(*X* = *k*) = 0** for any value *k*
  - Instead we define the **probability density function** f(*x*) for a specific value
  - We talk about the **probability** of X being within a **certain range**
- A continuous probability distribution can be represented by a continuous graph (the values for *X* along the horizontal axis and probability **density** on the vertical axis)
- The **area under the graph** between the points *x* = *a* and *x* = *b* is equal to **P(*a* ≤ *X* ≤ *b*)**
  - The **total area under the graph equals 1**
- As P(*X* = *k*) = 0 for any value k, it does not matter if we use strict or weak inequalities
  - P(X ≤ k) = P(X < k) for any value *k*

What is a normal distribution?

- A normal distribution is a **continuous probability distribution**
- The **continuous random variable** can follow a normal distribution if:
  - The distribution is **symmetrical**
  - The distribution is **bell-shaped**
- If *X* follows a normal distribution then it is denoted $X \sim N(\mu,\, \sigma^2)$
  - μ is the mean
  - σ^2 is the variance

- σ is the standard deviation

- If the **mean** changes then the graph is **translated horizontally**
- If the **variance** changes then the graph is **stretched horizontally**
  - A small variance leads to a tall curve with a narrow centre
  - A large variance leads to a short curve with a wide centre

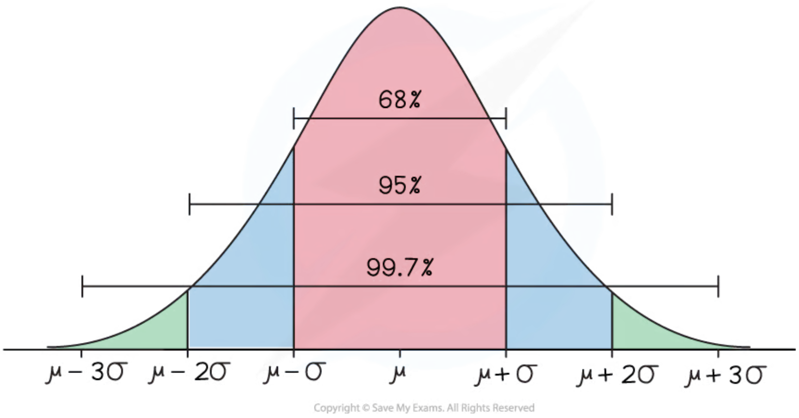


What are the important properties of a normal distribution?

- The **mean** is μ
- The **variance** is σ^2
  - If you need the standard deviation remember to square root this
- The normal distribution is symmetrical about $x = \mu$
  - Mean = Median = Mode = μ
- The normal distribution curve has two points of inflection
  - $x = \mu \pm \sigma$ (one standard deviation away from the mean)
- There are the results:
  - Approximately **two-thirds (68%)** of the data lies within **one standard deviation** of the mean ($\mu \pm \sigma$)
  - Approximately **95%** of the data lies within **two standard deviations** of the mean ($\mu \pm 2\sigma$)
  - Nearly **all of the data (99.7%)** lies within **three standard deviations** of the mean ($\mu \pm 3\sigma$)
- For any value x a z-score (or z-value) can be calculated which measures how many standard deviations x is away from the mean

Your notes

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



Modelling with a normal distribution

What situations can be modelled by a normal distribution?

- A lot of real-life continuous variables can be modelled by a normal distribution provided that the population is large enough and that the variable is **symmetrical** with **one mode**
- For a normal distribution X can take any real value, however values far from the mean (more than 4 standard deviations away from the mean) have a probability density of practically zero
  - This fact allows us to model variables that are not defined for all real values such as height and weight

What situations can not be modelled by a normal distribution?

- Variables which have **more than one mode** or **no mode**
  - For example, the number given by a random number generator
- Variables which are **not symmetrical**
  - For example, how long a human lives for

Worked Example

The random variable S represents the speeds (mph) of a certain species of cheetahs when they run. The variable is modelled using N(40, 100).

(a) Write down the mean and standard deviation of the running speeds of cheetahs.

Your notes

(b) State the two assumptions that have been made in order to use this model.

Answer:

a) $\mu = 40$ and $\sigma^2 = 100$

Square root to get standard deviation

Mean $\mu = 40$ mph

Standard deviation $\sigma = 10$ mph

b) We assume that the distribution of the speeds is

- symmetrical
- bell-shaped

Examiner Tips and Tricks

Remember the second number in N(x, y) is the variance, if you want the standard deviation then you need to square root it.

Your notes

Standard Normal Distribution

What is the standard normal distribution?

- The **standard normal distribution** is a normal distribution where the **mean is 0** and the **standard deviation is 1**
    - It is denoted by Z
    - $Z \sim \mathrm{N}(0, 1^2)$

Why is the standard normal distribution important?

- Calculating probabilities for the normal distribution can be difficult and lengthy due to its complicated probability density function
- The probabilities for the **standard normal distribution** have been calculated and laid out in the **table of the normal distribution** which can be found in your formula booklet
    - Nowadays, many calculators can calculate probabilities for any normal distribution, if yours does then the tables can be used as a check
- It is possible to map any normal distribution onto the standard normal distribution curve
- Mapping different normal distributions to the standard normal distribution allows distributions with different means and standard deviations to be compared with each other

How is any normal distribution mapped to the standard normal distribution?

- Any **normal distribution curve** can be transformed to the standard normal distribution curve by a **horizontal translation** and a **horizontal stretch**
    - Therefore, for $X \sim \mathrm{N}(\mu, \sigma^2)$ and $Z \sim \mathrm{N}(0, 1^2)$, we have the relationship:

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

- Probabilities are related by:
    - $\mathrm{P}(X < a) = \mathrm{P}\left(Z < \frac{a - \mu}{\sigma}\right)$
    - This is a very useful relationship for calculating probabilities for any normal distribution
    - As the normal distribution is a continuous distribution $\mathrm{P}(z_1 \le Z \le z_2) = \mathrm{P}(z_1 < Z < z_2)$ so you do not need to worry about whether the inequality is strict (< or >) or weak (≤ or ≥)

- A value of z = 1 corresponds with the *x*-value that is 1 standard deviation above the mean and a value of z = -1 corresponds with the *x*-value that is 1 standard deviation below the mean
- If a value of ***x*** is **less than the mean** then the *z* **-value will be negative**
- The function $\Phi(z)$ is used to represent $\mathrm{P}(Z < z)$

How is the table of the normal distribution function used?

- In your formula booklet you have the table of the normal distribution which provides probabilities for the standard normal distribution
  - The probabilities are provided for $\Phi(z) \; = \mathrm{P}(Z \leq z) = \mathrm{P}(Z < z)$
  - To find other probabilities you should use the symmetry property of the normal distribution curve
- The table gives probabilities for values of *z* between 0 and 4
  - For negative values of *z*, the symmetry property of the normal distribution is used
  - For values greater than *z* = 4 the probabilities are small enough to be considered negligible
- The tables give the probabilities to 4 decimal places
- To read probabilities from the normal distribution table for a *z* value :
  - Find the *z* value in a column labelled *z* at the top
  - The probability $\Phi(z)$ will be the probability directly to the right of the *z* value
  - So the value of $\Phi(1.23) = \mathrm{P}(Z \leq 1.23)$ would be found to the right of 1.23
    - $\mathrm{P}(Z \leq 1.23) = 0.8907$
- If the value of *z* is not listed then find the closest value on the table
  - *z* = 3.14 is not listed so use *z* = 3.15
  - *z* = 2.25 is not listed so use *z* = 2.24 or *z* = 2.26

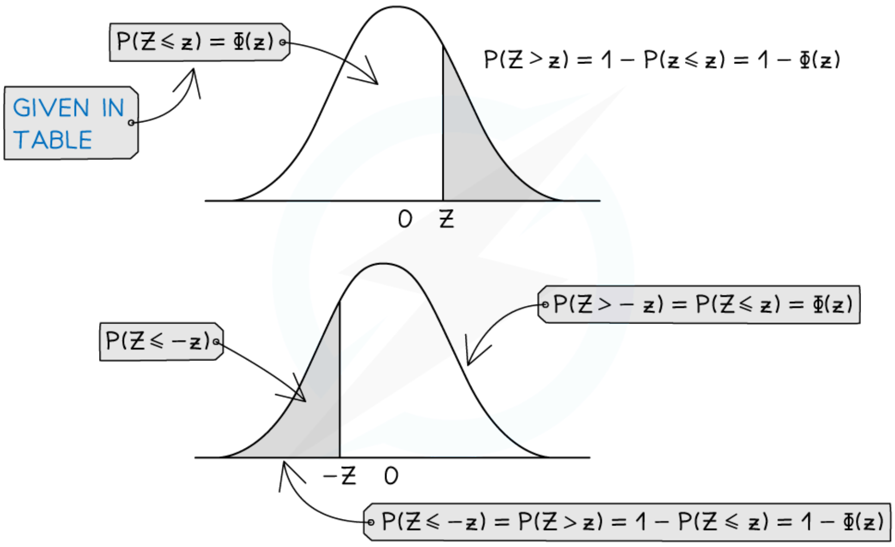
How is the table used to find probabilities that are not listed?

- The property that the area under the graph is 1 allows probabilities to be found for P(*Z* > *z*)
  - Use the formula $\mathrm{P}(Z > z) = 1 - \Phi(z)$
- The symmetrical property of the normal distribution gives the following results:
  - $\mathrm{P}(Z \leq z) = \mathrm{P}(Z \geq -z)$
  - $\mathrm{P}(Z \geq z) = \mathrm{P}(Z \leq -z)$
- This allows probabilities to be found for negative values of *z* or for $\mathrm{P}(Z > z)$

Your notes

- $\Phi(-z) = 1 - \Phi(z)$
- $P(Z > z) = 1 - \Phi(z)$

- In summary
  - $P(Z < z) = \Phi(z)$
  - $P(Z > z) = 1 - \Phi(z)$
  - $P(Z < -z) = 1 - \Phi(z)$
  - $P(Z > -z) = \Phi(z)$
- Drawing a sketch of the normal distribution will help find equivalent probabilities
  - Always consider whether the probability should be bigger or smaller then 0.5



How are z values found from the table of the normal distribution function?

- To find the value of *z* for which $P(Z \leq z) = p$ look for the value of *p* from within the table and find the corresponding value of *z*
  - If the probability is given to 4 decimal places most of the time the value will exist somewhere in the tables
- If your probability is 0.5 or greater look through the tables to find the corresponding *z* value
  - For $P(Z < z) \geq 0.5$ use the z value found in the table
  - For $P(Z > z) \geq 0.5$ take the negative of the z value found in the table
- If the probability is less than 0.5 you will need to **subtract it from one** before using the tables to find the corresponding *z* value

- For $\mathrm{P}(Z < z) < 0.5$ take the negative of the z value found in the table
- For $\mathrm{P}(Z > z) < 0.5$ use the z value found in the table
- **Always draw a sketch** so that you can see these clearly
- The formula booklet also contains a **table of the critical values of *z*** for $\mathrm{P}(Z > z) = p$
  - This gives *z* values to 4 decimal places for common probabilities
  - The probabilities in this table are 0.5, 0.4, 0.3, 0.2, 0.15, 0.1, 0.05, 0.025, 0.01, 0.005, 0.001 and 0.0005.

Worked Example

(a) By sketching a graph and using the table of the normal distribution, find the following:

(i) $\mathrm{P}(Z \le 0.96)$

(ii) $\mathrm{P}(Z > 0.96)$

(iii) $\mathrm{P}(Z \le -0.96)$

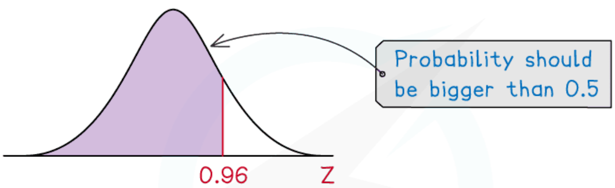
(iv) $\mathrm{P}(-0.96 < Z \le 0.96)$

(b) Find the value of Z such that $P(Z < z) = 0.7$

Answer:

For normal distribution the type of inequalities don't matter.
$P(Z < z) = P(Z \leqslant z)$

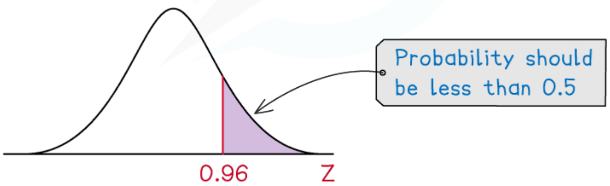
a) (i)



$P(Z < 0.96) = \Phi(0.96) = 0.8315$

$P(Z \leqslant 0.96) = 0.832$ (3.s.f.)

(ii)

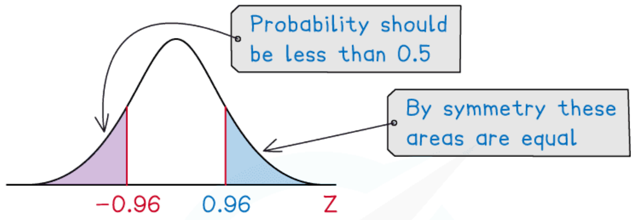


$P(Z > 0.96) = 1 - P(Z < 0.96) = 1 - 0.8315 = 0.1685$

$P(Z > 0.96) = 0.169$ (3.s.f.)

Your notes

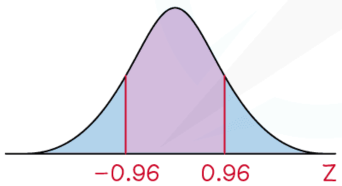
(iii)



$P(Z \leqslant -0.96) = P(Z > 0.96)$

$P(Z \leqslant -0.96) = 0.169$ (3.s.f.)

(iv)

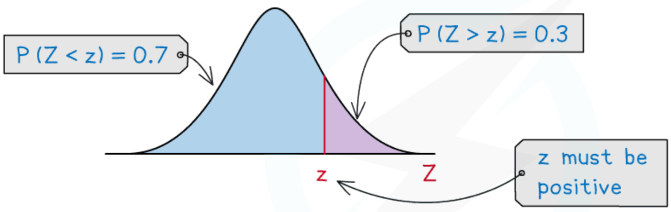


$P(-0.96 < Z \leqslant 0.96) = 1 - P(Z < -0.96) - P(Z > 0.96)$

$= 1 - 0.1685 - 0.1685$

$P(-0.96 < Z \leqslant 0.96) = 0.663$ (3.s.f.)

b)



Using the table of critical values

$p = 0.3$

$Z = 0.5244$

Examiner Tips and Tricks

- A sketch will always help you to visualise the required probability and can be used to check your answer
- Check whether the area shaded is more or less than 50% and compare this with your answer

Your notes

Calculations with Normal Distributions

Throughout this section we will use the random variable $X \sim \mathrm{N}(\mu, \sigma^2)$. For normal, X can take any real number. Therefore any values mentioned in this section will be assumed to be any real number.

Calculating Normal Probabilities

How do I find probabilities using a normal distribution?

- The **area under a normal curve** between the points $x = a$ and $x = b$ is equal to the **probability** $P(a < X < b)$
  - Remember for a normal distribution $P(a \le X \le b) = P(a < X < b)$ so you do not need to worry about whether the inequality is strict (< or >) or weak (≤ or ≥)
- The equation of a normal distribution curve is complicated so the **area must be calculated numerically**
- You will be expected to **standardise** all normal distributions to Z and **use the table of the normal distribution** to find the probabilities
  - It is likely that your calculator has a function that can find normal probabilities, if so it is a good idea to learn to use it so that you can check your probabilities
  - However you **must** show your calculations to get the z values and use the tables to get all the marks

How do I calculate the probability for a normal distribution?

- A random variable $X \sim \mathrm{N}(\mu, \sigma^2)$ can be coded to model the standard normal distribution $Z \sim \mathrm{N}(0, 1^2)$ using the formula

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

- You can calculate a probability $\mathrm{P}(X < x)$ using the relationship $\mathrm{P}(X < x) = \mathrm{P}\left(Z < \frac{x - \mu}{\sigma}\right)$
- **Always sketch** a quick diagram to visualise which area you are looking for
- Once you have determined the *z* value use the table of the normal distribution to find the probability
  - Refer to your sketch to decide if you need to subtract the probability from one
- The probability of a **single value** is **always zero** for a normal distribution
  - You can picture this as the area of a single line is zero
  - $\mathrm{P}(X = x) = 0$

- $\mathrm{P}(X < \mu) = \mathrm{P}(X > \mu) = 0.5$
  - You can look at which side of the mean *x* is on and the direction of the inequality to decide if your answer should be greater or less than 0.5
- As $\mathrm{P}(X = a) = 0$ you can use:
  - $\mathrm{P}(X < a) + \mathrm{P}(X > a) = 1$
  - $\mathrm{P}(X > a) = 1 - \mathrm{P}(X < a) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$
  - $\mathrm{P}(a < X < b) = \mathrm{P}(X < b) - \mathrm{P}(X < a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$

Your notes

Worked Example

The random variable $X \sim \mathrm{N}(20, 5^2)$. Calculate:

(a) $\mathrm{P}(X \le 22)$,

(b) $P(18 \le X \le 27)$

Answer:

a) $X \sim N(20, 5^2)$ $\quad \mu = 20, \quad \sigma = 5$

Read from the tables

$$P(X \leqslant 22) = P\left(Z \leqslant \frac{22-20}{5}\right) = P(Z \leqslant 0.4) = \Phi(0.4)$$
$$= 0.6554$$

$P(X \leqslant 22) = 0.655$ (3sf)

b) $P(18 \leqslant X < 27) = P\left(\frac{18-20}{5} \leqslant Z < \frac{27-20}{5}\right)$

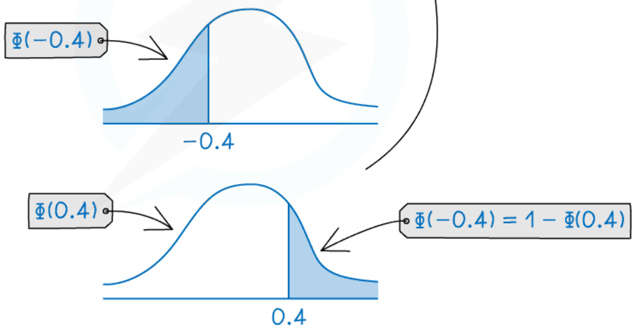
$= P(-0.4 \leqslant Z < 1.4)$

$= \Phi(1.4) - \Phi(-0.4)$

$= \Phi(1.4) - (1 - \Phi(0.4))$

$= 0.9192 - (1 - 0.6554)$

$= 0.5746$



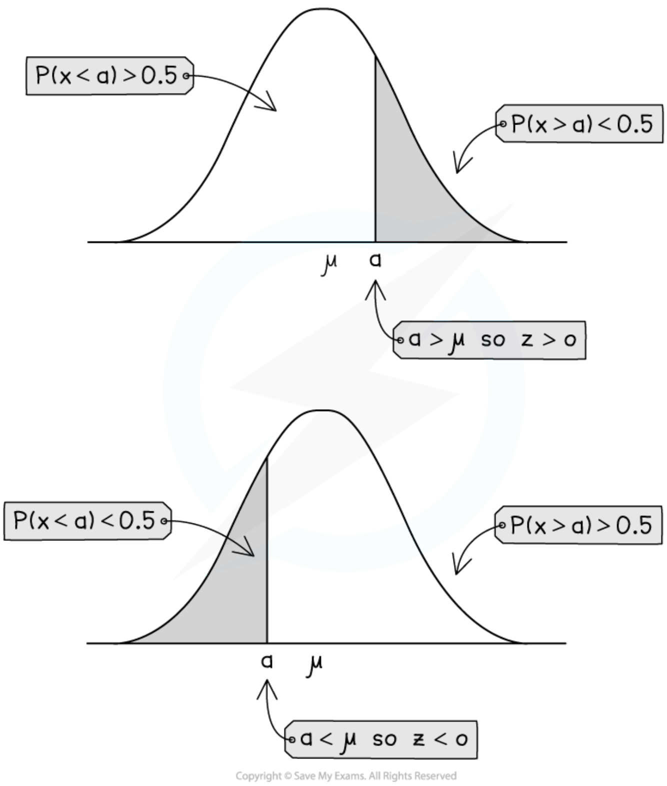
$P(18 \leqslant X < 27) = 0.575$ (3sf)

Your notes

Inverse Normal Distribution

Given the value of P(X < a) or P(X > a) how do I find the value of a?

- Given a probability you will have to look through the table of the normal distribution to locate the *z*-value that corresponds with that probability
- Look at whether your probability is **greater or less than 0.5** and the **direction of the inequality** to determine whether your *z*-value will be positive or negative
  - If $P(X < a)$ is **more than 0.5** or $P(X > a)$ is **less than 0.5** then *a* should be **bigger than the mean**
    - *z* will be positive
  - If $P(X < a)$ is **less than 0.5** or $P(X > a)$ is **more than 0.5** then *a* should be **smaller than the mean**
    - *z* will be negative
- You do not need to remember these, **a sketch will help you see it**
  - **Always sketch a diagram**

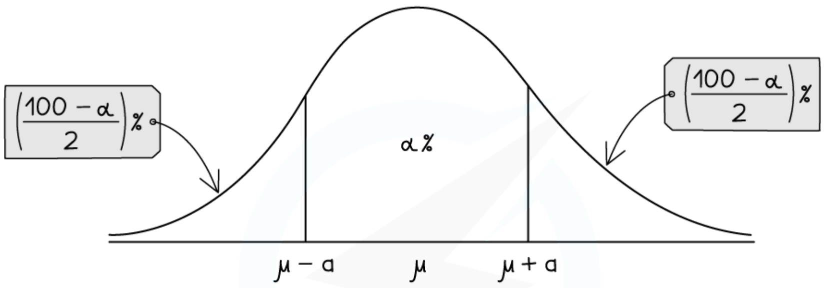


- If your probability is less than 0.5 you will need to **subtract it from one** to find the corresponding *z* value
  - Remember that the **position** of the z-value will not change, only the direction of the inequality
- Once you have the correct value substitute it into the formula $z = \frac{a-\mu}{\sigma}$ and solve to find the value of *a*
- Always check that your answer makes sense by considering where *a* is in relation to the mean

Given the value of P(μ– a < X < μ + a) I find the value of a ?

- A sketch making use of the symmetry of the graph is essential
- If you are given $P(\mu - a < X < \mu + a) = \alpha\%$ then $\mathrm{P}(X < \mu + a)$ will be $\left(\frac{100+\alpha}{2}\right)\%$
  - This is easier to see from a sketch than to remember
  - You can then look through the tables for the corresponding z-value and substitute into the formula $z = \frac{(\mu + a) - \mu}{\sigma} = \frac{a}{\sigma}$

Your notes



$$P(X < \mu + a) = \left(\alpha + \left(\frac{100-\alpha}{2}\right)\right)\%$$

$$= \left(\frac{2\alpha + 100 - \alpha}{2}\right)\%$$

$$= \frac{100+\alpha}{2}\%$$

Worked Example

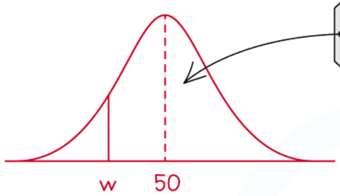
The random variable $W \sim \mathrm{N}(50, 36)$.

Find the value of w such that $P(W > w) = 0.7673$

Answer:

$W \sim N(50, 36)$ $\quad \mu = 50 \quad \sigma^2 = 36 \quad$ so $\sigma = 6$

$P(W > w) = 0.7673$



0.7673 > 0.5, w must be less than 50, z must be negative.

Find the z-value: $z = -$

Write the negative straight away.

$\Phi(0.73) = 0.7673$

$z = -0.73$

Substitute into the formula and solve for w:

$$-0.73 = \frac{w - 50}{6}$$

μ

σ

$w = -0.73(6) + 50$

$= 45.62$

Check the answer is the correct side of μ

$w = 45.6$ (3.s.f.)

Your notes

Your notes

Examiner Tips and Tricks

- The most common mistake students make when finding values from given probabilities is forgetting to check whether the z-value should be negative or not. Avoid this by checking early on **using a sketch** whether z is positive or negative and writing a note to yourself before starting the other calculations.

Your notes

Finding Sigma and Mu

How do I find the mean (μ) or the standard deviation (σ) if one of them is unknown?

- If the **mean** or **standard deviation** of the $X \sim N(\mu, \sigma^2)$ is **unknown** then you will need to use the **standard normal distribution**
- You will need to use the formula
  - $z = \frac{x-\mu}{\sigma}$ or its rearranged form $x = \mu + \sigma z$
- You will be given a **probability for a specific value** of $x (P(X < x) = p \text{ or } P(X > x) = p)$
- To find the unknown parameter:
- **STEP 1**: **Sketch** the normal curve
  - Label the known value and the mean
- **STEP 2**: **Find** the **z-value** for the given value of *x*
  - Use the **table of the Normal Distribution** to find the value of *z* such that $P(Z < z) = p$ or $P(Z > z) = p$
  - Make sure the direction of the inequality for *Z* is consistent with *X*
  - The table gives the z-value to four decimal places to **avoid rounding errors**
    - Use the sketch to help you decide whether your *z* value is positive or negative
    - You should use the 4 decimal places throughout your calculations so that your final answer can be rounded to 3 significant figures
- **STEP 3**: **Substitute** the known values into $z = \frac{x-\mu}{\sigma}$ or $\boldsymbol{x = \mu + \sigma z}$
  - You will be given x and one of the parameters (μ or σ) in the question
  - You will have calculated z in STEP 2
- **STEP 4**: **Solve** the equation

How do I find the mean (μ) and the standard deviation (σ) if both of them are unknown?

- If **both** of them are **unknown** then you will be given two probabilities for two specific values of *x*
- The process is the same as above

- You will now be able to **calculate two z-values**
- You can form **two equations** (rearranging to the form $x = \mu + \sigma z$ is helpful)
- You now have to **solve the two equations simultaneously**
- Be careful not to mix up which z-value goes with which value of x

Your notes

Worked Example

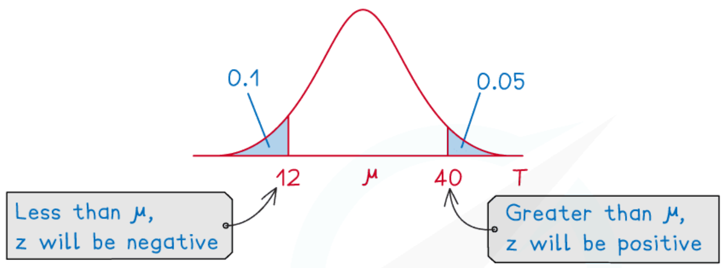
It is known that the times, in minutes, taken by students at a school to eat their lunch can be modelled using a normal distribution with mean μ minutes and standard deviation σ minutes.

Given that 10% of students at the school take less than 12 minutes to eat their lunch and 5% of the students take more than 40 minutes to eat their lunch, find the mean and standard deviation of the time taken by the students at the school.

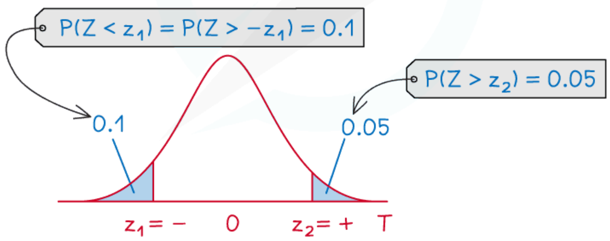
Answer:

Let the times in minutes, be $T \sim N(\mu, \sigma^2)$

Step 1: Draw a diagram to represent the distribution:



Step 2: Find the z-values for 12 and 40



Use the table of the critical values to find the z-values such that $P(Z > -z_1) = 0.1$ and $P(Z > z_2) = 0.05$

$z_1 = -1.2816$ $\quad z_2 = 1.6449$

Step 3: Substitute the z-values into the formula $z = \frac{T - \mu}{\sigma}$

1

$$-1.2816 = \frac{12 - \mu}{\sigma}$$

2

$$1.6449 = \frac{40 - \mu}{\sigma}$$

Step 4: Solve the equations to find μ and σ

1 $\longrightarrow$ $-1.2816\sigma = 12 - \mu$

2 $\longrightarrow$ $1.6449\sigma = 40 - \mu$ (subtract)

$$-2.9265\sigma = -28$$

$$\sigma = 9.5677$$

Substitute into 1 for μ:

$$-1.2816\,(9.5677...) = 12 - \mu$$

$$\mu = 12 + 1.2816\,(9.5677...)$$

$$= 24.2620...$$

$\mu = 24.3$ mins (3.s.f.)

$\sigma = 9.57$ mins (3.s.f.)

Your notes

Examiner Tips and Tricks

- These questions are normally given in context so make sure you identify the key words in the question. Check whether your z-values are positive or negative and be careful with signs when rearranging.