 Year 12 Mathematics Standard 2

| MS-S5 The normal distribution | Unit duration |
| --- | --- |
| The normal distribution was developed from a model originally propounded by Abraham de Moivre, an 18th century statistician and consultant to gamblers. The normal distribution curve is sometimes called the ‘bell curve’ or the ‘Gaussian curve’ after the mathematician Gauss, who played an important role in its development. Early on, the normal distribution was used to model and manage the errors received during astronomy investigations, though its applications are more wide ranging today. The normal distribution is the most important and widely used distribution in business, statistics and government. Indeed its importance stems primarily from the fact that the distributions of many natural phenomena are at least approximately normally distributed. The normal distribution unit will further develop students' understanding and use of the statistical investigation process; the collection, display, analysis and interpretation of data to identify and communicate key information. It will allow students to engage with the Work and Enterprise and Applications and Modelling priorities. Students will develop their skills to select and apply appropriate mathematical techniques and problem-solving strategies through work-related experiences. | 2 to 3 weeks |

| Subtopic focus | Outcomes |
| --- | --- |
| The principal focus of this subtopic is to develop an understanding of the properties of the normal distribution and the value of relative measure in the analysis and comparison of datasets arising from random variables that are normally distributed. Students develop techniques to analyse normally distributed data and make judgements in individual cases justifying the reasonableness of their solutions. | A student:* analyses representations of data in order to make inferences, predictions and draw conclusions MS2-12-2
* solves problems requiring statistical processes, including the use of the normal distribution, and the correlation of bivariate data MS2-12-7
* chooses and uses appropriate technology effectively in a range of contexts, and applies critical thinking to recognise appropriate times and methods for such use MS2-12-9
* uses mathematical argument and reasoning to evaluate conclusions, communicating a position clearly to others and justifying a response MS2-12-10

Related Life Skills outcomes: MALS6-2, MALS6-9, MALS6-13, MALS6-14 |

| Prerequisite knowledge | Assessment strategies |
| --- | --- |
| Students should have studied the topics of MS-S1 Data analysis, describing data arising from a single variable, and MS-S1 Relative frequency and probability prior to studying this topic. | **Is mathematical modelling better than guessing?** is an investigative task in which students interact with the statistical analysis process to solve to problem of their choice. |

All outcomes referred to in this unit come from [Mathematics Standard 2019](https://www.educationstandards.nsw.edu.au/wps/portal/nesa/11-12/stage-6-learning-areas/stage-6-mathematics/mathematics-standard-2017) Syllabus © NSW Education Standards Authority (NESA) for and on behalf of the Crown in right of the State of New South Wales, 2017

Glossary of terms

| Term | Description |
| --- | --- |
| Empirical rule | The empirical rule is a statistical rule that states that for a normal distribution, almost all data falls within three standard deviations (denoted by $σ$) of the mean (denoted by $µ$). Broken down, the empirical rule shows that 68% falls within the first standard deviation $(µ \pm σ)$, 95% within the first two standard deviations $(µ \pm 2σ)$, and 99.7% within the first three standard deviations $(µ \pm 3σ)$. |
| Normal distribution | The normal distribution is a type of continuous distribution where the mean, median and mode are equal and the scores are symmetrically arranged either side of the mean. The graph of a normal distribution is often called a ‘bell curve’ due to its shape, as shown below.graph demonstrating bell curve |
| Populations | A population is the complete set of individuals, objects, places, etc. that we want information about. A census is an attempt to collect information about the whole population. |
| Standard deviation | Standard deviation is a measure of the spread of a dataset. It gives an indication of how far, on average, individual data values are spread from the mean. |
| Z-score (standardised score) | A z-score is a statistical measure of how many standard deviations a raw score is above or below the mean. A $z$-score can be positive or negative, indicating whether it is above or below the mean, or zero. Also known as a standardised score. |

| Lesson sequence | Content | Suggested teaching strategies and resources  | Date and initial | Comments, feedback, additional resources used |
| --- | --- | --- | --- | --- |
| Introducing the normal distribution(1 or 2 lessons) | * Recognise a random variable that is normally-distributed, justifying their reasoning, and draw an appropriate ‘bell-shaped’ frequency distribution curve to represent it
	+ Identify that the mean and median are approximately equal for data arising from a random variable that is normally-distributed
 | Identifying the characteristics of the normal distribution* Introduce students to the idea of naturally occurring random events. These are events that happen *naturally* without bias and without limits.
* Staff may like to run this [Galton board simulation](https://www.mathsisfun.com/data/quincunx.html) as an example of a naturally occurring random event. Ask students to describe the shape of the frequency distribution table produced as the number of trials becomes large.
* Staff may like to run this [popcorn popping](https://www.youtube.com/watch?v=YmOsDTczOFs) activity as another example of a naturally occurring random event.
* Define the normal distribution

This is a graph demonstrating an example of a normal distribution.* Data that is normally distributed has the same mean, mode and median and is symmetrical about the mean. These characteristics only occur with a large number of data points (scores). Samples may be normally distributed but do not exhibit these characteristics.
* Students investigate the popular times function of a google search by entering a business name and a location, for example “Coles Newcastle West”. During this activity, students will be challenged to identify normally distributed patterns and develop an awareness of when and why these patterns exist or, conversely, when and why they do not.

A screenshot from Google showing the popular times for a Coles supermarket on a Wednesday. The display is normally distributed.The distribution above shows strong signs of a normal distribution.A screenshot from Google showing the popular times for a Coles supermarket on a Wednesday. The display appears normally distributed but contains an upward peak afer 9pm, indicating a trend for late night shopping.The distribution above shows signs of a normal distribution but there is a trend for late night shopping.**Resource**: the-shape-of-business-worksheet.DOCXUnderstanding the standard deviation* Students should investigate the idea of spread and how it is represented mathematically. Staff could show the [Bean Machine YouTube clip](https://www.youtube.com/watch?v=3m4bxse2JEQ) for different sized beans. Lead students into discussion regarding the spread of the large beans compared to the smaller beans and that the distribution for the larger beans will have a smaller standard deviation.

This is a graph that shows three different normal distributions with the same means but varying standard deviations. * This [interactive normal distribution link](http://www.intmath.com/counting-probability/normal-distribution-graph-interactive.php) allows students to interact with the standard deviation $σ$ by the sliding the value from small (less than 1) to large (greater than 1), and this will automatically change the shape of the bell-curve as they slide the standard deviation value.
 |  |  |
| Standardising population data by calculating z-scores(1 or 2 lessons) | * Calculate the z-score (standardised score) corresponding to a particular value in a dataset
	+ use the formula $z=\frac{x-μ}{σ}$, where $μ$ is the mean and $σ$ is the standard deviation  Information and communication technology capability icon
	+ describe the z-score as the number of standard deviations a value lies above or below the mean
	+ recognise that the set of z-scores for data arising from a random variable that is normally distributed has a mean of 0 and standard deviation of 1
 | Calculating a z-score* Lead students to understand that a z-score is standardised score on a scale between -4 and 4 (approximately). Each score is standardised using the mean and standard deviation as part of the calculation.
	+ A positive z-score indicates that the score is above the mean
	+ a zero z-score is equal to the mean
	+ a negative z-score is below the mean
* The z-score indicates how many standard deviations away from the mean the score is. For example, a z-score of -2 should be interpreted as a score which is 2 standard deviations below the mean.
* The z-score distribution is normally distributed with a mean $μ=0$ and standard deviation $σ=1$
* The heights of students in a class are normally distributed with a mean of 150cm and a standard deviation of 5cm, calculate the z-score of a student who is 165cm and another who is 140cm tall.

$x=165$ and $x=140$ respectively$$\overbar{x}=150$$$$s=5$$Student 1: $z=\frac{x-\overbar{x}}{s}\rightarrow \frac{165-150}{5}=3$Student 2: $z=\frac{x-\overbar{x}}{s}\rightarrow \frac{140-150}{5}=-2$* This how-to guide for calculator skills provides instructions for calculating the mean, standard deviation and z-score using a Casio scientific calculator.

**Resource**: how-to-guide-calculator-skills.DOCX* This how-to guide for spreadsheet skills provides instructions for calculating the mean, standard deviation and z-score using an Excel spreadsheet.

**Resource**: how-to-guide-spreadsheet-skills.DOCX**NESA exemplar question**1. The lifetime of a particular lightbulb is normally distributed with mean of $1020$ hours and standard deviation $85$ hours. Find the probability that a lightbulb of the same make chosen at random has a lifetime between $1003$ and $1088$ hours. **Note:** Normal distribution tables would be used to answer this question.
2. Find the probability that a person selected at random from a pool of people that took a test on which the mean was $100$ and the standard deviation was $15$ will have a score of:

**Note:** [normal distribution tables](http://www.ztable.net) would be used to answer this question.* 1. between $100$ and $120$
	2. of at most $120$
	3. of greater than $120$

**Resource**: ms-s5-nesa-examplar-question-solutions.DOCX |  |  |
| Comparing scores from different populations(1 lesson) | * Use calculated z-scores to compare scores from different datasets, such as comparing students’ subject examination scores **AAM**
 | Applying z-scores to compare scores from different populations relative to the population* Lead students to debate the following “Your exam mark from Class A was 91% and your mark from Class B was 82%. In which class did you perform better?”
* Lead students to the idea that the highest mark is not necessarily enough to gauge performance and that you need more information.
* Introduce the mean scores to further the debate. For Class A the mean was 93% and in class B the mean was 86%. Again, ask the question “In which class did you perform better?”. Students may reply that both scores are below the mean, but Class A is closer to the mean, therefore a better score.
* Finally introduce the standard deviation to further the date again. For Class A the standard deviation is 2 and for Class B is 5. Again, ask the question “In which class did you perform better?”. Lead students to the idea of standardising the distributions, through z-scores, to compare fairly.

Class A: $z=\frac{91-93}{2}=-1$Class B: $z=\frac{82-86}{5}=-0.8$ |  |  |
| Understanding the empirical rule(1 or 2 lessons) | * Use collected data to illustrate that, for normally distributed random variables, approximately 68% of data will have z-scores between -1 and 1, approximately 95% of data will have z-scores between -2 and 2 and approximately 99.7% of data will have z-scores between -3 and 3 (known as the Empirical Rule)
	+ Apply the Empirical Rule to a variety of problems
	+ Indicate by shading where results sit within the normal distribution, such as where the top 10% of data lies
 | Defining the percentile regions of the normal distribution: The empirical rule* Empirical means that conclusions or findings are determined from observations or data rather than theory. The empirical rule is a breakdown of a standardised normal distribution into regions that determine the likelihood of a future trial as a percentage. These percentages have been determined by running a normally distributed experiment with a large amount of trials, as close to infinity as possible.
* The key points are:

This is a graph demonstrating the empirical rule for a normal distribution graph.This is an example of a normal distribution graph with percentile regions indicated for each standard deviation.* + 68% of scores have a z-score between -1 and 1 (probable in this range)
	+ 95% of scores have a z-score between -2 and 2 (very probable in this range)
	+ 99.7% of scores have a z-score between -3 and 3 (almost certain in this range)
	+ 50% of scores have a z-score between 0 and $\pm \infty $ (think symmetry)
* The empirical rule is stated on the Standard 1 and 2 reference sheet.
* Students need to calculate the percentage of trials determined by a region by using the Empirical rule.
* Students may like to explore this [interactive curve](https://www.mathsisfun.com/data/standard-normal-distribution-table.html) showing the percentile breakdown using tabled values. For extension, students could be introduced to the z-scores table, i.e. the percentage of trials that lie between $z=0$ and $z=0.68$ can be read from the table by finding the intersection point of 0.6 and 0.08. In this example this is the value 0.2517, which means that 25.17% lies in the region stated. Please note that there is no explicit reference to the z-scores table but there is uncertainty whether students are expected to be able to read from it.
* Staff may like to us the following resources to investigate the empirical rule with authentic data from Statistics Online Computational Resource (SOCR)

**Resource**: demonstrating-the-empirical-rule-teacher-guide.DOCX, demonstrating-the-empirical-rule.XLSX, demonstrating-the-empirical-rule-teacher-guide-SOCR.DOCX, demonstrating-the-empirical-rule-SOCR.XLSX **NESA exemplar question**1. A machine is set for the production of cylinders of a mean diameter $5.00$cm, with standard deviation $0.020$cm. Assuming a normal distribution, between which values will $95\%$ of the diameters lie? If a cylinder, randomly selected from this production, has a diameter of $5.070$cm, what conclusions could be drawn?
 |  |  |
| Applying normal distribution models to quality assurance processes(1 lesson) | * Use z-scores to identify probabilities of events less or more extreme than a given event AAM
	+ use statistical tables to determine probabilities Literacy icon
	+ use technology to determine probabilities  Information and communication technology capability icon
* Use z-scores to make judgements related to outcomes of a given event or sets of data AAM
 | Applying normal distribution models to quality assurance processes* The normal distribution can be applied to model variation or errors. An example of this is within large scale manufacturing processes which may produce thousands of products each day. These products must meet certain standards regarding quality and quantity. Though the processes involved in manufacturing a product are as streamlined as possible, each product is susceptible to error. For example a manufacturing process aims to fill bags with 1kg of sugar; however, variation in the process can cause some bags to be filled with more or less than 1kg. By sampling products regularly during the manufacturing process, the mean and standard deviation of the sample are calculated and used to develop benchmark statements, like 97.5% of all bags will contain 1kg of sugar.

**NESA exemplar question**1. Packets of rice manufactured are each labelled as having a mass of 1kg. The mass of these packets is normally distributed with a mean of 1.02kg and a standard deviation of 0.01kg. Complete the following table:

This is a table for students to complete for the rice manufacturing example. It shows the mass in kilograms and their equivalent z-scores.* 1. What percentage of packets will have a mass less than 1.02kg?
	2. What percentage of packets will have a mass between 1.00 and 1.04kg?
	3. What percentage of packets will have a mass between 1.00 and 1.02kg?
	4. What percentage of packets will have a mass less than the labelled mass?
	5. Why do you think the mean is above the actual labelled mass?
	6. How many standard deviations is the mean above the labelled mass? Do you think this would be a good point for all manufacturers and why?

**Resource**: ms-s5-nesa-examplar-question-solutions.DOCX* The activity ‘to be sure’ challenges students to use the empirical rule to determine the likelihood of certain outcomes as percentages.

**Resource**: to-be-sure-activity.DOCX* This student activity [bags of sugar](http://www.mathsisfun.com/data/standard-normal-distribution.html) demonstrates the principles for meeting quality assurance statements.
* Solutions for NESA exemplar questions from the topic guidance can be found in this resource.
 |  |  |
| Investigating using normal distributions(1 lesson) |  | **Investigating using normal distributions*** Students can investigate the characteristics of the normal distribution through the ‘we have a winner’ activity. During this task, students are asked design a fair game using the Galton board.

**Resource**: we-have-a-winner-worksheet.DOCX* Staff may like students to investigate human and social characteristics from the [Australian Bureau of Statistics](http://www.abs.gov.au/) website. For example
	+ [Population by age and region](http://www.abs.gov.au/ausstats/subscriber.nsf/log?openagent&32350ds0001_nsw_2010_2015.xls&3235.0&Data%20Cubes&68CC4123E1B99078CA25801200168034&0&2015&18.08.2016&Latest): Students investigate their local population by age by creating a column graph representing this information. A suggested activity is to allow students to compare the populations of similar towns, for example Taree and Kempsey, and interpret their results. There will be opportunities to compare Indigenous and non-indigenous populations here.
	+ [Fertility rates by age from 2000 to 2010](http://www.abs.gov.au/AUSSTATS/subscriber.nsf/log?openagent&33010do001_2010.xls&3301.0&Data%20Cubes&186501EC1649A6FCCA25792F00160FE2&0&2010&25.10.2011&Previous): Students create column graphs to display the fertility rates by age for a given year. A suggested activity is to allow students to compare the fertility rates by age from 2000 to 2010. Students should comment on their findings and describe current trends.
* The following student activity, MVP breaking the bank, investigates whether the Cleveland Cavaliers and New England Patriots are paying too much for their star player compared to the rest of their roster.

**Resource**: mvp-breaking-the-bank-activity.DOCX |  |  |

Reflection and evaluation

Please include feedback about the engagement of the students and the difficulty of the content included in this section. You may also refer to the sequencing of the lessons and the placement of the topic within the scope and sequence. All information communication technologies (ICT), literacy, numeracy and group activities should be recorded in the ‘Comments, feedback, additional resources used’ section.