 Year 12 Mathematics Advanced

| MA-S3 Random variables | Unit duration |
| --- | --- |
| The topic Statistical Analysis involves the exploration, display, analysis and interpretation of data to identify and communicate key information.A knowledge of statistical analysis enables careful interpretation of situations and an awareness of the contributing factors when presented with information by third parties, including its possible misrepresentation.The study of statistical analysis is important in developing students' ability to recognise, describe and apply statistical techniques in order to analyse current situations or to predict future outcomes. It also develops an awareness of how conclusions drawn from data can be used to inform decisions made by groups such as scientific investigators, business people and policy-makers. | 4 weeks |

| Subtopic focus | Outcomes |
| --- | --- |
| The principal focus of this subtopic is to introduce students to continuous random variables, the normal distribution and its use in a variety of contexts.Students develop understanding of the probability density function, how integration or the area under the function determines probabilities to solve problems involving random variables, and an understanding of the normal distribution, its properties and uses. Students make connections between calculus skills developed earlier in the course and their applications in Statistics, and lay the foundations for future study in this area. | A student:* solves problems using appropriate statistical processes MA12-8
* chooses and uses appropriate technology effectively in a range of contexts, models and applies critical thinking to recognise appropriate times for such use MA12-9
* constructs arguments to prove and justify results and provides reasoning to support conclusions which are appropriate to the context MA12-10
 |

| Prerequisite knowledge | Assessment strategies |
| --- | --- |
| This topic builds upon the statistics and data concepts explored in Stage 5 and MA-S1, describing skewed data, symmetry in data and the introduction to random variables. Students will need to be familiar with the notations and definitions of integral calculus.  | Is a Normal Distribution Appropriate?**Resource**: ma-s3-assessment-is-a-normal-distribution-appropriate.DOCXorHow well can mathematics predict outcomes?**Resource**: ma-s2\_3-how-well-can-mathematics-predict-outcomes.DOCX(includes MA-S2 Descriptive Statistics and Bivariate Data Analysis) |

All outcomes referred to in this unit come from [Mathematics Advanced](http://educationstandards.nsw.edu.au/wps/portal/nesa/11-12/stage-6-learning-areas/stage-6-mathematics/mathematics-advanced-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 |
| --- | --- |
| Continuous random variable | A continuous random variable is a numerical variable that can take any value along a continuum. |
| Cumulative distribution function | Given a continuous random variable , the cumulative distribution function is the probability that . |
| Normal distribution | The normal distribution is a type of continuous distribution whose graph looks like this:Bell curve diagram showing a bell shaped line plotted along an x-y axis. The horizontal x-axis is labelled from 100 to 200 in increments of 20 from left to right. The y-axis is labelled from 0 to 50 in increments of 10 from the bottom to the top. Along the x-axis is a series of columns that rise to a peak halfway along the axis and fall back towards 0 on the far right of the x-axis. An orange ‘bell curve’ is laid over the peaks of the columns.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. |
| Probability density function | A function of a continuous random variable, whose integral across an interval gives the probability that the value of the variable lies within the same interval. |
| z-score | A -score is a statistical measurement of how many standard deviations a raw score is above or below the mean. A -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 continuous random variables(1 lesson) | **S3.1: Continuous random variables*** use relative frequencies and histograms obtained from data to estimate probabilities associated with a continuous random variable (ACMMM164)
 | **Building towards a continuous function*** Build the concept that the probability of an event is represented by the area of a rectangle in a histogram, rather than its height. A focus on discrete data can display this idea.
* Students investigate the idea that the area of the rectangles in a frequency histogram can represent the probabilities of events occurring, changing the scale of the *y* – axis to relative frequencies.

**Resource**: investigating-relative-frequencies.DOCX |  |  |
| Introducing probability density functions and cumulative distribution functions (1-2 lessons) | * understand and use the concepts of a probability density function of a continuous random variable **AAM**
	+ know the two properties of a probability density function: for all real and
	+ define the probability as the area under the graph of the probability density function using the notation , where is the probability density function defined on
	+ find the mode from a given probability density function
* obtain and analyse a cumulative distribution function with respect to a given probability density function
	+ understand the meaning of a cumulative distribution function with respect to a given probability density function
	+ use a cumulative distribution function to calculate the median and other percentiles
 | **Introducing probability density functions and cumulative distribution functions*** Explore the calculus relationship between probability density functions (PDF) and associated cumulative distribution functions (CDF).
* Students create frequency histograms from a cumulative frequency graph to establish the connections.

**Resource**: developing-the-relationship-between-PDF-and-CDF.DOCX* Demonstrate the application of integration to calculate the probability of an outcome occurring from within a range of values. In addition, students should be taught to solve problems, such as a given probability with an unknown value in the PDF.

Concepts such as and need to be explicitly explained. Techniques should also be identified to find the mode by identifying the maximum value of a PDF, the median and other percentiles of the distribution. **Resource:** probability-density-function-examples.DOCX, pdf-examples-solution.DOCX |  |  |
| Investigating continuous random variables(2 or 3 lessons) | * understand and use the concepts of a probability density function of a continuous random variable **AAM**
	+ examine simple types of continuous random variables and use them in appropriate contexts
	+ explore properties of a continuous random variable that is uniformly distributed
 | **Investigating uniformly distributed random variables*** Investigate the PDF of a uniformly distributed random variables and develop the formula for this PDF.
* For a uniformly distributed random variable defined on the interval , the PDF will be .

**Resource:** uniformly-distributed-random-variables.DOCX* Practical examples might include:
	+ A spinner where the important feature is the angle between the original position and the landing position (so interval of .)
	+ Raindrops landing on a ruler

**Investigating non-uniformly distributed random variables*** Students can develop the PDF for an example by researching a model (e.g. linear) and then test this model.
* Practical examples might include:
	+ Volume of Coca-Cola in a bottle marked as 600mL
	+ Length of a randomly selected telephone call
	+ Height of a plant grown under specific conditions
 |  |  |
| Introducing the normal distribution(1 or 2 lessons) | **S3.2: The normal distribution*** identify the numerical and graphical properties of data that is normally distributed
* calculate probabilities and quantiles associated with a given normal distribution using technology and otherwise, and use these to solve practical problems (ACMMM170) **AAM**
	+ identify contexts that are suitable for modelling by normal random variables, such as the height of a group of students (ACMMM168)
	+ recognise features of the graph of the probability density function of the normal distribution with mean and standard deviation , and the use of the standard normal distribution (ACMMM169)
	+ visually represent probabilities by shading areas under the normal curve, for example identifying the value above which the top 10% of data lies
 | **Introducing the normal distribution*** Large sets of naturally occurring random phenomena follow a pattern distributed symmetrically about the mean. We refer to this pattern as the normal distribution.
* The normal distribution is the representation of how things occur ‘normally’ in random natural events. Normal distributions are likely to occur in random variables that are free from bias and restrictions such as endpoints i.e. They are allowed to occur naturally.
* Examples of random natural phenomena that are normally distributed:
	+ Heights of people (males or females alone) and some other body measurements / physical or psychological traits.
	+ Weights for animal populations occur quite naturally. However, the weights of human populations do not occur naturally due to economic factors which influence diet. Weight is not random nor normally distributed.
	+ Errors in measurements in terms of manufacturing processes. Weight, volume, quantities of products produced by a machine vary according to a normal distribution.
	+ Lifetimes of products, ie) the lifetime of a lightbulb.
* When investigating a set of data, factors that could prevent it from being a perfect normal distribution should be considered. e.g. for the weight of a box of sultanas labelled 60g might have a mean of 62g when they come out of a machine in a factory, but if a quality control process eliminates all boxes that fall below 60g, this could skew the data that we then take by purchasing sultana boxes.
* Emphasise that sample sizes must be large to allow for data to come close to a normal distribution. Data also needs to be unrestricted to fall naturally into a normal distribution.
* Some worthwhile considerations are given in the clip [The Normal Distribution – An Introduction](https://www.youtube.com/watch?v=PC74JgrdztQ) (duration 4:51) or [Normal Distribution](https://www.mathsisfun.com/data/standard-normal-distribution.html).

Students could perform a large scale collection of data and analyse, calculating the mean, mode, median, standard deviation and investigating quartiles to establish the properties of data sets that are normally distributed. This could include short term surveys around heights of students of a similar age, or male or female adults in the community, weights of a food product, such as boxes of sultanas. To increase feasibility of the survey, longer term surveys that could be performed weeks before beginning the actual topic could include:* + Identifying a food product that a student brings most days and measuring its mass(reduces cost)
	+ A single student recording their daily travel time to arrive at school

**Resource:** dice-dropping-experiment.DOCX This investigation should take students from the introduction to the end of this topic |  |  |
| Understanding and calculating z-scores(1 or 2 lessons) | * understand and calculate the -score (standardised score) corresponding to a particular value in a dataset **AAM Paperclip icon**
	+ use the formula , where is the mean and is the standard deviation
	+ describe the -score as the number of standard deviations a value lies above or below the mean
* use -scores to compare scores from different datasets, for example comparing students’ subject examination scores **AAM**
 | **Considering the need for a z-score*** Teachers explain the importance of -scores as:
	+ raw scores by themselves do not necessarily provide information about the position in a distribution
	+ standardising different distributions allows us to make comparisons between these distributions.
* Explore two or more sets of scores before and after conversion to -scores in order to assist explanation of the advantages of using standardised scores.
* Consider the concept using the [Why do we need z-scores?](https://www.youtube.com/watch?v=fnU42Ue9utk) (duration 4:33)clip.

**Developing the formula of a z-score*** The formula for a *z*-score can be developed by considering the definition of a ­z­-score as being “the number of standard deviations above the mean that a score lies.”
* We can say that if a score lies “z” standard deviations above the mean, we could calculate the score itself by adding “z” standard deviations to the mean, i.e.

and rearranging we get**NESA suggested applications and exemplar questions*** Given the means and standard deviations of each set of test scores, compare student performances in the different tests to establish which is the ‘better’ performance.
* Packets of rice are each labelled as having a mass of 1 kg. The mass of these packets is normally distributed with a mean of 1.02 kg and a standard deviation of 0.01 kg. Complete the following table:

Packets of rice are each labelled as having a mass of 1 kg. The mass of these packets is normally distributed with a mean of 1.02 kg and a standard deviation of 0.01 kg.   Student to complete the table. * + What percentage of packets will have a mass less than 1.02 kg?
	+ What percentage of packets will have a mass between 1.00 and 1.04 kg?
	+ What percentage of packets will have a mass between 1.00 and 1.02 kg?
	+ What percentage of packets will have a mass less than the labelled mass?
 |  |  |
| Verifying the results under a normal curve(1 lesson) | * use collected data to illustrate the empirical rules for normally distributed random variables
	+ apply the empirical rule to a variety of problems
	+ sketch the graphs of and the probability density function for the normal distribution using technology
	+ verify, using the Trapezoidal rule, the results concerning the areas under the normal curve
 | **Verifying the normal distribution empirical rules using the trapezoidal rule*** Consider the need to use the trapezoidal rule to verify the probabilities in a normal distribution, given that the probability density function is too complex to integrate.
* As students use the trapezoidal rule, they will need to be supported tot consider why the percentages will not match perfectly and whether they should be over or underestimates.
* Students use Microsoft Excel and paper methods to investigate the empirical rules. **Resources:** trapezoidal-rule-pdf.XLSX
* Students use Microsoft Excel to input data they have collected in previous lessons and construct a relative frequency histogram. Students can then calculate the mean and standard deviation of their data, and construct the probability density function of a normally distributed random variable on the same axes to compare their results. The following sheet explains how to graph a normal distribution based on its mean and standard deviation, and then how to place a polygon of your data on the same graph to compare.

**Resource:** creating-a-histogram-in-excel.DOCX, creating-a-normal-distribution-in-excel.DOCX * After creating the graphs in Excel, students should discuss a range of questions including:
	+ Why do these graphs not match perfectly?
	+ Is my data normally distributed?
* The concept we are leading the students to is that we use a normally distributed random variable to model a real world situation, but it will rarely be perfectly normally distributed.
 |  |  |
| Using z-scores(1 or 2 lessons) | * use -scores to identify probabilities of events less or more extreme than a given event **AAM Paperclip icon**
	+ use statistical tables to determine probabilities
	+ use technology to determine probabilities
* use -scores to make judgements related to outcomes of a given event or sets of data **AAM**
 | **Using z-scores*** Given a normally distributed variable with a mean and standard deviation, students should be exposed to a variety of questions of the form, “What is the probability that a person chosen at random will have a score greater/less than \_\_\_\_?”
* Ideally, such questions will be developed based around surveys that the students have conducted in previous lessons.
* Students also need the opportunity to draw clear conclusions from data sets and results of individuals compared to normally distributed data sets. They should be encouraged to use z-scores to justify their conclusions.

**NESA suggested applications and exemplar questions*** A machine is set for the production of cylinders of 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 conclusion could be drawn?
* 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:
	+ between 100 and 120
	+ of a least 120
	+ of greater than 120

**Note:** Normal distribution tables would be used to answer this question.* The lifetime of a particular make of lightbulb is normally distributed with mean 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. Normal distribution tables would be used to answer this question.
 |  |  |

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 ICT, literacy, numeracy and group activities should be recorded in the ‘Comments, feedback, additional resources used’ section.