 Year 11 Mathematics Standard

| MS-S1 Data Analysis | Unit duration |
| --- | --- |
| Statistical Analysis involves the collection, exploration, display, analysis and interpretation of data to identify and communicate key information.  Knowledge of statistical analysis enables the careful interpretation of situations and raises awareness of contributing factors when presented with information by third parties, including the possible misrepresentation of information.  Study of statistics is important in developing students’ understanding of the contribution that statistical thinking makes to decision-making in society and in the professional and personal lives of individuals. | 5 weeks |

| Subtopic focus | Outcomes |
| --- | --- |
| The principal focus of this subtopic is planning and management of data collection, classification and representation of data, calculation of summary statistics for single datasets and their use in the interpretation of data.  Students develop awareness of the importance of statistical processes and inquiry in society.  Within this subtopic, schools have the opportunity to identify areas of Stage 5 content which may need to be reviewed to meet the needs of students. | A student:   * represents information in symbolic, graphical and tabular form MS11-2 * develops and carries out simple statistical processes to answer questions posed MS11-7 * uses appropriate technology to investigate, organise and interpret information in a range of contexts MS11-9 * justifies a response to a given problem using appropriate mathematical terminology and/or calculations MS11-10   **Related Life Skills outcomes**: MALS6-2, MALS6-9, MALS6-13, MALS6-14 |

| Prerequisite knowledge | Assessment strategies |
| --- | --- |
| The material in this topic builds on content from the Statistics and Probability Strand of the K–10 Mathematics syllabus, including the Stage 4 content of measures of location and range and the Stage 5 content for comparing sets of data. | Summative assessment: MS-S1 Data Analysis – How far would you go for fashion?  Students investigate what are the attitudes of teenagers towards fashion and how far would they go for fashion? |

All outcomes referred to in this unit come from [Mathematics Standard Stage 6](https://syllabus.nesa.nsw.edu.au/mathematics-standard-stage6/) 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 |
| --- | --- |
| blood alcohol content | Blood alcohol content measures the amount of alcohol present in the bloodstream, and may be used for legal purposes. |
| box-plots | A box-plot is a graphical display of a five-number summary.  In a box-plot, the ‘box’ (a rectangle) represents the interquartile range (IQR) with ‘whiskers’ reaching out from each end of the box towards maximum and minimum values in the dataset. A line in the box is used to indicate the location of the median. Also known as a box-and-whisker plot. |
| categorical variable | A categorical variable is a variable whose values are categories.  Examples include major blood type (A, B, AB or O) or principal construction type (brick, concrete, timber, steel, other).  Categories may have numerical labels, for example postcodes, but these labels have no numerical significance, and they merely serve as labels. |
| continuous | Continuous data is data associated with continuous variables and is a type of numerical data. |
| cumulative frequency graphs | A cumulative frequency graph or 'ogive' is a curve or series of straight lines representing the cumulative frequency for a given dataset. |
| deciles | Deciles divide an ordered dataset into ten equal parts. |
| discrete | Discrete data is data associated with discrete variables and is a type of numerical data. |
| five-number summary | A five-number summary is a method for summarising a dataset using five statistics: the minimum value, the first quartile, the median, the third quartile and the maximum value. |
| interquartile range | The interquartile range is a measure of the spread within a numerical dataset. It is equal to the difference between the upper quartile Q3 and the lower quartile Q1; that is, IQR=Q3−Q1. |
| measures of central tendency | Measures of central tendency are the values about which the set of data values for a particular variable are scattered. They are a measure of the centre or location of the data.  The two most common measures of central tendency are the mean and the median. |
| measures of spread | Measures of spread describe how similar or varied the set of data values are for a particular variable.  Common measures of spread include the range, combinations of quantiles (deciles, quartiles, percentiles), the interquartile range, variance and standard deviation. |
| modality | Modality describes the number of peaks in a set of data.  For example data can be unimodal (having one peak), bimodal (having two peaks) or multimodal (having many peaks). |
| nominal | Nominal data is a type of categorical data that has no natural order in which the categories may be placed, for example eye colour. |
| ordinal | Ordinal data is a type of categorical data where the possible categorical responses have a natural order. For example level of happiness: very unhappy, unhappy, neutral, happy, very happy. |
| outliers | An outlier in a dataset is a data value that appears to be inconsistent with the remainder of that dataset. |
| Pareto charts | A Pareto chart is a type of chart that contains both a bar and a line graph, where individual values are represented in descending order by the bars and the cumulative total is represented by the line graph. |
| percentiles | Percentiles divide an ordered dataset into 100 equal parts. See also quantiles.  More formally, it is a statistical measure indicating the value below which a given percentage of observations in a group of observations lie. For example the 20th percentile is the value below which 20% of the observations may be found. |
| quantiles | Quantiles are a set of values that divide an ordered dataset into equal groups. Examples include quartiles, deciles and percentiles.  Formally in statistics, quantiles are cutpoints dividing the range of a probability distribution into continuous intervals with equal probabilities, or dividing the observations in a sample in the same way. |
| 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. |

| Lesson sequence | Content | Suggested teaching strategies and resources | Date and initial | Comments, feedback, additional resources used |
| --- | --- | --- | --- | --- |
| Use appropriate collection methods for populations and samples (2 lessons) | S1.1 Classifying and representing data (grouped and ungrouped)   * describe and use appropriate data collection methods for a population or samples ◊ * investigate whether a sample obtained from a population may or may not be representative of the population by considering different kinds of sampling methods: systematic sampling, self-selected sampling, capture-recapture, simple random sampling and stratified sampling * investigate the advantages and disadvantages of each type of sampling * describe the potential faults in the design and practicalities of data collection processes, eg surveys, experiments and observational studies, misunderstandings and misrepresentations, including examples from the media | * The use of spreadsheets and/or other software is encouraged in this topic to enable students to produce a variety of graphs and tables, create data displays and to calculate measures of location and spread.   Collecting data   * Teacher to discuss with students reasons for collecting and analysing data e.g. In manufacturing, agriculture, medicine, government * Student activity: Students complete the [Bias in data worksheet](https://topdrawer.aamt.edu.au/Statistics/Good-teaching/Data-collection/Bias-in-data/Biased-data) to look at how to avoid bias when collecting data.   Population vs Sample   * Teacher to discuss with students when it is appropriate to use a sample rather than a census * Student activity: Students visit the [You Gov](https://au.yougov.com/results/) website and look at the types of polls conducted. Students should consider the size of samples, how these samples may have been chosen and the use of small samples to make generalisations about how people will and how sometimes these samples are disposed towards a particular viewpoint   Sampling techniques   * Teacher to discuss and identify each type of sampling technique as well as relevant examples of each (for example; census - The National Census, stratified - national opinion of Malcolm Turnbull of 20-30 year olds, systematic - the number of faulty products per day that come out of a production line) * Teacher to lead a discussion on the positives and negatives of each technique and its desired purpose (for example; create questions such as ‘How many girls in the class like the colour blue?’ Then talk about who you would ask as well as which techniques you could use and why/why not). * Teacher to discuss with students effective questionnaire/survey design including; the use of simple language; unambiguous questions; requirements for privacy; freedom from bias * Student activity: Students perform basic class surveys (pets owned, hair colour, transportation to school) based on the various types of sampling. * Student activity: Students create various surveys online via [Survey Monkey](https://www.surveymonkey.com/), Google Forms or Microsoft Forms. Questioning could be based around common issues in society or within the school. This could be done after discussing the correct ways to ask survey style questions. * Student activity: Students could research and discuss pop culture references of sampling techniques, for example; the book/movie ‘The Hunger Games’ represents stratified sample (must be a girl and a boy). |  |  |
| Classify categorical and numerical data (1 lessons) | * classify data relating to a single random variable ◊ **Paperclip icon** * classify a categorical variable as either ordinal, eg income level (low, medium, high) or nominal, eg place of birth (Australia, overseas) * classify a numerical variable as either discrete, eg the number of rooms in a house, or continuous, eg the temperature in degrees Celsius | Classifying data   * Teacher to use a [KWL Organiser](https://goalbookapp.com/toolkit/strategy/kwl-chart) to review student’s background knowledge of categorical and numerical variables. * Teacher to discuss the different types of variables. [What are variables?](http://www.abs.gov.au/websitedbs/a3121120.nsf/home/statistical+language+-+what+are+variables) * Student activity: Using the [2016 Household Census form](https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2900.02016?OpenDocument), students find examples (about 5 each) of questions that involve nominal, ordinal, continuous and discrete variables. * Student activity: Using [101 questions](http://www.101qs.com/) , students think of as many random variables as they can, for the given stimuli. Generate a class discussion to determine whether the variables are categorical (nominal or ordinal) or numerical (discrete or continuous). A discussion regarding the variable ‘postcode’ is useful as an example of a variable which appears to be numerical but is actually categorical. * Student activity: In groups, students select a variable, and consider the ways its classification can change, depending on context and how the variable is measured. In the activity student groups may be given a different classification of age (ordinal, discrete, categorical) and they could discuss how age can be classified in that way. [An example, with ‘age’ as the random variable](http://www.theanalysisfactor.com/level-of-measurement-not-obvious/) |  |  |
| Display data in table and graphical forms (3 lessons) | * review how to organise and display data into appropriate tabular and/or graphical representations **AAM** ◊ **Paperclip icon**  Information and communication technology capability icon Literacy icon * display categorical data in tables and, as appropriate, in both bar charts or Pareto charts * display numerical data as frequency distribution tables and histograms, cumulative frequency distribution tables and graphs, dot plots and stem and leaf plots (including back-to-back where comparing two datasets) * construct and interpret tables and graphs related to real-world contexts, including: motor vehicle safety including driver behaviour, accident statistics, blood alcohol content over time, running costs of a motor vehicle, costs of purchase and insurance, vehicle depreciation, rainfall, hourly temperature, household and personal water usage Sustainability icon Civics and citizenship icon | Misleading graphs   * Teacher to use the various graphs on the [Misleading Graphs - Real Life Examples](http://www.statisticshowto.com/misleading-graphs/) website to lead a discussion about how graphs can become misleading.   Types of data displays   * Student activity: Students use the [Graph Investigator](http://splash.abc.net.au/res/i/L5904/index.html) to research the different types of graphs and the types of data they are suitable for. Alternatively, they could watch a short video on [Choosing the right graph](https://www.youtube.com/watch?v=Ka5pGmHJENI) * Student activity: Students could look through the many different graphs on the [Data visualisation catalogue](https://datavizcatalogue.com/) page, focussing on the more unusual styles of graphs * Teacher should make a clear distinction between histograms and column graphs or bar graphs, particularly when using spreadsheets to generate graphical representations. * Student activity: Students use [Singapore’s Public Data](https://data.gov.sg/) website to look at the different data displays, name the data display and explain why this is a preferable choice of data display for the data set. * Student activity: Students should use a spreadsheet or [CreateAGraph](https://nces.ed.gov/nceskids/createagraph/Default.aspx) to create a variety of charts from data sourced online ([Australian Bureau of Statistics](http://www.abs.gov.au/)) or collected from and/or by the students. * Student activity: Students use the graphical tool on [Gapminder](http://www.gapminder.org/tools/#_chart-type=bubbles) to see a visual representation of how to graph different types of data. Students should discuss, for example, how data changes over time and how this change is reflected in the graph. * The Gapminder site, whilst included here, is more often used to display bivariate data. However, the site does include some examples of single variable data. The website should be explored before opening in class. * Student activity: Students use current statistics on road safety/crash analysis to construct tables and graphs (for example; over-time analysis of deaths on NSW roads, Pareto charts of causes of accidents, and so on).   Resources:   * [Centre for Road Safety – Interactive Crash Statistics](http://roadsafety.transport.nsw.gov.au/statistics/interactivecrashstats/index.html) (Transport for NSW) * [NSW Road Statistics](http://roadsafety.transport.nsw.gov.au/statistics/index.html) * [Notes on constructing a Pareto Chart](http://asq.org/learn-about-quality/cause-analysis-tools/overview/pareto.html) * Student activity: Students use a spreadsheet or graphing software to plot and create graphs of car maintenance/costs or the amount owing on a loan versus value of the car over time   **Additional Resources**   * [More data sets through “Kaggle”:](https://www.kaggle.com/) * Note: The Kaggle website has an extensive number of datasets, but the site should be explored before opening it in class. The amount of data is overwhelming, however, it can be difficult to locate useful data. It may be worth downloading specific datasets and tailoring it to suit the needs of the class. |  |  |
| Compare data in tables and graphical forms (2 lessons) | * interpret and compare data by considering it in tabular and/or graphical representations **AAM** ◊ **Paperclip icon**  Information and communication technology capability icon Literacy icon * choose appropriate tabular and/or graphical representations to enable comparisons * compare the suitability of different methods of data presentation in real-world contexts, including their visual appeal, eg a heat map to illustrate climate change data or the median house prices across suburbs Sustainability icon Ethical understanding icon Difference and diversity icon | Comparing data   * Student activity: Students investigate the [StatTrek](http://stattrek.com/statistics/charts/compare-data-sets.aspx?Tutorial=AP) website to identify and create a summary of the key aspects of data comparison e.g. centre, spread, shape, outliers * Students compare different data displays for the same data to determine the suitability of data displays and to identify the key features of different data displays. (e.g. climate change, yearly analysis/reviews from bank websites, surveys/analysis from the Australian Bureau of Statistics, road statistics). * Student Activity: Students look at the charts showing the trend in Median property prices. ([House price data](https://www.realestate.com.au/invest/house-in-newcastle,+nsw+2300)) Students can compare these data sets for different locations, using correct metalanguage. * Student activity: Students look at examples of heat maps, such as land surface temperature ([Earth Science Data Visualisations](https://www.jpl.nasa.gov/edu/teach/activity/earth-science-data-visualizations-how-to-read-a-heat-map/)) or [Sydney house values by suburb](http://cdn.newsapi.com.au/image/v1/855f18b4fa3634a4e471a169b5a3dc8a), [BOM temperature change](http://www.bom.gov.au/climate/history/temperature/), [BOM rainfall change](http://www.bom.gov.au/climate/history/rainfall/) and discuss the effectiveness of this type of display for the data. |  |  |
| Calculate population and sample means and standard deviations (1 lesson) | S1.2 Summary Statistics   * describe the distinguishing features of a population and sample ◊ * define notations associated with population values (parameters) and sample-based estimates (statistics), including population mean , population standard deviation , sample mean and sample standard deviation | Mean and standard deviation   * Teacher to revise the difference between a population and a sample * Student activity: Using existing data sets from the [Australian Bureau of Statistics](http://www.abs.gov.au/), [United Nations data](http://data.un.org/), or [Google Public Data Explorer](https://www.google.com/publicdata/directory) students could analyse the differences between samples and populations, using notations to explain findings to other students. * Student activity: Using 2016 [Census Quick Stats](http://www.abs.gov.au/websitedbs/D3310114.nsf/Home/2016%20QuickStats) students can compare population mean and standard deviation (calculated from census data) with the sample mean and standard deviation (calculated from class data of the same topic/issue). For example: Family Composition - Single Parent, Average Children per Family, Gender, and so on. |  |  |
| Calculates measures of central tendency and measures of spread (3 lessons) | * summarise and interpret grouped and ungrouped data through appropriate graphs and summary statistics **AAM** ◊ **Paperclip icon** * discuss the mode and determine where possible * calculate measures of central tendency, including the arithmetic mean and the median (ACMEM050) * investigate the suitability of measures of central tendency in real-world contexts and use them to compare datasets Critical and creative thinking icon Civics and citizenship icon * calculate measures of spread including the range, quantiles (including quartiles, deciles and percentiles), interquartile range (IQR) and standard deviation (calculations for standard deviation only required using technology)  Information and communication technology capability icon | Grouped data   * Teacher to lead a discussion on the inappropriateness of ungrouped data for some variables. * Discussion should consider how data can be grouped, in regards to continuous data, and the appropriateness of graphs to represent the data. * Discussion should take place regarding the uniformity of the size of groups, and whether the first group should include 0. * Discussion should also consider group boundaries for continuous data, for example, if groups are 30 - 34, then 35 - 39, where would the value of 34.5 belong?   Measure of spread   * Quartiles can be determined for data sets containing odd and even numbers of data values. In calculating the first and third quartiles, the median scores are excluded. Students should be aware that the second quartile is the median. * Student activity: Students conduct surveys to collect grouped and ungrouped data in teams. They should carefully choose topics of interest such as “Number of hours spent on social media by teenagers (grouped in hours)” or “Number of hours spent on social media (grouped in age groups)”. * Students should discuss the pros and cons of the measures of central tendency and the appropriateness of each, for each of their questions * Students should use technology to create graphs and to calculate measures of central tendency and spread. Most spreadsheet programs contain similar functions for summary statistics: sum, minimum, maximum, mean, mode, median, quartile, standard deviation. * Other areas of interest to investigate: advertising (particularly for mode), sporting events (crowds, hospitality, transport, and so on), suitability of central tendency measure (mean vs median for house prices or national incomes) * Teacher to lead a discussion regarding the benefits of measures of spread. Students need to understand what a small or large standard deviation might indicate. For example, factory production would require small standard deviations to ensure uniformity in the production line. In contrast, a dog shelter may collect data on the weight of all dogs in the shelter. A high standard deviation may indicate a large range in the size of the dogs, justifying the need to stock a variety of care products.   **Exemplar Question**  A data set of nine scores has a median of 7. The scores 6, 6, 12 and 17 are added to the data set. What is the median of the data set now?  Resource: ms-s1-nesa-exemplar-question-solutions.DOCX |  |  |
| Determine outliers and their effect on the mean and median (2 lessons) | * investigate and describe the effect of outliers on summary statistics ◊ **Paperclip icon** * use different approaches for identifying outliers, including consideration of the distance from the mean or median, or the use of and as criteria, recognising and justifying when each approach is appropriate * investigate and recognise the effect of outliers on the mean and median | Introducing outliers   * Students should develop an understanding of the impact of outliers on data and how the measures of mean and median are impacted. Students, as mathematicians, should not only determine if data is an outlier, through calculation, but should analyse the relevance and cause of outliers. * Student activity: As students enter the classroom, have them put a post-it note on a dot plot scale indicating their shoe size. Use this as a discussion to start the lesson about outliers. If there are no outliers present, then add one and demonstrate different approaches for identifying outliers. * Student activity: Students investigate salt levels in various foods. [Beware of outliers (student worksheet)](https://topdrawer.aamt.edu.au/Statistics/Downloads/Beware-of-outliers-Student-worksheet) * Student activity: Students research a data set of interest to them, identifying outliers and calculating statistics; e.g. local prices of recently sold houses or prices of a car they are interested in buying. Teacher to then lead a discussion of how various factors can impact results. Websites such as [Gumtree](http://gumtree.com.au) and [CarSales](http://carsales.com.au) allow users to specify a price range which can therefore influence results. * Student activity: Students visit a website for a local sporting association and use data from the last season to determine if an outlier exists. For example, students could visit the [Newcastle District Cricket Association](http://mycricket.cricket.com.au/common/pages/asphost.aspx?id=HBA&entityid=2975) website and visit the Hall of Fame section. Students can download various types of data. They can try to predict if a dataset has an outlier before then using the formula to verify their decision. * Alternatively, they could complete the [Outstanding cricketer worksheet](https://topdrawer.aamt.edu.au/Statistics/Downloads/An-outstanding-cricketer-Student-worksheet) |  |  |
| Investigate misuse of statistics (2 lessons) | * investigate real-world examples from the media illustrating appropriate and inappropriate uses or misuses of measures of central tendency and spread (ACMEM056) **AAM** **Paperclip icon** | Misuse of statistics   * Teacher leads a discussion of how [Fox News](https://www.businessinsider.com.au/fox-news-charts-tricks-data-2012-11?r=US&IR=T) creates misleading graphs to deceive its viewers. * Teacher uses the [Examples of misleading statistics](https://www.quora.com/What-are-good-examples-of-misleading-statistics) and [Misleading statistics and data](https://www.datapine.com/blog/misleading-statistics-and-data/) websites to look at how poor sampling and/or bias creates misleading statistics. * The “Examples of misleading statistics2” website outlines an activity involving chocolates which would be useful as an introduction to this topic. * Student activity: Groups are assigned different perspectives of a current issue (Climate change, autism and vaccines, average wages) to research. They need to present their case using statistics to support their arguments. * Mock debate: students can the debate against another group using their argument and ‘false’ statistics. * A discussion should be held at the end of the debate, explaining how they manipulated the statistics to suit their purpose. |  |  |
| Describe datasets according to their spread and measures of central tendency (1 lesson) | * describe, compare and interpret the distributions of graphical displays and/or numerical datasets and report findings in a systematic and concise manner **AAM** ◊ **Paperclip icon** Critical and creative thinking icon  Information and communication technology capability icon Literacy icon * identify modality (unimodal, bimodal or multimodal) * identify shape (symmetric or positively or negatively skewed) * identify central tendency, spread and outliers, using and justifying appropriate criteria * calculate measures of central tendency or measures of spread where appropriate | Describing data sets   * Student activity: Students complete a very simple quiz, such as the addition of two single digit numbers, using Google Forms, where the results will be graphed automatically. Then students complete a harder quiz. The teacher then compares the graphs, discussing similarities and differences (the simpler one should be negatively-skewed and the harder one should be more symmetrical or positively-skewed). * Student activity: Students can choose a social issue or product and using [Google Trends](https://trends.google.com/trends/?geo=AU), can discuss trends in the searches performed for that item. * Students should consider possible reasons for dips and peaks * They could compare Australian data with data from another country. Do they follow a similar pattern? Do they experience peaks and troughs at roughly the same time? |  |  |
| Use parallel box plots to compare data (1 lesson) | * construct and compare parallel box-plots **AAM** **Paperclip icon**  Information and communication technology capability icon * complete a five-number summary for different datasets (ACMEM058) * compare groups in terms of central tendency (median), spread (IQR and range) and outliers (using appropriate criteria) * interpret and communicate the differences observed between parallel box-plots in the context of the data | Using box plots to compare data   * Teacher to revise features of a box plot * Student activity: Students complete this online Scootle activity [Use box plots to compare data sets](http://www.scootle.edu.au/ec/viewing/R12073/index.html). It uses the context of sampling and quality control in manufacturing. Back-to-back (parallel) box and whisker plots are constructed and used to compare datasets. In both contexts, the significance of quartiles and medians is highlighted * Student activity: Students match cumulative frequency graphs to the matching box plot [Box plot match](https://nrich.maths.org/11002) * Student activity: Students can use [Statista](https://www.statista.com/markets/424/internet/) to research how many people in different age groups use Facebook compared to Instagram (or similar). They can then draw parallel box plots to compare the data.   **Additional Resources**   * The following resources are tools to support drawing boxplots * [5 Number Summary Calculator](https://www.easycalculation.com/statistics/five-number-summary.php) * [Statistics Calculator: Box Plot](http://www.alcula.com/calculators/statistics/box-plot/) * [Create Box and Whisker Chart](https://www.meta-chart.com/box-and-whisker)   **Exemplar Questions**   1. Using the box-plot, what percentage of drivers in this sample have reaction times of three or more seconds? What percentage of drivers in this sample have reaction times between four and nine seconds? What is the interquartile range for this data set?   Reaction time in seconds prior to braking –  drivers over 55  1  0  10  9  2  3  4  5  6  7  8   * 1. The box-plots show the distribution of the ages of children in Numbertown in 2002 and 2012.   There are two box-plots drawn above each other, above a number line indicating age in years. The number line begins at 0 and ends at 18.  The higher box-plot is labelled 2012. The whiskers begin at 0 and end at 10. The box begins at 2 and ends at 12 with a dividing line at 6.  The lower box-plot is labelled 2002. The whiskers begin at 0 and end at 10. The box begins at 8 and ends at 16 with a dividing line at 12.   * 1. The number of children aged 12–18 years was the same in both 2002 and 2012. By considering the data, provide advice to town planners about recreational facilities that should be offered, giving statistical reasons.   2. Describe and explain any differences expected between the two data sets containing the finishing times in the Olympic Marathon (elite athletes) and in the Sydney Marathon (open entry) in a particular year.   Resource: ms-s1-nesa-exemplar-questio-solutions.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 ICT, literacy, numeracy and group activities should be recorded in the ‘Comments, feedback, additional resources used’ section.