UNIT 4 Numerical Statistics

Recommended Prior Knowledge. Students will have encountered much of the work in this Unit in previous courses, and will have drawn and used many of the diagrams included in this unit**.**

Context. This unit is independent of Units 1 to 3 (Pure units from P1), but should be covered before Unit 5 (Probability). For schools where two teachers are teaching the AS course, there is no problem in running Units 4 and 5 alongside Units 1 to 3.

Outline. This Unit looks at the topic of Statistics, more or less from first principles, though many students will have encountered many of theses ideas before. It introduces statistics as the collection, representation and analysis of data and examines different aspects of each. It concludes with a study of permutations and combinations; a good understanding of these is necessary before work on Probability (Unit 5) is introduced.

Topic	Learning Outcomes	Suggested Teaching Activities	Resources	On-Line Resources
1	**Representation of Data** • Select a suitable way of presenting raw statistical data, and discuss advantages and / or disadvantages that particular representations may have.	General introduction of the subject of statistics as:- • the collection of data • the representation of data • calculations with data • analysis of data. Look at the different forms of raw data – numeric , nominal etc. Spend some time differentiating between discrete and continuous data. Students will already have encountered different means of representing data in O'Level mathematics, in sciences and in social sciences as well as in real life – in newspapers etc. Ask students to suggest different ways of representing data and obtain a list of different types – pictographs, pie charts, frequency polygons, line graphs, histograms, cumulative frequency diagrams, box and whisper plots, stem and leaf diagrams etc. Have all of these clearly drawn on OHPs. For each type of diagram, discuss with students whether the diagram is for discrete or continuous data and discuss	OHP to list the four components of the study of statistics. OHPs showing accurate diagrams of all the different sorts of statistical diagrams.	

		generally the advantages and / or disadvantages of each.		
	• Construct and interpret stem-and-leaf diagrams, box-and-whisper plots, histograms and cumulative frequency graphs.	Look at stem and leaf diagrams and box and whisper plots in detail. In particular, stress the need for the inclusion of a scale explaining what the numbers on the axis of each type of diagram represents. It is worth giving the students several questions on both types of diagram, ensuring, certainly at first, that the total frequency is, say something like 39, (i.e. of the form $4N$-1, since this makes the median and the quartiles easy whole numbers). Work on both diagrams should be covered at the same time as "measures of central tendency" since students need to appreciate how the median and quartiles can be calculated from the diagrams. Students should be encouraged to comment on the deductions to be made from different box-and-whisper diagrams and these could be available on OHPs. It is a worthwhile exercise asking students to sketch frequency diagrams and other types of diagram for given box-and-whisper plots.	OHPs showing different types of box-and-whisper plots (+ve skew, −ve skew etc).	
	• Understand and use different measures of central tendency (mean, median, mode) and variation (range, interquartile range, standard deviation), e.g. in comparing and contrasting sets of data.	General discussion on measures of "central tendency" (mean, median and mode), including the uses of each one. Look at the calculation of each statistic when the data is discrete and then when it is continuous. This may be the best time to look at different class intervals for continuous data. Students need to be aware that the interval $10 \leq L < 20$ means that L can take any value from 10.0 up to 19.9 recurring and that the mid-point of this interval is 15 and the upper limit is 20. They should also realise that "10- , 20- etc" is another way of expressing $10 \leq L < 20$, $20 \leq L < 30$ etc. Similarly, if continuous data has been expressed to the nearest integer, then the interval 10–14 is equivalent to $9.5 \leq L < 14.5$, with a mid-point of 12 and an upper limit of 14.5. Confusion over the use of such intervals is widespread and it is worth spending time specifically on such intervals at an early stage.		

		Consider two sets such as {1,2,3,4,5,5,5,6,7,8,9} and {3,3,4,4,5,5,5,6,6,7,7} which have the same mean but a different variance. Discussion of the differences between the two sets of data is a good starting point to introduce the idea of variation. Introduce the terms range, interquartile range, standard deviation and variance. Similarly, students should be able to deduce how, in everyday terms, performances of one person compares with those of another, given the means and standard deviations of the performances of the two people.		
	• Use a cumulative frequency graph to estimate the median value, the quartiles and the interquartile range of a set of data.	Many students will have met the idea of cumulative frequency. Look at its use with examination marks, in decisions on pass marks and grade boundaries. (OHP showing a set of results would be useful). Students should not only be aware of the need to plot at the upper class boundary (this is still the most common error), but also understand that this is because an ordinate at *x* represents the number of readings "up to *x*". Give the students plenty of practice drawing graphs and using them to calculate the median, upper and lower quartiles, interquartile range etc.	OHP with a cumulative frequency curve showing examination results.	
	• Calculate the mean and standard deviation of a set of data (including grouped data) either from the data itself or from given totals such as Σx and Σx^2, or $\Sigma(x-a)$ and $\Sigma(x-a)^2$.	General discussion of the methods for calculating the mean and standard deviation of a set of results (discrete or continuous data). The exact meaning of standard deviation is an extremely difficult one for students to fully appreciate. However it is important that the students can calculate the standard deviation (or the variance) from a set of raw data or from given totals such as Σx and Σx^2, or $\Sigma(x-a)$ and $\Sigma(x-a)^2$. Students should also be able to perform these calculations on a calculator, but should be aware that showing full working is still a requirement of questions on this topic.		

2	**Permutations and Combinations** • Understand the terms permutation and combination, and solve simple problems involving selections.	This is a difficult topic for students to fully understand. If time allows, students should be encouraged to find results for themselves. Certainly students should be asked to deduce for themselves the number of different permutations of, say, *ABC,* of ABCD, of ABCDE etc and to deduce the general result of "*N*!", where *N* is the number of letters.. The "*r*-*s*" principle –(that is if there are *r* ways of selecting one option and *s* ways of achieving a second independent option then there are *rs* ways of achieving both options) is a useful introduction to the idea of "5×4×3×2×1". A similar method can then be used to find the number of permutations of, say 5 objects, if only 3 are to be selected. Students should realise that the answer is "5×4×3" and that this is the same as $\frac{5!}{(5-2)!}$, leading to the general result for ${}_nP_r$ $(=\frac{n!}{(n-r)!})$. Following practice in using this formula, introduction of the term "selection" or "combination" should follow naturally. Students should be aware that if a selection of *r* objects is being made from a total of *n* objects, then the number of such selections will be smaller than the corresponding number of permutations (unless r = 1). The number of selections is obtained by dividing the corresponding number of permutations by *r*!, leading to the formula for *n*C*r* of $\frac{n!}{r!(n-r)!}$.		

	• Solve problems about arrangements of objects in a line, including those involving repetition (e.g. the number of ways of arranging the letters of the word "NEEDLESS"), restriction (e.g. the number of ways several people can stand in a line if 2 particular people must – or must not – stand next to each other).	General discussion can then follow on the number of permutations of a collection of items when some of the items are identical. A good introduction to this is to look at the number of permutations of the letters of a word, such as CANADA. Students should realise that the total number of ways is $\frac{6!}{3!}$. A series of restrictions can now be applied to the problem e.g. the number of selections • that start with C, • that start with A, • that start and finish with A, • in which the letter A is never adjacent to an A, • in which all three A's come together, • that do not begin with A .		