THE ACADEMY OF ST NICHOLAS	Y12 Applied	CH03 2.1, 2.2, 2.3, 2.4	Statistics – Representing Data	Lessons 5

Essential Knowledge Milestones

- Understand how to identify outliers in data sets and how to 'clean' the data set
- Understand how to draw and interpret box plots
- Understand how to draw and interpret cumulative frequency diagrams
- Understand how to draw and interpret histograms knowing that the height of a bar can be represented as:

  area of bar = k x frequency
- Understand how to compare two data sets

Success Criteria

- ❑ Be able to follow a rule to decide which points are considered to be outliers and then decide how to deal with them
- ❑ Be able to draw a box plot, including the labeling of outliers. Understand what measures to compare between two box plots
- ❑ Be able to draw cumulative frequency diagrams and extract quartiles and percentiles from it
- ❑ Be able to draw a histogram, split class widths appropriately and find the height and width of missing bars
- ❑ Be able to compare two data sets by making a comment on a measure of location and a measure of spread

Assumed Prior Knowledge/ Links / Interleaving

GCSE (9-1) in Mathematics at Higher Tier

- S2 Interpret and construct tables, charts and diagrams, including frequency tables, bar charts, pie charts and pictograms for categorical data, vertical line charts for ungrouped discrete numerical data and know their appropriate use
- S3 Construct and interpret diagrams for grouped discrete data and continuous data, i.e. histograms with equal and unequal class intervals and cumulative frequency graphs, and know their appropriate use

Teaching Points

- Students should be familiar with and be able to interpret histograms, frequency polygons, box and whisker plots and cumulative frequency diagrams. These should have been covered at GCSE but it is worth a recap for consistency of methods. Also cover calculating summary statistics from diagrams, including the mean and standard deviation from a histogram.
- Students will often say 'The area of the bar is the frequency'. They need to be dissuaded of that notion early on. Clarify that what they will have seen at GCSE was indeed area of bar = k x frequency but k = 1. It is important to explore this deeply. This is particularly important when students are faced with a histogram and are told how many people it represents. Finding the k value has to be the first move.
- Students will need explicit instruction on 'splitting groups'. Ie. If there are 16 people in the 30-40 group how many are in the 30-35 group. This raided a number of points in terms of assumptions but some proportional reasoning too. Students will find it easier when faced with a histograms as they will say 'well that just splits it in half' but will find that question more difficult form a tabular format.
- Outliers will need to be identified and interpreted from data sets and statistical diagrams. Any rules to be used will be given in the question, for example Q1 – 1.5 × IQR, Q3 + 1.5 × IQR. Make students aware that these are just rules and can vary from question to question, they are not to be learned as a common definition.
- When calculating outliers all working out must be shown to gain all marks. Despite it looking easy, students must define the boundaries of outliers and use this in justifications as to whether a point is an outlier or not.
- Ensure you complete Ex3AQ4. It's a good example to test understanding of summary statistics. If students didn't do the standard deviation investigation in the last chapter, their manipulation of Σx^2 may be interesting!
- When delivering box plots ensure that students understand the key new information regarding how to display outliers and then what that then means to the 'whisker' there are two options that students need to consider!
- Students will be expected to select an appropriate diagram or critique the choice of one which is used. They should also be able to clean data by identifying possible outliers (box plots and scatter diagrams). They may also be asked to fill in missing data using a regression line.
- When comparing data sets students should be encouraged to choose a measure of location AND a measure of spread to discuss and then do so in context. Students must choose complimentary measures ie median and IQR or mean and sd. They will find the contextual references the most difficult.

Potential Barriers to Access /Misconceptions	Opportunities for Reasoning/Problem Solving/Proofs
• Knowing how to interpret statistics students have calculated is sometimes found challenging, and often discriminates between students in exam questions. Full and clear reasons for interpretations and decisions need to be given for marks to be awarded. • Students will often say 'The area of the bar is the frequency'. They need to be dissuaded of that notion early on. Clarify that what they will have seen at GCSE was indeed *area of bar = k x frequency* but k = 1. It is important to explore this deeply. • Many students have difficulties calculating the sizes of bars in histograms, as commented on by one examiner: 'Most were able to state the correct width of the bar but few used frequency densities correctly to find the height, some finding the frequency density of but then calculating $\frac{1}{3} \times 2.5$ rather than $2.5 \div \frac{1}{3}$. Some identified that 1.5 cm^2 represented 10 customers but were then unable to use this correctly to find the height … some students had an incorrect class width because they did not realize that the lower class boundary was 70 not 69.5.'	• There is opportunity for further use of the large data set here. Summary statistics of elements from the data set can be calculated and then used to compare and interpret for both location and variation statistics. This will further expand understanding of the LDS whilst building on their comparison skills. • The reasoning aspect here is huge. All conclusions when comparing data sets must have a reason.

Key Mathematical Vocabulary	Histogram, box plot, probability density function, cumulative distribution function, continuous random variable, outlier, skewness, symmetrical, positive skew, negative skew.

Personal Development	Notes	Resources
Pupils are taught that they must 'respect' each other's opinions and well-being when working collectively in class. Pupils to learn that mathematicians have 'ambition' to push boundaries when aiming to solve new problems **Resilience** – never giving up! Building confidence across the problem solving aspects of the course. **Ambition** – living life to the full – fulfilling dreams and aspirations – linking to future career and ambition plans. **Respect** – respect for others – the 9 protected characteristics **Personal Best** – First Work – Best Work every time		• I have drawn together a number of resources from over the years to create a pack that is hopefully informative, a good range of questions and able to be drawn/written on for speed.