|  |
| --- |
| **UNIT 14: Statistics and sampling, cumulative frequency and histograms** |

[Return to Overview](#HOverview)

**SPECIFICATION REFERENCES**

S1 infer properties of populations or distributions from a sample, whilst knowing the limitations of sampling apply statistics to describe a population

S3 **interpret and construct 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**

S4 interpret, analyse and compare the distributions of data sets from univariate empirical distributions through:

* Appropriate graphical representation involving discrete, continuous and grouped data, **including box plots**
* appropriate measures of central tendency (median, mean, mode and modal class) and spread (range, including consideration of outliers, **quartiles and inter-quartile range)**

S5 apply statistics to describe a population

**PRIOR KNOWLEDGE**

Students should understand the different types of data: discrete/continuous.

Students should have experience of inequality notation.

Students should be able to multiply a fraction by a number.

Students should understand the data handling cycle.

**KEYWORDS**

Sample, population, fraction, decimal, percentage, bias, stratified sample, random, cumulative frequency, box plot, histogram, frequency density, frequency, mean, median, mode, range, lower quartile, upper quartile, interquartile range, spread, comparison, outlier

|  |  |
| --- | --- |
| **14b. Cumulative frequency, box plots and histograms**(S3, S4, S5) | **Teaching time**5-7 hours |

**OBJECTIVES**

By the end of the sub-unit, students should be able to:

* Use statistics found in all graphs/charts in this unit to describe a population;
* Know the appropriate uses of cumulative frequency diagrams;
* Construct and interpret cumulative frequency tables, cumulative frequency graphs/diagrams and from the graph:
* estimate frequency greater/less than a given value;
* find the median and quartile values and interquartile range;
* Compare the mean and range of two distributions, or median and interquartile range, as appropriate;
* Interpret box plots to find median, quartiles, range and interquartile range and draw conclusions;
* Produce box plots from raw data and when given quartiles, median and identify any outliers;
* Know the appropriate uses of histograms;
* Construct and interpret histograms from class intervals with unequal width;
* Use and understand frequency density;
* From histograms:
* complete a grouped frequency table;
* understand and define frequency density;
* Estimate the mean and median from a histogram with unequal class widths or any other information from a histogram, such as the number of people in a given interval.

**POSSIBLE SUCCESS CRITERIA**

Construct cumulative frequency graphs, box plots and histograms from frequency tables.

Compare two data sets and justify their comparisons based on measures extracted from their diagrams where appropriate in terms of the context of the data.

**OPPORTUNITIES FOR REASONING/PROBLEM SOLVING**

Interpret two or more data sets from box plots and relate the key measures in the context of the data.

Given the size of a sample and its box plot calculate the proportion above/below a specified value.

**COMMON MISCONCEPTIONS**

Labelling axes incorrectly in terms of the scales, and also using ‘Frequency’ instead of ‘Frequency Density’ or ‘Cumulative Frequency’.

Students often confuse the methods involved with cumulative frequency, estimating the mean and histograms when dealing with data tables.

**NOTES**

Ensure that axes are clearly labelled.

As a way to introduce measures of spread, it may be useful to find mode, median, range and interquartile range from stem and leaf diagrams (including back-to-back) to compare two data sets.

As an extension, use the formula for identifying an outlier, (i.e. if data point is below
LQ – 1.5 × IQR or above UQ + 1.5 × IQR, it is an outlier). Get them to identify outliers in the data, and give bounds for data.