

Higher Tier GCSE Statistics

1.1 Introduction to Types of Data

- a Primary and Secondary Data
- b Qualitative and Quantitative Variables
- c Discrete and Continuous Data
- d Independent (explanatory) and Dependent (response) variables.
- e Classifying data, defining class limits and intervals
- f Consideration of implications of grouping for loss of accuracy in presentation and calculation.
- g Bivariate data

1.2 Planning Investigations and Obtaining Data

- a Specifying research questions and hypotheses
- b Determining the data required and appropriate collection methods. Justifying choice of methods.
- c Obtaining data by counting or measuring. Consider accuracy. Using appropriate and efficient methods of recording data.
- d Questionnaires, Pilot studies, pre-testing and opinion scales.
- e Problems of design. Wording, biased questions, definitions, obtaining truthful answers, random response techniques. Closed and open questions
- f Awareness of problems arising through: identifying the population, questionnaire distribution and collections, non-response, errors in recording answers, missing data.
- g Interview data. Advantage and disadvantages over questionnaires
- h Data logging and simulations. Use of dice, random numbers and ICT.
- i Using secondary data, sources, reliability, accuracy, relevance and bias. Differences between census and sample data.
- j Designing and obtaining data from simple statistical experiments. Identifying variables, matching pairs of groups, "before and after" experiments. Use of control groups. Identifying extraneous variables and control methods. Using surveys. The capture/recapture method for obtaining data.

2.1 Sampling and Census Data

- a Census data from a small-well defined population
- b The implications of obtaining census data from a large population. The National Census.
- c The purpose of sampling, and the benefits of sampling compared to a census.
- d Randomness. Generating and using random numbers
- e Sampling methods from a well-defined population. Be familiar with each of the following methods and situations in which they are used. Simple random sampling, Systematic Sampling, Stratified sampling with up to two sets of categories, Cluster and quota sampling
- f Compare sampling methods in terms of their strengths and weaknesses. Also consider the criteria used for selecting sample members and associated bias.

2.2 Tabulation

- a Constructing frequency tables by tallying raw data
- b Use of class intervals and open-ended classes
- c Simplifying tables by combining categories and reducing the number of significant figures. The effects of over and under simplification resulting from choice of group size, and the effects on readability, masking of patterns and trends.
- d Reading and interpreting data presented in tabular or graphical form

- e Design of tables to summarise data effectively. Including listing all outcomes for single events and two successive events in a systematic way. Design and use of appropriate two-way tables.

3.1 Diagrammatic Representation

- a Qualitative data: bar and pie charts, pictograms, multiple and composite bar charts, comparative pie charts (area proportional to frequency)
- b Discrete data: vertical line graphs, cumulative frequency step polygons
- c Continuous data: grouped frequency diagrams with equal class intervals. Histograms with equal or unequal class intervals. Frequency Polygons, cumulative frequency polygons, and population pyramids.
- d Stem-and-leaf diagrams, ordered, unordered, comparative
- e Shading maps (choropleth maps)
- f Transforming data presentation from one form to another.
- g The shapes and simple properties of frequency distributions and the normal frequency distribution
- h Bivariate data and scatter diagrams
- I Time series data and line graphs
- j Other diagrammatic representations for comparisons of data using length. Use of area and volume, and comparisons of various methods
- k Visual misrepresentation, misuse or omission of origin or scale. Broken, incorrect or changed scales. Incomplete definitions and labelling. Misuse or length, area and volume.
- l Interpretation of information in diagrams. Distinctions between well and poorly presented data. Spotting outliers that do not fit a general pattern.

4.1 Measures of Location

- a Calculate the mean, median and mode for raw data. Use of a change of origin when calculating the mean, e.g. the mean of 103 and 105 is 100 plus the mean of 3 and 5. The effect on the average of changes to the sample, e.g. addition or withdrawal of a member.
- b Mean, median and mode for discrete frequency distributions. Modal class, median and mean for grouped frequency distributions
- c Advantages and disadvantages of each of the three measures in a given situation. Choice of an appropriate measure of location
- d The Geometric Mean

4.2 Measures of Spread

- a The range.
- b Quartiles for discrete data. Quartile, deciles and percentiles, for grouped frequency distributions.
- c Interquartile range for discrete and continuous data. Interpercentile ranges.
- d Variance and standard deviation, including grouped frequency distributions. Awareness of calculator methods. Use sigma notation
- e Advantages and disadvantages of these measures of spread.
- f Construction of box and whisker plots. Use of plots to identify outliers. Outliers defined as less than $1.5 \times \text{IQR}$ below the LQ or more than $1.5 \times \text{IQR}$ above UQ.
- g Calculation and interpretation of standardised scores.
- h Use of tabulated data, diagrams, measures of location and measures of spread to compare data sets. Use of standardised scores to compare values from different frequency distributions.

4.3 Other Summary Statistics, Time Series and Quality Assurance.

- a Simple index numbers. Weighted index numbers. Chain base numbers. General Index of Retail Prices (RPI).
- b Crude and Standardised rates, e.g. birth, death and unemployment.
- c Drawing a trend line by eye and using it for prediction. Evaluating and plotting appropriately chosen moving averages.
- d Identification of seasonal variation. Trend line based on moving averages. Seasonal effect at a given data point. Average seasonal effect. Prediction of future values. All by graphical methods only.
- e Plotting sample means, medians or ranges over time to view consistency and accuracy against a target value.

5.1 Correlation, Regression and Estimation.

- a Scatter diagrams. Recognition by eye of correlation.
- b The distinction between correlation and causality.
- c Spearman's rank correlation coefficient; its calculation and limitation in interpretation. Include tied ranks. Formula given.
- d Fitting a line of best fit through (\bar{x} mean, \bar{y} mean). Obtain the equation of the line of best fit in the form $y = mx + c$. Interpreting m and c . Non-linear data. Appreciation of whether line is appropriate.
- e Interpolation and Extrapolation. Including dangers of inappropriate extrapolation.
- f Interpretation of bivariate data presented in the form of a scatter diagram. Comparison of the degree of correlation between two or more pairs of data sets with reference to scatter diagrams and rank correlation coefficients.
- g Estimation of population mean from a sample. Estimation of a population proportion from a sample; particularly in opinion polls. Variability in estimates from different samples and the effect of sample size. Estimation of population size bases on the capture/recapture method. Quantitative appreciation of appropriate sample size. Including the concept that to halve the variability in an estimate, four times the sample size is required.

5.2 Probability

- a Probability of an event, impossible events, certain events. Use of words such as possible, likely. Putting events into order of probability. Probability of a scale from 0 to 1.
- b Probability as the limit of relative frequency as the number of observations increases. Equally likely events.
- c Sample space: pictorial representation; probability by counting. Use of Venn diagrams, tables and Cartesian grids. Listing all possible outcomes.
- d Exhaustive and Mutually exclusive events. The general addition law. The sum of the probabilities of all mutually exclusive outcomes.
- e Independent events and the general multiplication law.
- f Tree diagrams up to three stages, with and without replacement.
- g An intuitive approach to conditional probability, in simple cases.
- h Expected frequencies. Comparison of actual frequencies with expected frequencies.
- I Simple cases of the binomial and discrete uniform distribution. Including the use of simulation methods.