

Subject: Statistics (Higher)

Department Staff – Mr E DeSouza, Mr E Gakpo, Mr A Hosseinian,
Mr P Kelly, Mr S McCann, Mr K Pillay,
Mr A Sztranyovsky, Ms D Thompson

Year 10 - Academic Year 2010-2011

Contents: what you will study

Data Collection

Types of data

Primary and secondary data sources. Qualitative and quantitative variables. Discrete and continuous data.

Classification of data; the need for precise definitions; class limits and intervals. Implications of grouping for loss of accuracy in presentation and calculation.

Bivariate data: discrete, continuous; ungrouped and grouped.

Census data

Obtaining information from a well-defined small population.

Obtaining information from a large population.

Sampling

Purpose of sampling; variability between samples.

Randomness. Random numbers from tables, calculators and computers.

Sampling from a well-defined population. Sample frame. Simple random sampling; the condition that all members of the population are equally likely to be included in the sample. Stratified sampling with one set of categories.

Systematic sampling. Stratified sampling with no more than two sets of categories.

Cluster sampling and quota sampling with particular reference to their use in conducting large scale opinion polls. Strengths and weaknesses of the various sampling methods, including the dangers of convenience sampling. The criteria used for selecting sample members in national opinion polls: geographical area, sex, age group, social and economic backgrounds. Associated sources of bias.

Biased samples arising from sampling from a wrong population or non-random choice of individual elements.

Planning and Investigation

Specifying a research question to be investigated and breaking it down into sub-questions as necessary. Specifying a hypothesis to be tested.

Determining the data needed in an investigation and selecting an appropriate method for obtaining the data.

Justifying the choice of method by comparing it with possible alternatives.

Obtaining Data

Obtaining data by counting or measuring; accuracy of such measures. Design and use of efficient methods of recording data, appropriate to the purpose for which it will be used.

Obtaining primary data by questionnaire. Pilot studies and pre-testing. Problems of design, wording, biased questions, definitions, obtaining truthful answers. The advantages and disadvantages of closed and open questions.

The use of opinion scales. The technique of random response, in its simplest form, for obtaining truthful answers to sensitive questions.

Awareness of the problems that may arise through:

- identifying the population;
- questionnaire distribution and collection;
- non-response;
- errors in recording answers;
- missing data.

Subject: Statistics (Higher)

Department Staff – Mr E DeSouza, Mr E Gakpo, Mr A Hosseinian,
Mr P Kelly, Mr S McCann, Mr K Pillay,
Mr A Sztranyovsky, Ms D Thompson

Obtaining data by interview. Advantages and disadvantages of interviews compared with written questionnaires.

Obtaining data by data logging.

Simulation. Use of, for example, dice, random number tables, ICT.

Using secondary data; sources, reliability, accuracy, relevance and bias. Difference between sample and census data.

Designing and obtaining data from simple statistical experiments.

Explanatory and response variables; identification of the variables to be investigated.

Use of a control group; use of random allocation to experimental and control groups.

Matched pairs of groups; “before and after” experiments. Identification of extraneous variables and methods of controlling them: the need to hold extraneous variables constant for both groups.

Surveys.

The capture/recapture method for obtaining data. Conditions for this method need to be appropriate.

Tabulation and Representation

Tabulation

Construction of frequency tables by tallying raw data.

Class intervals. Open-ended classes.

Simplifying tables by combining categories and reducing the number of significant figures; resulting effects on readability: identifying or masking of patterns/trends; loss of detail.

Problems of under and over simplification resulting from unsuitable choice of group size or number of significant figures.

Reading and interpreting data presented in tabular or graphical form.

Design of tables to summarise data effectively. Design and use of appropriate two-way tables.

Diagrammatic Representation

Qualitative data: bar and pie charts, pictograms. Multiple and composite bar charts.

Comparative pie charts (area proportional to total frequency).

Discrete data; vertical line graphs. Cumulative frequency step polygons.

Continuous data; grouped frequency diagrams with equal class intervals. Frequency polygons. Cumulative frequency polygons. Population pyramids.

Histograms with equal or unequal class intervals.

Stem and leaf diagrams.

Shading maps (choropleth maps)

Transforming data presentation from one form to another.

The shapes and simple properties of frequency distributions; symmetrical, positive and negative skew.

The shape and simple properties of the normal frequency distribution.

Bivariate data: scatter diagrams.

Time series: line graphs.

Other diagrammatic representations for comparisons of data using length.

Use of area and volume. Comparison of the various diagrammatic representations using area or volume, including their advantages and disadvantages.

Visual misrepresentation: misuse or omission of origin or scale. Broken, incorrect or changed scales. Incomplete definitions and labelling.

Misuse of length, area and volume in pictorial comparison.

Subject: Statistics (Higher)

Department Staff – Mr E DeSouza, Mr E Gakpo, Mr A Hosseinian,
Mr P Kelly, Mr S McCann, Mr K Pillay,
Mr A Sztranyovsky, Ms D Thompson

Interpretation of information presented in diagrammatic form; distinction between well and poorly presented data.

Spotting possible errors in a data set by recognising outliers that do not fit a general pattern.

Data Analysis

Calculations refer to both continuous and discrete data. Where numerical calculations are involved they should be carried out to a sensible degree of accuracy. Candidates are required to analyse written and statistical evidence as appropriate, as well as consider and check results, modifying their approach if necessary. Calculations and/or statistical results should be interpreted in the initial context of the data where appropriate.

Measures of Location

Mean, median and mode for raw data.

Use of change of origin when calculating the mean. Effect on the average of changes in the sample, eg, the addition or withdrawal of a sample member.

Mean, median and mode for discrete frequency distributions. Modal class for grouped frequency distributions. Median for grouped frequency distributions. Mean for grouped frequency distributions.

Advantages and disadvantages of each of the three measures of location in a given situation.

Reasoned choice of a measure of location appropriate to the nature of the data and the purpose of the analysis.

Geometric mean.

Measures of Spread

Range

Quartiles for discrete data. Quartiles and percentiles, for grouped frequency distributions.

Deciles

Interquartile range for discrete and continuous data.

Interpercentile ranges.

Variance and standard deviation.

Advantages and disadvantages of each of these measures of spread.

Construction of box and whisker plots. Use of box and whisker plots to identify outliers.

Calculation and interpretation of standardised scores.

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.

Other Summary Statistics

Simple index numbers. Weighted index numbers. Chain base numbers. General Index of Retail Prices. (RPI).

Crude rates. Standardised rates.

Time Series

Drawing a trend line by eye and using it for prediction.

Evaluating and plotting appropriately chosen moving averages.

Identification of seasonal variation. Trend line based on moving averages. Seasonal effect at a given data point. Average seasonal effect. Prediction of future values.

Subject: Statistics (Higher)

Department Staff – Mr E DeSouza, Mr E Gakpo, Mr A Hosseinian,
Mr P Kelly, Mr S McCann, Mr K Pillay,
Mr A Sztranyovsky, Ms D Thompson

Quality Assurance

Plotting sample means, medians or ranges over time to view consistency and accuracy against a target value.

Correlation and Regression

Scatter diagrams. Recognition by eye of positive correlation, negative correlation, lack of correlation.

The distinction between correlation and causality.

Spearman's rank correlation coefficient as a measure of agreement; its calculation and limitation in interpretation.

Fitting a straight line by eye through (\bar{x}, \bar{y}) to the plotted points on a scatter diagram.

Obtaining the equation of the fitted line in the form $c + mx = y$; the interpretation of m and c .

Non-linear data.

Interpolation and extrapolation.

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/or rank correlation coefficients.

Estimation

Estimation of population mean from a sample. Estimation of a population proportion from a sample; the use of this method of estimation in opinion polls. Variability in estimates from different samples and the effect of sample size.

Estimation of population size based on the capture/recapture method.

An elementary quantitative appreciation of appropriate sample size.

Probability

Probability

Probability of an event, impossible events, certain events. Use of words such as possible, likely. Putting events into order of probability. Probability on a scale from 0 to 1.

Probability as the limit of relative frequency as the number of observations increases.

Equally likely events.

Sample space: pictorial representation; probability by counting. Use of Venn diagrams,

tables and Cartesian grids.

Exhaustive events.

Mutually exclusive events, the addition law. The general addition law.

Independent events, the multiplication law. The general multiplication law.

Tree diagrams.

An intuitive approach to conditional probability.

Expected frequencies. Comparison of actual frequencies with expected frequencies.

Discrete Probability Distributions

Simple cases of the binomial and the discrete uniform distribution.

Subject: *Statistics (Higher)*

Department Staff – Mr E DeSouza, Mr E Gakpo, Mr A Hosseinian,
Mr P Kelly, Mr S McCann, Mr K Pillay,
Mr A Sztranyovsky, Ms D Thompson

National Curriculum levels at which you will work

The Higher tier exam enables pupils to achieve a grade C up to an A* grade.

Assessment: how you will be tested this term

Autumn Term

Classwork and homework assignments (2 per week). There will be a Mock GCSE in December 2011. There will be on-line home work.

Spring Term

Classwork and homework assignments (2 per week). You will have an end of half term test and an end of term test.

You will do a Controlled Assessment which will form part of your final GCSE grade.

Summer Term

Classwork and homework assignments (2 per week). You will take GCSE exams in either May or June 2012.