

TIGM

Session 2, August 1 – 5, 2005

Data Analysis

Preliminary Course/HSC Course – Some suggestions

Introduction

Consensus seems to be that the Data Analysis section of the General Mathematics course is the most time consuming part of the course.

The syllabus states, “The focus of this unit is the importance of statistical processes and inquiry in society.” (General Mathematics (syllabus), NSW Board of Studies, 1999.)

It calls for the following outcomes:

- P1 develops a positive attitude to mathematics and appreciates its capacity to provide enjoyment and recreation*
- P9 determines an appropriate form of organisation and representation of collected data*
- P11 justifies his/her response to a given problem using appropriate mathematical terminology.*

It then goes on to state the following:

Students learn about and acquire the following skills, knowledge and understanding

- *the importance of analysing data in planning and decision-making by governments and businesses*
- *the process of statistical inquiry, including the following steps:*
 - *posing questions*
 - *collecting data*
 - *organising data*
 - *summarising and displaying data*
 - *analysing data and drawing conclusions*
 - *writing a report*
- *the role of statistical methods in quality control in manufacturing industries*
- *issues of privacy and ethics in data collection and analysis*
- *organisations that collect and/or use statistics, including the Australian Bureau of Statistics (ABS), the United Nations (UN), the World Health Organisation (WHO).*

TIGM session aims

The syllabus writings are admirable. One could wonder if the examination-focused setting we seem to live and work in assists in trying to deliver a rich experience for our students.

So, this provides the TIGM people with a challenge. Is it possible to facilitate a rich and meaningful experience for our students that is much wider than the final examination requires and still ensure they will do their best in an examination where a 'teach them the minimum / drilling preparation seems so attractive? This leads us to the aims of this session.

The aims are:

- To describe a 'pyramid-type' approach to the teaching and learning of Data Analysis in General Mathematics that allows for a rich, meaningful and contemporary student experience that narrows to allow for a highly focused run toward the assessment part of the course.
- To provide the materials that enable one to realise a pyramid-type approach without super-human levels of input from masses of individual teachers (foiling the re-invention of the wheel).
- To have you proficient in the use of the statistical functions of the Casio *CFX 9850GB PLUS*

One could argue that such an approach should result in students finding the examination questions a relative 'walk-in-the-park'. Wouldn't it be nice if this was true for all?

There is little doubt in my mind it will lead to a vastly more enjoyable time for the teacher and student than is the case with a traditional text-book driven course of study.

I would like to thank Stuart Palmer (Head of Mathematics, PLC Croydon) for challenging my enduring 'rose-coloured glass' view of the world of education and for him agreeing to provide some of his resources to assist the delivery of TIGM.

The role of electronic technology

Statistical computation is labourious and largely 'drone work'. The philosophy of our pyramid-type approach is that 'semi-advanced' electronic technology (like a graphics calculator) allows for a large time saving. Saved time can be devoted to the rich, meaningful and contemporary trappings, as asked for in the syllabus. Managing the use of electronic technology so it is not abused and simply used as a black box is not a simple task. I believe materials provided in this session will support you in managing the use in an appropriate manner.

The session materials

The materials take four different forms. They are:

- Open(ish) problems that require students to form strategies to investigate the problems and through which the need for 'content' is revealed.
- Data that allows investigation of the problems to take place. The data is supplied in two forms:
 - MS Excel spreadsheet
 - .cat files – ready for uploading to the Casio *CFX-9850 GB PLUS*
- Succinctly written calculator 'quick reference' sheets which provide the essentials steps that are likely to assist students while working in assessment mode (and learning mode of course). These were prepared by Stuart Palmer (Head of Mathematics, PLC Croydon) and are used extensively by students at PLC and other schools. We greatly thank him for this great contribution.

Supplementary materials

Further support materials are available that are written by Anthony Harradine and Barry Kissane.

They are:

- A replacement unit of work on Statistics (aimed at Year 9, but also highly suitable for Year 11 General. Booklet and data available for download at <http://www.casioed.net.au/learnmaterials/runits.htm#stats>
- Booklet of problems and learning activities written specially for General Mathematics (Preliminary Course). Booklet and data is available for download at <http://www.casioed.net.au/learnmaterials/mathinteractions.htm>
- A document describing clearly how to transfer data from one calculator to another, from a spreadsheet to a calculator and other related processes. This is available for download at <http://www.casioed.net.au/learnmaterials/getstarted/getstarted.htm>

The syllabus content

The syllabus 'skills, knowledge and understanding' are reprinted on the following pages for reference.

DA1: Statistics and society

The focus of this unit is the importance of statistical processes and inquiry in society. The material treated is essential background for the later units in Data Analysis, and may be integrated with the later units if desired.

Students learn about and acquire the following skills, knowledge and understanding

- the importance of analysing data in planning and decision-making by governments and businesses
- the process of statistical inquiry, including the following steps:
 - posing questions
 - collecting data
 - organising data
 - summarising and displaying data
 - analysing data and drawing conclusions
 - writing a report
- the role of statistical methods in quality control in manufacturing industries
- issues of privacy and ethics in data collection and analysis
- organisations that collect and/or use statistics, including the Australian Bureau of Statistics (ABS), the United Nations (UN), the World Health Organisation (WHO).

DA2: Data collection and sampling

The principal focus of this unit is the planning and management of data collection. In some cases, complete data sets may be available, and in other cases a survey may be needed that involves the use of a sample. Although the emphasis is on quantitative data, students should be aware of processes related to categorical data.

Students learn and acquire the following skills, knowledge and understanding

- identification of the target population to be investigated
- determining whether data for the whole population is available (eg the results of a round of a sporting competition), or whether sampling is necessary
- recognising that the purpose of a sample is to provide an estimate for a particular population characteristic when the entire population cannot be accessed
- classification of data as:
 - Quantitative, either discrete or continuous
 - Categorical
eg gender (male, female) is categorical; height (measured in cm) is quantitative, continuous; quality (poor, average, good, excellent) is categorical; school population (measured in individuals) is quantitative, discrete
- distinguishing between the following sample types:
 - random
 - stratified
 - systematic
- determination of which of the above sample types is appropriate for a given situation
- relating sample selection to population characteristics, eg if 20% of the Australian population is aged under 20, your sample should include 20% of under 20s
- generating random numbers with a table or a calculator to assist in establishing random samples
- describing and using the 'capture-recapture' technique for estimating the size of populations, eg the number of fish in a lake
- recognising the effect of sample size in estimating the nature of the population, eg using the number of boys and girls in a particular Year 11 class to estimate the gender ratio in Year 11 across NSW
- use of the principles for effective questionnaire design, such as
 - simple language
 - unambiguous questions
 - respect for privacy
 - freedom from bias
 - consideration of number of choices, if given, eg an even number of choices may force an opinion in particular circumstances, but on occasions it may be appropriate to allow a neutral choice.

DA3: Displaying single data sets

In this unit, students prepare a variety of data displays and consider the appropriateness of each for the stated purpose. The power of statistical displays both to inform and misinform should be emphasised.

Students learn and acquire the following skills, knowledge and understanding

- creation of tally charts and frequency tables to organise ungrouped and grouped data
- creation of dot plots, sector graphs (pie charts), bar graphs, histograms and line graphs, with attention being paid to the scale on each axis
- selection of a suitable scale for each axis of a graph
- noting the capacity of statistical displays for misrepresentation, particularly in the selection of the scale used on the axes
- creation of a stem-and-leaf plot to illustrate a small data set
- drawing a radar chart to display data such as sales figures, temperature or rainfall readings (see example at end of unit)
- division of data into deciles and quartiles
- determination of the range and interquartile range as measures of the spread of a data set
- creation of frequency graphs and cumulative frequency graphs (ogives)
- determining the median and upper and lower quartiles of a data set from a cumulative frequency polygon
- establishment of a five number summary for a data set (lower extreme, lower quartile, median, upper quartile, upper extreme)
- development of a box-and-whisker plot from a five number summary
- linking types of data with appropriate displays, eg continuous quantitative data is best represented by a histogram; categorical data is best represented with a bar graph or sector graph (pie chart)
- describing the strengths and/or weaknesses of sector graphs, bar graphs, histograms, frequency polygons and radar charts, including suitability for data represented.

Technology that may be used in support of this unit

- use of a spreadsheet or a graphing calculator to create frequency tables and statistical graphs

DA4: Summary statistics

The principal focus of this unit is the calculation of summary statistics for single data sets and their use in interpretation.

Students learn and acquire the following skills, knowledge and understanding

- calculation of the mean of small data sets, using the formulae
$$\bar{x} = \frac{\sum x}{n}, \bar{x} = \frac{\sum fx}{\sum f}$$
 where \bar{x} , represents the mean of the sample
- determination of the mean for larger data sets of either grouped and ungrouped data using the statistical functions of a calculator
- calculation of the means of a range of samples from a population
- informal description of standard deviation as a measure of the spread of data in relation to the mean
- determination of the population standard deviation using the σ_n button of a calculator and the sample standard deviation as an estimate of the population measure, using the σ_{n-1} button
- determination of the median and mode(s) of a data set, either from a list or from a frequency table
- selection and use of the appropriate statistic (mean, median or mode) to describe features of a data set, eg median house prices, modal shirt size
- comparison of the summary statistics of various samples from the same population.

Terminology introduced in this unit

mean

median

mode

standard deviation

summary statistic.

Technology that may be used in support of this unit

- calculate summary statistics from inbuilt statistical functions of a graphing calculator
- use a spreadsheet to create frequency tables and calculate summary statistics

DA5: Interpreting sets of data

The principal focus of this unit is the use of data displays, measures of location and measures of spread to summarise and interpret one or more sets of data.

Students learn and acquire the following skills, knowledge and understanding

- identifying measures of location as mean and median
- identifying measures of spread as range, interquartile range and standard deviation
- investigating outliers in small data sets and their effects on the mean, median and mode
- describing the general shape of a graph or display which represents a given data set, eg in terms of smoothness, symmetry or number of modes
- making judgements about the data based on observed features of the display such as shape and skewness
- displaying data in double (back-to-back) stem-and-leaf plots
- displaying data in two box-and-whisker plots drawn on the same scale
- displaying two sets of data on a radar chart
- preparing an area chart to illustrate and compare different sets of data over time (see example at end of unit)
- using multiple displays to describe and interpret the relationships between data sets
- interpreting data presented in two-way table form, eg male/female versus exercise/no exercise
- comparing summary statistics from two sets of data.

DA6: The normal distribution

In this unit, students will apply the properties of the standard normal distribution to the solution of real problems.

Students learn and acquire the following skills, knowledge and understanding

- describing the z-score (standardised score) corresponding to a particular score in a set of scores as a number indicating the position of that score relative to the mean
- using the formula $z = \frac{x - \bar{x}}{s}$ to calculate z-scores, where s is the standard deviation
($s = \sigma_n$ for a population, $s = \sigma_{n-1}$ for a sample)
- using calculated z-scores to compare scores from different data sets
- identifying the properties of data that are normally distributed, ie
 - the mean, median and mode are equal
 - if represented by a histogram, the resulting frequency graph is 'bell shaped'
- using collected data to illustrate that, for normally distributed data:
 - approximately 68% of scores will have z-scores between -1 and 1
 - approximately 95% of scores will have z-scores between -2 and 2
 - approximately 99.7% of scores will have z-scores between -3 and 3
- using these measures to make judgements in individual cases.

DA7: Correlation

In this unit, students investigate the strength of association of data through examining a scatterplot of ordered pairs. Where appropriate, students find the equation of a line of fit and use the equation to make predictions.

Students learn and acquire the following skills, knowledge and understanding

- plotting ordered pairs of data onto a scatterplot
- recognising from the scatterplot:
 - whether the points appear to form a mathematical pattern
 - whether the pattern appears to be linear
- establishing a median regression line to give a line of fit on a scatterplot with a ruler and pencil
- measuring the gradient of the line of fit drawn, with ruler and pencil
- noting the vertical intercept of the line of fit drawn
- establishing the equation of the resulting line of fit in form $y = mx + b$ (see AM2)
- using this equation to make predictions.

The remaining points relate to correlation. Students will not be required to calculate correlation coefficients.

- interpreting the strength of association using a given correlation coefficient
- interpreting the sign of a given correlation coefficient
- recognising that a high degree of correlation does not necessarily imply causality, eg there is a very high correlation between the sizes of one's left and right feet, but one does not cause the other.