QUANTITATIVE TECHNIQUES A
LECTURER RESOURCE DOCUMENT 2014
(First Edition: 2012)

The Independent Institute of Education (Pty) Ltd is registered with the Department of Higher Education and Training as a private higher education institution under the Higher Education Act, 1997 (reg. no. 2007/HE07/002). Company registration number: 1987/004754/07.

DID YOU KNOW?

There are two (2) IIE websites available that lecturers may find useful:

www.iieconnect.co.za

This website provides discussion forums, in which you can interact with lecturers from all the brands as well as faculty members from The IIE to share information, suggest content, discuss module specific issues and find announcements related to the online supplementary content for your students, and generally interact and network.

IIEConnect is also where you can provide feedback, as often as you like, on the modules on which you lecture. Good ideas and strong impressions fade with time and your contribution will be much less detailed if you only share this at the end of the module. We all want the same thing – so the more you can tell us about the module, the more we will be able to make the changes that need to be made!

www.facultybytes.co.za
or
www.facultybytes.net

This website offers a space in which you can publish research articles or practice papers.

Student Portal

The full-service Student Portal provides students with access to their academic administrative information – this includes online calendars, timetables, academic results, module content, financial accounts, etc. It also includes online Module Guides or Module Manuals, assignments and supplementary online materials.

Please encourage your students to engage with the supplementary online materials that are made available to them via their Student Portals.

SAM Admin Portal

The SAM Admin Portal interacts directly with the Student Portal in real time. As a lecturer you have access, via the SAM Admin Portal, to the module offers or classes that you teach. Just as students do, you have access to the Module Information page, which displays the 'Module Purpose' and 'Textbook Information' including online 'Module Guides' or 'Module Manuals' and assignments for the student and the 'Lecturer Resource Guide' for each module offer or class that you teach. For certain modules, electronic supplementary material is available to you via the 'Supplementary Module Material' link. By downloading relevant software you can view, print and annotate these related PDF documents. You can also use the 'Module Discussion Forum' to discuss module topics with your module offer or class such as any supplementary materials, ICE tasks, etc.

IIE Library Online Databases

The following Library Online Databases are available to lecturers. Please contact your librarian if you are unable to access any of these.

EBSCOhost

This database contains full text online articles.
http://search.ebscohost.com/
User name and password: Please ask the librarian

Inmagic

The Online Public Access Catalogue. Here you will be able to search for books that are available in all the IIE campus libraries.
https://library.iie.ac.za/InmagicGenie/opac.aspx
No password required

SABINET

This database will provide you with books available in other libraries across South Africa.
http://reference.sabinet.co.za/sacat
User name and password: Please ask the librarian

e-Brary

This database provides access to full text online books, some of which form part of the IIE recommended reading lists.
http://site.ebrary.com/lib/iieza
Password: Use your Employee Number

Table of Contents

Using this Guide

This guide has been developed to support your use of the prescribed material for this module. There may be occasions when the prescribed material does not provide sufficient detail regarding a particular idea or principle. In such instances, additional detail may be included in the guide. This guide should not, however, be used as a stand-alone textbook, as the bulk of the information that you will need to engage with will be covered in the prescribed material. You will not pass this module if you only use the module guide to study from.

Various activities and revision questions are included in the learning units of this guide. These are designed to help you to engage with the subject matter as well as to help you prepare for your assessments.

Introduction

The primary objective of numeracy concepts in the business environment is to provide essential tools to deal with the increasing amounts of information you are bombarded with, not only in the business environment but also in your everyday life through newspapers, magazines, television, radio and the Internet. The Internet literally puts data at our fingertips. By pressing a few buttons on a computer information from across the world can be found and accessed within minutes.

Numeracy concepts in the business environment include some mathematical calculations and statistical methods. A sound knowledge of mathematics and statistics is the cornerstone for various quantitative activities and good decision making in economics, business, industry, education, defence, science, medicine and engineering. It is included in this degree program because it plays a vital role in all the business fields including financial analysis, economics, accounting, auditing, banking, production and marketing research.

The reason for this is because data is everywhere and no matter what your career; you will at some stage be required to make decisions that involve data using statistical techniques. An understanding of statistical methods will help you make these decisions more effectively.

Your knowledge of numeracy concepts in business will also develop many other fundamental skills, including computing, logical thinking, decision making and communication skills, which are highly valued by employers in all sectors and are essential to most of the other subjects in your course.

H. G. Wells, a science fiction author, wrote in 1903: "Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write." He was quite prophetic because anyone who lacks fundamental numeracy, statistical reasoning, and thinking skills may find they are unprepared to meet the needs of future employers or to understand information presented in the news and media. From experience, most companies and organisations have realised that success depends on the level of analytical, quantitative and statistical skills of their workforce and they therefore seek employees with a sound mathematical training.

Business Statistics can therefore be defined as the science of good decision making in the face of uncertainty. It is used to provide, organise and analyse data using various quantitative analysis tools, testing the significance of data and using the results to make informed business decisions.

Please take note of the formulae sheet that has been inserted in the guide. This is what you will receive in the examinations and tests.

Module Resources	
Prescribed Book for this Module	Wegner, T. 2012. *Applied business statistics methods and Excel-based application*. 3rd edition. Pretoria: JUTA. ISBN: 9780702177743 *Please note that this Module Guide is intended to support your learning – the content of this module is in the prescribed textbook. You will not succeed in this module if you focus on this Module Guide only.*

Module Purpose
This module addresses a selection of numeracy skills and basic concepts and principles of Statistics. It also explains the application of key statistical concepts and techniques.

Module Outcomes	
MO1	Demonstrate an understanding of mathematical principles and relationships;
MO2	Use rules and principles to solve statistical problems;
MO3	Apply statistical techniques in a variety of contexts.

Module Pacer					
Code	**QUAT5112**	**Hour Sessions**	**40**	**Credits**	**6**
Learning Unit 1		**Theme: Mathematical Calculations**		*Notes on this LU*	
Sessions: 1—3 Related Outcomes: • MO1		Learning Objectives: • Compute and interpret percentages, proportions and ratios; • Apply percentages, proportions and ratios in business contexts; • Determine discounts and commissions.			
Learning Unit 2		**Theme: Introduction to Statistics**		*Notes on this LU*	
Sessions: 4—6 Related Outcomes: • MO2 • MO3		Learning Objectives: • Describe the meaning of data and statistics and the role that "uncertainty" plays in decision making; • Distinguish between various terms and concepts utilised in statistical analysis; Distinguish between descriptive and inferential statistics and identify situations in which each is appropriate.			
Learning Unit 3		**Theme: Ungrouped data**		*Notes on this LU*	
Sessions: 7—12 Related Outcomes: • MO2 • MO3		Learning Objectives: • Organise and Describe data using charts and tables; • Compute and Interpret Measures of Location and Dispersion; • Construct the box-and-whiskers plot to indicate skewness and outliers in a given data set.			
25% of ICE should be completed by this point					
Learning Unit 4		**Theme: Grouped data**		*Notes on this LU*	
Sessions:13—20 Related Outcomes: • MO2 • MO3		Learning Objectives: • Organise and Describe data using charts and tables; • Compute and Interpret Measures of Location and Dispersion.			
50% of ICE should be completed by this point					

Learning Unit 5	**Theme: Regression and Correlation Analysis**	*Notes on this LU*
Sessions: 21—27 Related Outcomes: • MO2 • MO3	Learning Objectives: • Identify scenarios where the use of correlation and regression is appropriate. • Plot and interpret scatter diagrams; • Compute and interpret the coefficients of correlation and determination; • Compute and interpret the components of a linear regression function • Forecast values based on the linear regression function	
75% of ICE should be completed by this point		
Learning Unit 6	**Theme: Index Numbers**	*Notes on this LU*
Sessions:28—33 Related Outcomes: • MO2 • MO3	Learning Objectives: • Explain the significance of "index numbers " in the real world; • Differentiate between simple and composite index numbers; • Differentiate between unweighted and weighted index numbers; • Interpret index numbers.	
100% of ICE should be completed by this point		

Learning Unit 7	**Theme: Financial Mathematics**	*Notes on this LU*
Sessions: 34—40 Related Outcomes: • MO2 • MO3	Learning Objectives: • Explain the concepts used in interest calculations; • Identify the principles and notation of simple interest; • Calculate all possible components in simple interest and compound interest; • Calculate unknown present values and future values using compound interest; • Determine the values of different types of annuities: ○ Ordinary annuity certain; ○ Ordinary annuity due; ○ Deferred annuity. • Determine Net Present Value	

Assessments

Integrated Curriculum Engagement (ICE)	
Minimum number of ICE activities to complete	4
Weighting towards the final module mark	10%

Tests/ Examination	Test 1	Test 2	Examination
Weighting	25%	30%	35%
Duration	1 hour	1 hour	2 hours
Total marks	60 marks	60 marks	120 marks
Open/ closed book	Closed book	Closed book	Closed book
Resources required	Calculator, Formulae sheet will be handed out	Calculator, Formulae sheet will be handed out	Calculator, Formulae sheet will be handed out
Learning Units covered	LU1–2	LU3–5	LU1–7

Assessment Preparation Guidelines		
	Format of the Assessment ***(The Focus/ Approach/ Objectives)***	**Preparation Hints** ***(How to Prepare, Resources to Use, etc.)***
Test 1	This test will deal with both practical and theory. A formulae sheet will be handed out.	To prepare effectively for this test you can include the following in your preparation: • Ensure that you work through all the review questions in the LUs tested; • Check if you are confident that you could answer questions relating to all of the Learning Objectives for the LUs tested; • Work through Mock Assessments or previous assessments; • Work through the examples in the prescribed book.
Test 2	This test will deal with both practical and theory. A formulae sheet will be handed out.	To prepare effectively for this test you can include the following in your preparation: • Ensure that you work through all the review questions in the LUs tested; • Check if you are confident that you could answer questions relating to all of the Learning Objectives for the LUs tested; • Work through Mock

		Assessments or previous assessmentsl • Work through the examples in the prescribed book.
Examination	This examination will deal with both practical and theory. A formulae sheet and statistical tables will be handed out.	To prepare effectively for this examination you can include the following in your preparation: • Ensure that you work through all the review questions in the LUs tested; • Check if you are confident that you could answer questions relating to all of the Learning Objectives for the LUs tested; • Work through Mock Assessments or previous assessments; • Work through the examples in the prescribed book.

Glossary of Key Terms for this Module

Term	Definition	*My Notes*
Addition rule (or rule)	P (A or B) means A happens or B happens or both (A and B) can happen.	
Bar graph	A graphical representation of a categorical distribution with a bar drawn for each category. The length of the bar depends on the frequency of a category.	
Binomial distribution	Discrete distribution that gives the probability of "x" successes in "n" trials.	
Box-and-whisker plot	A graph that consists of a box that represents the interquartile range and inside the box a vertical line shows the position of the median. The whiskers extend from the sides of the box to the smallest value on the left and to the largest value on the right.	
Categorical variable	Consists of non-numerical measurements.	
Central Limit Theorem	As the sample size increases, regardless of the population distribution from which it was taken, the sampling distribution of the mean becomes more normal.	
Central tendency	Summary measures that describe the middle of a data set using a single number.	
Chi-square statistic	Tests a hypothesis by comparing a set of expected frequencies with a set of observed frequencies.	
Chi-square test of independence	A hypothesis test for association between two categorical variables.	
Cluster sampling	A sampling method by which a population is divided into groups or clusters. A random sample of n clusters is selected from which the elements of the sample are then selected.	
Coefficient of variation	The ratio of the standard deviation to the mean, expressed as a percentage.	
Complement rule	The complement of an event A is the collection of all possible outcomes not contained in event A.	
Compound (or composite) Index	Measures change for a basket of related items from the base period to the current period.	

Term	Definition	*My Notes*
Conditional probability	The probability that an event B will occur given (or if) that event A has already happened.	
Confidence interval	An interval developed from sample data such that if all possible intervals with the same width were constructed, a proportion of these intervals, known as the confidence level, would contain the true population parameter.	
Confidence level	The proportion of all possible confidence intervals from a sampling distribution that will contain the true population parameter.	
Confidence limits	The proportion of all possible confidence intervals that will contain the true population parameter.	
Contingency table	A table used to classify sample observations according to two or more variables. Also known as a cross-classification table.	
Continuous data	Data obtained through a measurement, not a count, and can assume any value in an interval.	
Correlation coefficient	Measures the strength of a relationship between two variables.	
Critical value	The value that corresponds to a significance level in hypothesis testing to determine the cut-off point between the rejection area of the null hypothesis and the not to reject area.	
Counting rule	These rules involve counting the number of successful outcomes and the total number of outcomes in an experiment.	
Cumulative frequency	Total frequency up to the upper class boundary of given interval.	
Cyclical component	A wavelike pattern in a time series that repeats itself over periods longer than one year.	
Data	A collection of observations about a variable.	
Dependent event	Two events A and B are dependent if the occurrence of event A impacts on the probability of event B occurring.	
Descriptive statistics	Summarises sample data.	

Term	Definition	*My Notes*
Discrete variable	A variable which can be measured only by means of counting or whole numbers.	
Distribution	A collection of measurements.	
Dispersion	The spread or variability of the data around an average.	
Elements	The people or objects about which information is collected.	
Event	One outcome or a collection of outcomes from an experiment.	
Expected frequencies	The pattern of expected responses in a chi-square test.	
Experiment	An *experiment* or *investigation* is a repeatable process or action that causes an event to happen.	
Factorial	$n! = n(n-1)(n-2)(n-3)\ldots 3 \cdot 2 \cdot 1$	
Forecasting period	The time period for which forecasts must be made.	
Frequency	The count or number of observations in a category or interval.	
Frequency distribution	A classification of data elements into class intervals in a manner that displays the number of observations (frequency) per class interval.	
Grouped data	Data summarised into distinct groups or class intervals.	
Histogram	A continuous series of rectangles or bars constructed across class intervals.	
Independent events	Events whose occurrence does not affect the probability of occurrence of another event.	
Independent variable	An X-variable whose values impact the values of another variable known as the dependent (Y) variable.	
Index number	A value that is used to measure change in price and/ or quantity from one time period to another.	
Inferential statistics	Analysis and estimation of population parameters based on sample statistics.	
Interval scale	Quantitative data to which you can add or subtract values to obtain the differences, but no true zero exists.	
Interquartile range	A measure of dispersion that is determined by subtracting the first quartile value from the third quartile value of a set of data.	

Term	Definition	*My Notes*
Irregular variation	Random fluctuations in a time series caused by unpredictable occurrences.	
Least squares criterion	The sum of the squared deviations of the data values from the line of best fit is minimised.	
Linear trend	An increase or decrease trend in a time series in which the rate of change is constant.	
Margin of error	A measurement of how close a point estimate is expected to be to a population parameter within a specified level of confidence. This determines the value that has to be added to or subtracted from a point estimate to determine a confidence interval.	
Mean	A numerical measure of central tendency obtained by dividing the sum of all the data values by the number of data elements.	
Median	A numerical measure of central; obtained by identifying the centre value of a number of data values after having arranged the data values in a decreasing or increasing numerical order.	
Mode	A numerical measure of central tendency obtained by identifying the data value that appears most often in a data set.	
Multiplication rule (AND rule)	P (A and B): Both A and B happen at the same time.	
Mutually exclusive events	Two events A and B are mutually exclusive if the occurrence of event A precludes the occurrence of event B.	
Nominal scale	Defines some categorical characteristics such as gender, occupation, etc. in no specific order.	
Normal distribution	A symmetrical or bell-shaped distribution.	
Non-probability sampling	A non-random criterion is used in the sampling method.	
Null hypothesis	The statement about the population parameter that will be tested.	
Observed frequencies	The pattern of observed responses obtained from a sample.	

Term	Definition	*My Notes*
Ogive	A graph depicting the total number of observations per interval that have values less than the upper boundary of each interval.	
One-tailed test	A hypothesis test in which the rejection area is located in one tail of the sampling distribution – one direction only.	
Ordinal scale data	Nominal data ranked in one or other order.	
Outcome	The result of a single trial in an experiment.	
Outlier	Unusually large or small value.	
Parameter	A numerical measure of description computed from the entire population.	
Percentiles	Percentiles divide a data set into 100 segments. A specific n^{th} percentile can be interpreted as n% of the values in the data set that are smaller than this value, the rest (100 – n%) are higher than this value.	
Pie chart	A circle graph divided into "slices". Each slice represents a category to be displayed.	
Point estimate	A single sample statistic that is used to estimate its corresponding population parameter.	
Poisson distribution	The probability of "x" occurrences of an event over a given interval of time or measurement.	
Polygon	It is a line graph that plots the frequency for each class interval at the midpoint of each class interval.	
Population	A set of all objects or individuals of interest from which a sample can be drawn.	
Primary data	New data collected from scratch.	
Probability	The chance that a particular event will occur/ not occur.	
Proportion	The fraction of values from a data set that has a specific characteristic.	
Probability distribution	A list of all possible outcomes of a course of action together with their corresponding probabilities.	
Qualitative data	Data for which the measurement scale is categorical.	
Quantitative data	Numerical measurements that can be used in calculations.	

Term	Definition	*My Notes*
Quartiles	Dividing a data set into four equal sized segments.	
Questionnaire	List of questions that are used to gather data from respondents.	
Random sample	A sample chosen in such a way that each element in the population has the same (equal) chance of being included in the sample.	
Ratio scale	Highest measurement scale on which you can perform any calculations because of the presence of a true zero.	
Raw data	List of the original observations of a variable.	
Regression analysis	A technique which is used to identify relationships between variables for the purpose of predicting future values.	
Relative frequency	The proportion of the total observations that fall within a given interval.	
Sample	A proportion or subset of a population.	
Sample space	The set of all possible outcomes of an experiment.	
Sampling distribution	A distribution of the statistics obtained from all possible samples of the same size that can be selected from a population. It describes how a particular sample statistic varies about its true, unknown population parameter.	
Sampling error	The difference between a statistic computed from a sample and a parameter computed from a population.	
Scatter plot	A graph of plotted points in which the vertical axis represents the dependent variable (Y) and the horizontal axis represents the independent variable (X). Each plotted point has coordinates whose values are obtained from the X and Y variables.	
Seasonal variation	A wavelike pattern that repeats itself within a period of one year.	
Seasonal index	A number that measures the effect of the seasonal component in time series.	
Secondary data	Already available data. The user is not the one that collected the original data.	

Term	Definition	*My Notes*
Significance level	The maximum allowable probability of committing a Type I error denoted by α.	
Simple index number	Index number calculated for one product only.	
Standard deviation	The square root of the average of the sum of squared differences between the mean and each data value.	
Standard error	A value that measures the spread of the sample means around the population mean.	
Statistic	A summary measure computed from an analysis of a sample, such as an average or standard deviation.	
Statistics	The subject that can be defined as the science of data.	
Stratified sampling	A sampling method in which the population is divided into homogeneous groupings. A representative sample is then selected proportionally at random from each group.	
Systematic sampling	A sampling technique that involves determining a *k* value = $\frac{population\ size}{sample\ size}$. Select the first element at random from value number 1 to the k^{th} value; thereafter select every k^{th} item in the population.	
Test statistic	A value calculated by making use of sample statistics to provide a comparison basis for testing hypothesis.	
Time series data	Measurements collected over a period of time.	
Trend component	The general direction (increasing or decreasing) in which the time series is moving.	
Trial	A single execution of an experiment.	
Two-tailed test	A hypothesis test in which the rejection area is split into both tails of the sampling distribution.	
Uncertainty	A decision environment in which the decision maker does not know what outcome will occur if there is more than one possibility.	
Unweighted index number	No consideration is given to the relative importance of the different commodities in a basket.	

Term	Definition	*My Notes*
Variable	A variable describes an event and is said to be the characteristic of interest.	
Variance	The square of the relative distance between the data points and the mean of a data set.	
Variation	Variation in a set of data if all the values in the data set are not the same value.	
Weighted index number	A composite index that includes a measure of importance (weight) into the calculation.	

Learning Unit 1: Mathematical Calculations	
Material used for this Learning Unit: • Mathematical calculations are not covered in the prescribed textbook.	**My notes**
How to prepare for this Learning Unit: • Before the first class, be sure that you read Learning Unit 1 in this module guide. • As you read these sections, see if you can find the answers to the following questions: o How do you calculate percentages? o Why do you want to do percentage conversions or calculations? o Can you calculate percentages? o Can you determine the increase or decrease in percentages	

1 Introduction

"Percentage" is derived from the Latin words "per centum" meaning "per hundred". The basis of comparison is therefore always 100 and uses the symbol "%". Once a value is expressed in terms of its portion of a 100, the result is indicated as a percentage by adding the percentage sign "%" after the result. It is used as a universal basis for comparison, even if the measurement units of the original data differ.

2 Percentage

Percentages are used widely in business to determine discounts, taxes, interest rates and numerous other comparisons. It is recommended that you always change a percentage to a decimal or a fraction first, before you use it in a calculation.

2.1 Converting Percentages to Fractions and Decimals

Divide the percentage by 100 and drop the % sign. The fraction can be converted to a decimal by dividing the numerator by the denominator.

Example:

Express 75% as a fraction and a decimal.

$$75\% = \frac{75}{100} = 0.75$$

Find the fraction and decimal equivalents of:

1. 44%;

2. 83%;

3. 126%.

2.2 Converting a Fraction or Decimal into a Percentage

Multiply the fraction or decimal by 100 and add the % sign.

Example 1:

Express $\frac{6}{20}$ as a percentage:

$\frac{6}{20} \times 100 = 30\%$

Example 2:

Express 0.14 as a percentage:

0.14 × 100 = 14%

1. Express $\frac{25}{15}$ as a percentage.

2. Express 0.165 as a percentage.

2.3 Calculating the Percentage Amount

Convert the percentage rate (e.g.10%) to a fraction (10/100) or a decimal number (0.10) and then multiply the result by the amount (0.10 × R500). The result is the percentage amount. (R50)

Example:

Calculate 5% of R200:

$$\frac{5}{100} \times R200 = R10$$

Calculate:

1. 14% of 430;

2. 5% of 684.

2.4 Calculating the Percentage Rate

To express one amount as a percentage of another, express the first amount as a fraction of the second and multiply the result by 100.

Example:

What percentage is 8 of 18?

$$\frac{8}{18} \times 100 = 44.44\%$$

1. What percentage is 9 of 18?

2. What percentage is 3 of 67?

2.5 Calculating the Base Amount

If the percentage amount of the base and the percentage rate are known, the base amount can be determined by dividing the percentage amount by the percentage rate.

Example:

If 14% of an amount equals 90, what is the base amount?

$$\frac{90}{14} \times 100 = 642.86$$

If 29% of an amount is 569, what is the base amount?

2.6 Percentage Rate Increase or Decrease

A percentage rate increase or decrease between two amounts is the difference between the two amounts, divided by the initial amount

$$Percentage\ increase\ = \frac{increase}{base\ value}\ \times 100$$

$$Percentage\ decrease\ = \frac{decrease}{basevalue} \times 100$$

Example:

The daily sales of a business has increased from R5 000 per day to R5 500 per day. The difference between the two amounts is R500. The initial amount is R5 000. The percentage rate increase is:

$$\frac{500}{5000} \times 100 = 10\%$$

When you return a rented suit expecting to receive your entire deposit of R2 500, but instead only receive R2 000. What percentage of your deposit is retained by the shop?

2.7 Adding to or Subtracting from a Base Amount Percentage Values of the Base Amount.

For some problems you have to add to the base amount or subtract from the base amount a percentage amount of base amount in order to determine a final amount.

Example 1:

If your salary is R800 per week and you have 21% payroll deductions, what is your net salary per week?
Your deductions are: $\frac{21}{100} \times 800\ = 168$
Your net salary is: $\mathrm{R}800\ - \mathrm{R}168\ = \mathrm{R}632$

Alternatively, if your deductions are 21%:
Your net salary is $100\%\ - 21\%\ = 79\%\ \mathrm{of}\ \mathrm{R}800$

$$\frac{79}{100} \times 800\ = \mathrm{R}632$$

Example 2:

This month's sales are 12% more than last month's sales. If last month's was R20 521, calculate this month's sales.

Percentage amount with last month's sales as base = $\frac{12}{100} \times 20521 = \text{R}2\,462.52$

This month's sales = R20 521 + R2 462.52 = R22 983.52

Alternatively one can reason that if last month's sales is the base (100%), then this month's sales would be 112% (100% + 12%).

This month's sales = $\frac{112}{100} \times 20521 = \text{R}22983.52$

1. The total weight of a packaged article is 25% more than its net weight of 6kg. Determine the total packaged weight of the article.

2. If an article, including the packaging material that increased the weight by 20%, weighs 9kg after it is packed, how much did the article weigh before packaging?

3 Ratio and Proportion Calculations

A fraction, ratio and proportion are different ways to describe the same relationship. A ratio compares two quantities expressed in the same measurement unit and is written as a to b or a: b or $\frac{a}{b}$.

100 kilometres per hour (100 km/ hour), 2.3 children per family (number of children/number of family members) and consumption of 18 kilometres per litre of petrol (18km/ litre) are all examples of ratios.

Example 1:

Comparing two bottles of fruit juice, one with a volume of 1.5 litres and the other with a volume of 2 litres, one can state that the ratio between the two bottles as 1.5:2.
It means that the smaller bottle has $\frac{1.5}{2}$ of the volume of the big bottle, which is also $\frac{1.5}{2} \times 100 = 75\%$ of the big bottle.

Example 2
If a statistics test normally results in a failure rate of 17 out of 200 students, how many failures can be expected if 600 students write the examination?

$$\frac{17}{200} \times 600 = 51 \text{ students}$$

4 Calculating Commission

Commission is a method of compensation for services rendered, such as selling products or services, e.g. leasing, insurance and real estate. It is a type of compensation that is usually associated with sales representatives, agents or brokers.

Sales people are often paid on a straight commission basis. Their total earnings are computed as a percentage of sales.

Commission = Amount sold × Commission rate:

Example 1:

An insurance broker is paid a commission of 15% of premiums written for an annuity. What is his commission on a premium of R205?

Commission = $205 \times \frac{15}{100} = \text{R}30.75$

Example 2:

A sales person's commission is 10% of his total weekly income. If his commission is R9 048 for a particular week, what is his total income?

Income = $\frac{9\,048}{10} \times 100 = \text{R}90\,480$

5 Mark-up, Markdown and Discounts

While a mark-up means adding to a base price, markdown means reducing a selling price. A markdown differs from a discount in the sense that a discount is a reduction in selling price because of method of payment (cash) or due to volumes purchased or early settlement of debtor accounts, while a markdown is a reduction in a current selling price to a new and lower selling price.

A markdown therefore is not influenced by method of payment or volume purchased.

To encourage prompt payment, companies allow their customers a discount if they pay their account within a specified number of days. The terms (the percentage discounts and the number of days) must be clearly stated on the invoice.

Discount = Amount of sale × Rate of discount:

Example 1
Shoprite buys a certain brand of washing powder for R12.50 per kilogram and sells it for R16.90 per kilogram. Calculate the mark-up on cost price that Shoprite applies to this product.

Mark-up percentage on cost price = (16.90 – 12.50/ 12.50) × 100
Mark-up percentage on cost price = (4.40/ 12.50) × 100
Mark-up percentage on cost price = 0.352 × 100
Mark-up percentage on cost price = 35.2%

Example 2:

A company sells imitation leather handbags for R500 each. Due to the slow sales of these handbags, the Sales Manager decides to mark the price of the bags down to R440 each. Calculate the markdown percentage on the original sales price.

Markdown percentage on original sales price:

Because the markdown percentage is unknown, call the unknown percentage "x".
x/ 100 = (500 – 440)/ 500
x/ 100 = 60/ 500
500x = 6 000
x = 12 or 12%

Alternatively the markdown percentage can be calculated as follows:
440/ 500 = 1 – x
0.88 = 1 – x
x = 0.12 × 100
x = 12%

Example 3:

A carpenter receives a trade discount of 20% and a cash discount of 3% from the hardware store. The list price for materials he purchased totalled R4 532 before discount. Find the invoice price and the actual amount paid in cash for the material. What percentage of the list price did he save?

$4\,532 \times \frac{20}{100} = 906.40$

4 532 – 906.40 = R3 625.60 before cash discount.

Cash discount = $3\,625.60 \times \frac{3}{100} = 108.77$
He saved: 906.40 + 108.77 = 1 015.17

That is $\frac{1015.17}{4532} \times 100 = 22.4\%$ of the list price.

Example 4:

The Boot Company allows a 3% discount on payments made within 10 days. What is the amount due if the allowable discount is R6?

Amount due = $\frac{6}{3} \times 100 = R200$

6 Recommended Additional Reading

You should engage extensively with your Recommended Additional Reading as it is an essential part of the learning experience. It will give you different perspectives and engagement opportunities with different authors and schools of thought. This section will give you deeper insight into this discipline and is designed to add value to the module.

Ratio, Rate and Proportion 1. 2009. [Video online]. Available at: http://youtu.be/avjovNP60Nk [Accessed 24 May 2014].

Willemse, I. 2009. *Statistical methods and calculation skills*. 3rd edition. Pretoria: Juta.

7 Recommended Digital Engagement and Activities

Your lecturer will provide you with the blog/ Facebook details for your class from which you can obtain information or give comments. Digital activities will also be available on the blog.

8 Interactive Work Space

8.1 Izimvo Exchange 1

Bring some advertisements to class that advertise products or services offered at discounts. Calculate the discounts offered and the after discount prices of the goods. Where relevant calculate what the prices will be excluding VAT. If an article is offered at a markdown price, calculate what the original or markdown price is.

Notes on Izimvo Exchange1:
Encourage the students to participate by discussing a product on their advertisement.

8.2 Izimvo Exchange 2

This is an exercise you can attempt in class together with a class mate.

Tesco receives a delivery of 12 cases of sugar from one of its distributors. The VAT invoice indicates a total of R2 736.00 (inclusive of VAT) to be paid to the distributor.

Upon checking the consignment, the owner of Tesco finds that three cases are water damaged. He returns the three cases to the distributor and requests a credit for them.

1. How much did Tesco pay per case (Incl VAT)?

2. How much VAT was charged by the distributor on the original invoice?

3. How much credit must the distributors grant Tesco (VAT excluded)?

4. How much VAT is involved in the credit amount?

5. How much must the adjusted VAT amount on the invoice be?

8.3 Activity 1

Notes on Activity 1:
Do this acticivity in class.

Purpose:
The purpose of this activity is to apply ratios and proportions in business.

Task:

1. Three friends decide to contribute R30, R15, and R10 respectively to buy tickets for the national lottery. They agree the division of winnings will be in the same ratio as their contributions. If their winnings amount to R600, what sum of money will each one receive?

2. A tennis player receives 10% of his income from sponsorships, double the percentage from coaching, and the rest from tournaments. If his total income for the year is R300 000, how much does he get from each source?

Commentary Related to Activity Design:
Before attempting the activity, revise the seven different variations of percentage calculations.

8.4 Activity 2

Purpose:
The purpose of this activity is to apply the concept of "commission" in businesses.

Task:

1. Lucky sells R7 300 worth of jerseys to a clothing shop. The clothing shop manager finds three damaged jerseys and returns them to Lucky. The three damaged jerseys cost R300 each. If Lucky's commission is 12%, how much commission does he receive?

2. A salesperson receives R4 500 in commission on R150 000 of sales. What is his commission rate on these sales?

3. Determine the value of goods sold when a commission of R898, based on a rate of 6%, is paid.

4. A salesperson sells goods to the amount of R37 000 for the month. He receives 5% commission on the first R15 000, 6% on the next R10 000 and 7% on the remainder. What is his commission for the month?

5. A paint store pays a basic salary to ensure a fixed income plus a small commission as an incentive feature. Mary receives a salary of R650 per week plus 1% commission on her total sales. What are her total earnings for the week if she sold R900 worth of paint?

Commentary Related to Activity Design:
N/ A

8.5 Activity 3

Purpose:
The purpose of this activity is to apply markdowns and discounts.

Task:

1. Determine the amount of discount on an invoice of R200 dated 15th March, if you pay on the 20th March. The discount is 2% if you pay within 10 days.

2. A customer pays R72 as settlement of an invoice for R80. Calculate the discount rate.

3. Calculate the amount due to the Dodo Company, if the 8% discount allowed amounts to R64.

4. Lady Lace allows discount of 2% on lingerie purchases paid within 10 days and 8% on sportswear purchases paid within 10 days. How much is due for lingerie if the discount allowed is R80? How much is due for sportswear if the discount allowed is R40?

5. The price of a jersey is reduced from R199.99 to R177.99. Calculate the mark-down percentage.

6. A department store is selling a suitcase for R153, after allowing 10% discount. Calculate the original selling price.

8.6 Activity 4

Purpose:
Although percentages have many applications in many disciplines, in business they are usually applied to the pricing of goods or services, to determine final prices after adding profit margins or allowing for discounts. Percentages are used in stock control, returns on investments, cost of capital, growth rates, share prices and interest rates. The purpose of this activity is to look into some more business applications.

Task:

1. An article costs R32 and is sold at a profit of 15%. Calculate the selling price.

2. By selling an article for R63.50, the profit is 25%. Calculate is the cost price.

3. An article costs R250 and is sold for R300. Calculate the % mark-up.

Commentary Related to Activity Design:
N/ A

8.7 Activity 5

Purpose:
The purpose of this activity is to apply Value-Added Tax (VAT) in business calculations.

Task:

1. KLM Stores sells 20 cases of soft drinks to a customer. KLM applies the VAT exclusive method. This means that the VAT is not included in the price, but added separately on the invoice. The price per case is R120.00 and there are no discounts.

    1.1 How much is the invoice amount of this purchase exclusive of VAT?

    1.2 How much VAT is charged on this purchase?

    1.3 Calculate the VAT inclusive amount the customer pays.

2. Joseph wants to determine the actual purchase price his wife paid for a microwave oven excluding the VAT. The amount on the invoice states “R2 508 VAT inclusive.” Help Joseph calculate this.

3. Patio Stores advertises garden furniture sets at R3 200 per set (VAT inclusive).

    3.1 What is the price of a set, exclusive of VAT?

    3.2 How much VAT is charged per set?

4. Sarah purchases a hair dryer for R110 excluding VAT. She negotiates a 10% discount for cash payment.

    4.1 How much does she pay for the hair dryer before VAT?

    4.2 Calculate the discount amount.

    4.3 How much VAT does she pay?

    4.4 State the total invoice amount.

Commentary Related to Activity Design:
The obligation by qualifying vendors to pay Value-Added Tax (VAT) is regulated by the Value-Added Tax Act No 89 of 1991 as amended. This Act stipulates that a tax, known as value-added tax shall be levied:

- On the supply by any vendor of goods or services in the course of the furtherance of his/ her business;
- On the importation of any goods into the Republic of South Africa by any person;
- On the supply of any imported services by any person. The VAT rate is currently 14% on the value of the goods or service as applicable.

8.8 Activity 6

Purpose:
The purpose of this activity is to recognise the correct method to use to do various percentage calculations.

Task:

1. In the 10 years to 2011, employment in a company fell by 780 to 3 240. What is the percentage decline in employment over the decade?

2. Determine the percentage commission if the commission earned on weekly sales of R50 000 is R475.

3. Giftware often carries a mark-up cost of 50%. If the cost of a vase to a retailer is R84, what will the vase retail for?

4. A pair of shoes priced at R793 is marked-up on cost by 30%. What is the cost price to the retailer?

5. The cost to the retailer of a tennis racquet sold for R165 is R110. Calculate the mark-up as a percentage based on cost and the profit percentage based on the list price.

6. The cost of a garment to a boutique is R400.

    6.1 If a profit of 17% is made on the retail price, calculate the retail price.

    6.2 If a loss of 2.5% is made on the retail price, calculate the retail price.

7. All candles in a particular gift shop are priced at R14.25 after a mark-up of 25%. Calculate the cost of the candle to the proprietor.

8. Goods are bought for R30. What must the selling price be in order to yield 10% profit and after a trade discount of 10% is allowed?

9. A suit is marked at R999.99. A trader allows 2.5% discount and still makes 20% profit. How much did the suit cost the trader?

10. You buy a Hi-Fi set and a DVD player for R3 400 and R2 800 respectively. How much VAT in total, will you pay on these transactions?

Commentary Related to Activity Design:
N/ A

8.9 Revision Exercise 1

1. A carpenter is eligible for a 15% trade discount on all purchases from a wholesaler. If R2 400 is the total list price of goods purchased, how much do the goods cost the carpenter?

2. If Mark deposits R1 000, which is 8% of his salary into a savings account at the end of a month, how much does he earn?

3. If 20% of a number equals 80, what is the number?

4. If you earn R25 per hour after a pay increase of 7%, how much did you earn before the increase?

5. This month's sales dropped by 9% compared to last month. If this month's sales are R32 953, calculate last month's sales.

6. If you earn R20 per hour and you receive a pay increase of 15%, how much is your hourly wage after the increase?

7. If the retail price of a bicycle is R189 including VAT and the cost price of the bicycle is R120 excluding VAT, calculate the retail price before including VAT.

8. Based on the information from Question 7. How much is the mark-up percentage on cost price if the retail price is R189 excluding VAT?

9. If the price of a bicycle is reduced in a sale from R189 to R175.50, calculate the percentage decrease in the price to the customer.

10. Sarah buys a dress with a price of R299.99 inclusive of VAT. Because she is a retailer, she is entitled to a trade discount of 25% before VAT. How much does she pay for the dress inclusive of VAT?

8.10 Revision Exercise 2

1. Tom, Dick and Harry are partners in a business worth R10 250. Tom owns 68% of the business and Dick and Harry have equal shares. Calculate the Rand value of their shares.

2. If 600 textbooks are to be divided between two schools in the ratio of 1 to 2, how many textbooks will each school get?

3. Eight slabs of chocolate cost R32. How much do three slabs cost?

4. John takes 30 minutes to walk from his home to school at a speed of 4km/ h. How long will he take if he cycles at 10km/ h?

5. A lecturer takes three hours to mark the books of all students in her class. How long will it take three lectures to mark the same books if they all work at the same pace?

6. At the end of 2001 there were 101 stores open in South Africa, 39 in Gauteng, 33 in Cape Town, 19 in Natal and 10 in the Free State. Calculate the percentage of stores in each province in relation to the total number of stores.

Note to Lecturer:
Please make it clear to the students that these must add up to 100%.

7. Each section in a department store is given a target for the year. Jack's section has a target increase of 25% over last sales. If last year's sales were R1.5 million, how much is Jack's target sales?

8. It takes three markers 120 hours to mark student examination scripts. Assuming they all work at the same pace, calculate how long it will take if there are:

   8.1 6 markers;

   8.2 10 markers;

8.3 20 markers.

9. Travelling at 50km/ h, I can complete a journey in six hours. How long will it take to complete the same journey at 40km/ h?

10. A farmer buys enough chicken feed to last 200 chickens for a week. How long will the same amount of feed last 350 chickens? (Assume all chickens eat the same amount each day.)

Solutions to Exercises

Izimvo Exchange 2

Questions:	Model Solutions:
This is an exercise you can attempt in class together with a class mate. Tesco receives a delivery of 12 cases of sugar from one of its distributors. The VAT invoice indicates a total of R2 736.00 (inclusive of VAT) to be paid to the distributor. Upon checking the consignment, the owner of Tesco finds that three cases are water damaged. He returns the three cases to the distributor and requests a credit for them.	
1. How much did Tesco pay per case (Incl VAT)?	R2 736 ÷ 12 = $R228$
2. How much VAT was charged by the distributor on the original invoice?	$\frac{2\,736}{114} \times 14 = \text{R}336$
3. How much credit must the distributors grant Tesco (VAT excluded)?	$\frac{228}{114} \times 100 \times 3 = \text{R}600$
4. How much VAT is involved in the credit amount?	$600 \times \frac{14}{100} = \text{R}84$
5. How much must the adjusted VAT amount on the invoice be?	R336 – R84 = R252

Activity 1

Questions:	Model Solutions:
1. Three friends decide to contribute R30, R15, and R10 respectively to buy tickets for the national lottery. They agree the division of winnings will be in the same ratio as their contributions. If their winnings amount to R600, what sum of money will each one receive?	This is a ratio calculation. 30 + 15 + 10 = 55 • $\frac{30}{55} \times 600 = \text{R}327.27$ • $\frac{15}{55} \times 600 = \text{R}163.64$ • $\frac{10}{55} \times 600 = 109.09$
2. A tennis player receives 10% of his income from sponsorships, double the percentage from coaching, and the rest from tournaments. If his total income for the year is R300 000, how much does he get from each source?	Sponsorships = 10% $\frac{10}{100} \times 300\,000$ = R30 000 Training = 20% $\frac{20}{100} \times 300\,000$ = R60 000 Tournaments = 70% $\frac{70}{100} \times 300\,000$ = R210 000

Activity 2

Questions:		Model Solutions:
1.	Lucky sells R7 300 worth of jerseys to a clothing shop. The clothing shop manager finds three damaged jerseys and returns them to Lucky. The three damaged jerseys cost R300 each. If Lucky's commission is 12%, how much commission does he receive?	Calculate the final invoice amount: R7 300 – (R300 × 3) = R6 400 Commission: $\frac{12}{100} \times 6\,400 = \text{R}768$
2.	A salesperson receives R4 500 in commission on R150 000 of sales. What is his commission rate on these sales?	$\frac{4\,500}{150\,000} \times 100 = 3\%$
3.	Determine the value of goods sold when a commission of R898, based on a rate of 6%, is paid.	898 = 6%, how much is 100% $\frac{898}{6} \times 100 = \text{R}14\,966.67$
4.	A salesperson sells goods to the amount of R37 000 for the month. He receives 5% commission on the first R15 000, 6% on the next R10 000 and 7% on the remainder. What is his commission for the month?	R15 000 @ 5% $\frac{5}{100} \times 15\,000 = 750$ R10 000 @ 6% $\frac{6}{100} \times 10\,000 = 600$ R12 000 @ 7% $\frac{7}{100} \times 12\,000 = 840$ Commission = R750 + R600 + R840 = R2 190
5.	A paint store pays a basic salary to ensure a fixed income plus a small commission as an incentive feature. Mary receives a salary of R650 per week plus 1% commission on her total sales. What are her total earnings for the week if she sold R900 worth of paint?	R650 + $\left(\frac{1}{100} \times 900\right) = \text{R}659$

Activity 3

Questions:		Model Solutions:
1.	Determine the amount of discount on an invoice of R200 dated 15th March, if you pay on the 20th March. The discount is 2% if you pay within 10 days.	$\frac{2}{100} \times 200 = \text{R}4.$ The invoice after discount is R196.
2.	A customer pays R72 as settlement of an invoice for R80. Calculate the discount rate.	Difference = R8 and the base is R80. $\frac{8}{80} \times 100 = 10\%$
3.	Calculate the amount due to the Dodo Company, if the 8% discount allowed amounts to R64.	If R64 = 8%, how much is 100%? $\frac{64}{8} \times 100 = \text{R}800$
4.	Lady Lace allows discount of 2% on lingerie purchases paid within 10 days and 8% on sportswear purchases paid within 10 days. How much is due for lingerie if the discount allowed is R80? How much is due for sportswear if the discount allowed is R40?	Lingerie: 2% = R80, how much is 100% $\frac{80}{2} \times 100 = \text{R}4\,000$ Sportswear: 8% = R40, how much is 100%? $\frac{40}{8} \times 100 = \text{R}500$
5.	The price of a jersey is reduced from R199.99 to R177.99. Calculate the mark-down percentage.	Difference = R22 on the base of R199.99 $\frac{22}{199.99} \times 100 = 11\%$
6.	A department store is selling a suitcase for R153, after allowing 10% discount. Calculate the original selling price.	Original price = 100% – 10% discount = 90% = R153 $\frac{153}{90} \times 100 = \text{R}170$

Activity 4

Questions:		Model Solutions:
1.	An article costs R32 and is sold at a profit of 15%. Calculate the selling price.	$32 + \left(\frac{15}{100} \times 32\right) = \text{R}36.80$
2.	By selling an article for R63.50, the profit is 25%. Calculate is the cost price.	R63.50 = cost price (100%) + 25% = 125% $\frac{63.50}{125} \times 100 = \text{R}50.80$
3.	An article costs R250 and is sold for R300. Calculate the % mark-up.	Difference = R50 on the base of R250 $\frac{50}{250} \times 100 = 20\%$

Activity 5

Questions:	**Model Solutions:**
1. KLM Stores sells 20 cases of soft drinks to a customer. KLM applies the VAT exclusive method. This means that the VAT is not included in the price, but added separately on the invoice. The price per case is R120.00 and there are no discounts. 1.1 How much is the invoice amount of this purchase exclusive of VAT? 1.2 How much VAT is charged on this purchase? 1.3 Calculate the VAT inclusive amount the customer pays.	1.1 20 cases × R120.00 per case = R2 400.00, VAT exclusive. 1.2 R2 400 purchase amount × 14% VAT $= \text{R2 400} \times \frac{14}{100} = \text{R336.00}$ 1.3 R2 400 purchase amount + R336 VAT = R2 736.00
2. Joseph wants to determine the actual purchase price his wife paid for a microwave oven excluding the VAT. The amount on the invoice states "R2 508 VAT inclusive." Help Joseph calculate this.	VAT paid on purchase = R2 508 × 14/ 114 VAT paid on purchase = R2 508 × 0.1228 VAT paid on purchase = R308 Actual Purchase price = Invoice amount – VAT Actual Purchase price = R2 508 – R308 Actual Purchase price = R2 200
3. Patio Stores advertises garden furniture sets at R3 200 per set (VAT inclusive). 3.1 What is the price of a set, exclusive of VAT? 3.2 How much VAT is charged per set?	3.1 $\frac{3\,200}{114} \times 100 = \text{R2 807.02}$ 3.2 R3 200 – R2 807 = R393
4. Sarah purchases a hair dryer for R110 excluding VAT. She negotiates a 10% discount for cash payment. 4.1 How much does she pay for the hair dryer before VAT? 4.2 Calculate the discount amount. 4.3 How much VAT does she pay? 4.4 State the total invoice amount.	4.1 $\frac{10}{100} \times 110 = \text{R11}$ She pays R110 – R11 = R99 4.2 R11 4.3 $\frac{14}{100} \times 99 = \text{R13.86}$ 4.4 R99 + R13.86 = R112.86

Activity 6

Questions:		Model Solutions:
1.	In the 10 years to 2011, employment in a company fell by 780 to 3 240. What is the percentage decline in employment over the decade?	The employment was: 780 + 3 240 = 4 020 (base) $\frac{780}{4\,020} \times 100 = 19.4\%$
2.	Determine the percentage commission if the commission earned on weekly sales of R50 000 is R475.	$\frac{475}{5000} \times 100 = 9.5\%$
3.	Giftware often carries a mark-up cost of 50%. If the cost of a vase to a retailer is R84, what will the vase retail for?	$\left(\frac{50}{100} \times 84\right) + 84 = \text{R}126$
4.	A pair of shoes priced at R793 is marked-up on cost by 30%. What is the cost price to the retailer?	R793 = 130% $\frac{793}{130} \times 100 = \text{R}610$
5.	The cost to the retailer of a tennis racquet sold for R165 is R110. Calculate the mark-up as a percentage based on cost and the profit percentage based on the list price.	Difference = R55 profit $\frac{55}{110} \times 100 = 50\%$
6.	The cost of a garment to a boutique is R400. 6.1 If a profit of 17% is made on the retail price, calculate the retail price. 6.2 If a loss of 2.5% is made on the retail price, calculate the retail price.	6.1 $\left(\frac{17}{100} \times 400\right) + 400 = \text{R}468$ 6.2 $400 - \left(\frac{2.5}{100} \times 400\right)$ = R390
7.	All candles in a particular gift shop are priced at R14.25 after a mark-up of 25%. Calculate the cost of the candle to the proprietor.	$\frac{14.25}{125} \times 100 = \text{R}11.40$
8.	Goods are bought for R30. What must the selling price be in order to yield 10% profit and after a trade discount of 10% is allowed?	R30 – 10% = $\frac{90}{100} \times 30 = \text{R}27$ R27 +10% = $\left(\frac{10}{100} \times 27\right) + 27 = \text{R}29.70$
9.	A suit is marked at R999.99. A trader allows 2.5% discount and still makes 20% profit. How much did the suit cost the trader?	R999 – 2.5% = $\frac{97.5}{100} \times 999 = \text{R}974.02$ R974.02 = 120% of cost price. The cost price is the base. $\frac{974.02}{120} \times 100 = \text{R}811.68$

10.	You buy a Hi-Fi set and a DVD player for R3 400 and R2 800 respectively. How much VAT in total, will you pay on these transactions?	R3 400 + R2 800 = R6 200 $\frac{14}{100} \times 6\,200 = \text{R}868$

Revision Exercise 1

Questions:		**Model Solutions:**
1.	A carpenter is eligible for a 15% trade discount on all purchases from a wholesaler. If R2 400 is the total list price of goods purchased, how much do the goods cost the carpenter?	R2 400 – 15% = $2\,400 - \left(\frac{15}{100} \times 2\,400\right) = \text{R}2\,040$
2.	If Mark deposits R1 000, which is 8% of his salary into a savings account at the end of a month, how much does he earn?	$\frac{1000}{8} \times 100 = \text{R}12\,500$
3.	If 20% of a number equals 80, what is the number?	$\frac{80}{20} \times 100 = 400$
4.	If you earn R25 per hour after a pay increase of 7%, how much did you earn before the increase?	R25 = 100% + 7% = 107% $\frac{25}{107} \times 100 = \text{R}23.36$
5.	This month's sales dropped by 9% compared to last month. If this month's sales are R32 953, calculate last month's sales.	R32 953 = 100 – 9% = 91% of last months' sales $\frac{32\,953}{91} \times 100 = \text{R}36\,212.09$
6.	If you earn R20 per hour and you receive a pay increase of 15%, how much is your hourly wage after the increase?	$\frac{115}{100} \times 20 = \text{R}23$
7.	If the retail price of a bicycle is R189 including VAT and the cost price of the bicycle is R120 excluding VAT, calculate the retail price before including VAT.	$\frac{189}{114} \times 100 = \text{R}165.79$
8.	Based on the information from Question 7. How much is the mark-up percentage on cost price if the retail price is R189 excluding VAT?	R165.79 – R120 = R45.79 profit $\frac{45.79}{120} \times 100 = 38.16\%$
9.	If the price of a bicycle is reduced in a sale from R189 to R175.50, calculate the percentage decrease in the price to the customer.	R189 – R175.50 = R13.50. The base is R189 $\frac{13.50}{189} \times 100 = 7.14\%$

10.	Sarah buys a dress with a price of R299.99 inclusive of VAT. Because she is a retailer, she is entitled to a trade discount of 25% before VAT. How much does she pay for the dress inclusive of VAT?	R299.99 – 14%=$\frac{299.99}{114} \times 100 = \text{R}263.15$ R263.15 – 25% = $263.15 - \left(\frac{25}{100} \times 263.15\right) = \text{R}197.36$ R197.36 + 14% = $\frac{114}{100} \times 197.36 = \text{R}224.99$

Revision Exercise 2

Questions:		**Model Solutions:**
1.	Tom, Dick and Harry are partners in a business worth R10 250. Tom owns 68% of the business and Dick and Harry have equal shares. Calculate the Rand value of their shares.	Tom: $\frac{68}{100} \times 10\,250 = \text{R}6\,970$ R10 250 – 6 970 = R3 280 Dick and Harry = $\frac{3\,280}{2}$ = R1 640 each
2.	If 600 textbooks are to be divided between two schools in the ratio of 1 to 2, how many textbooks will each school get?	1 + 2 = 3 $\frac{1}{3} \times 600 = 200$ For school A $\frac{2}{3} \times 600 = 400$ For school B
3.	Eight slabs of chocolate cost R32. How much do three slabs cost?	$\frac{32}{8} \times 3 = \text{R}12$
4.	John takes 30 minutes to walk from his home to school at a speed of 4km/ h. How long will he take if he cycles at 10km/h?	$\frac{4}{60} \times 30 = 2\,\text{km}$: The distance he walks. $\frac{60}{10} \times 2 = 12\,\text{min}:$ The time to cycle to school is 12 min.
5.	A lecturer takes three hours to mark the books of all students in her class. How long will it take three lectures to mark the same books if they all work at the same pace?	$\frac{180}{3} = 60\,\text{min}(\text{or } 1\,\text{hour})$
6.	At the end of 2001 there were 101 stores open in South Africa, 39 in Gauteng, 33 in Cape Town, 19 in Natal and 10 in the Free State. Calculate the percentage of stores in each province in relation to the total number of stores.	Gauteng: $\frac{39}{101} \times 100 = 38.61\%$ Cape Town: $\frac{33}{101} \times 100 = 32.67\%$ Natal: $\frac{19}{101} \times 100 = 18.81\%$ Free State: $\frac{10}{101} \times 100 = 9.91\%$
7.	Each section in a department store is given a target for the year. Jack's section has a target increase of 25% over last sales. If last year's sales were R1.5 million, how much is Jack's target sales?	$\frac{125}{100} \times 1.5 = \text{R}1.875\,\text{million}$

8.	It takes three markers 120 hours to mark student examination scripts. Assuming they all work at the same pace, calculate how long it will take if there are: 8.1 6 markers; 8.2 10 markers; 8.3 20 markers.	8.1 6 markers: $\frac{3}{6} \times 120 = 60$ hours 8.2 10 markers: $\frac{3}{10} \times 120 = 36$ hours 8.3 20 markers: $\frac{3}{20} \times 120 = 18$ hours
9.	Travelling at 50km/h, I can complete a journey in six hours. How long will it take to complete the same journey at 40km/h?	$\frac{300}{40} = 7.5 \text{ hours}$
10.	A farmer buys enough chicken feed to last 200 chickens for a week. How long will the same amount of feed last 350 chickens? (Assume all chickens eat the same amount each day.)	$\frac{200}{350} \times 7 = 4 \text{ days}$

Learning Unit 2: Introduction to Statistics	
Material used for this Learning Unit: • Prescribed text pp.3—24.	**My notes**
How to prepare for this Learning Unit: • Before the first class, be sure that you read pp.3–24 in the textbook. • As you read these sections, see if you can find the answers to the following questions: o What is Statistics? o Do you know the terminology used in Statistics? o Can you explain the different types of data?	

1 What is Statistics?

The purpose of the subject “Statistics” is to help you make sense of all the information, also known as data, we are bombarded with every day through the Internet, newspapers, magazines, radio, television, and other sources of information.

Statistics can therefore be defined as the science of collecting, organising, analysing and interpreting data in order to make decisions.

We make decisions in the face of uncertainty. That means, facing a situation where we know that it is possible that things could turn out in different ways but we simply do not know how probable each possibility is. Our need to cope with this risk or the “chance” that it will happen, leads us to the study and use of probability theory. Coping with this uncertainty or risk is when we know there are different outcomes but we also have some idea of how likely each one is to occur. That means we can never be certain that our inferences are correct but we can predict the probability that the event will occur.

1.1 The Statistical Cycle

Refer to p.4 in the textbook. The three key components in identifying problems are:

- Think systematically;
- Look for connections and relationships;
- Understand why data values differ from one another.

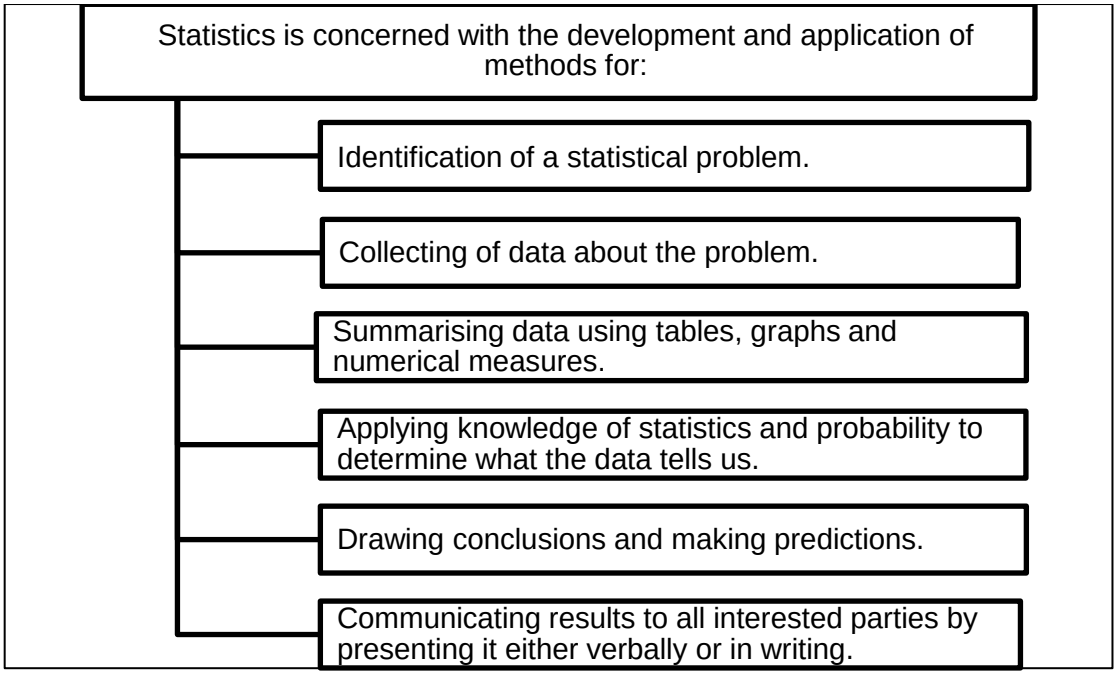


Diagram 2.1: Statistics in summary

2 The Language of Statics

Read pp.5–6 of the prescribed textbook.

3 Basic Components of Statistics

Read pp.7–8 in the textbook.

3.1 Descriptive and Inferential Statistics

The tools and techniques used to collect, classify and summarise the data are known as descriptive statistics, while interpreting the results and drawing valid conclusions form part of the inferential field of statistics. Descriptive statistics forms the basis of all statistical investigations and must therefore be done with the utmost care and accuracy.

3.2 Sample versus Population

- A sample is a portion or a sub-set of a population;
- A statistic is a summary measure for a sample and a parameter for a population;
- If data are collected on all the elements of the population, it is referred to as a census.

3.3 Data versus Variable

The characteristic or topic that the data is collected for is known as the variable of the study and can assume different values. These values or observations are the data.

4 Types of Data

Read pp.9–10 in the textbook.

Quantitative data are numerical values and the variable is of a descriptive nature. Non-numerical data is known as qualitative data. Data such as age, distance and money are all quantitative in nature while type of transport, the make of your shirt, gender, political affiliation and the different subjects you are enrolled for are all examples of qualitative data. Qualitative data is sometimes coded to make it appear quantitative, but will have no meaning on a number line.

Complete Concept Questions 1–5, p.9 in the textbook.

Discrete data result from a count and continuous data from a measurement.

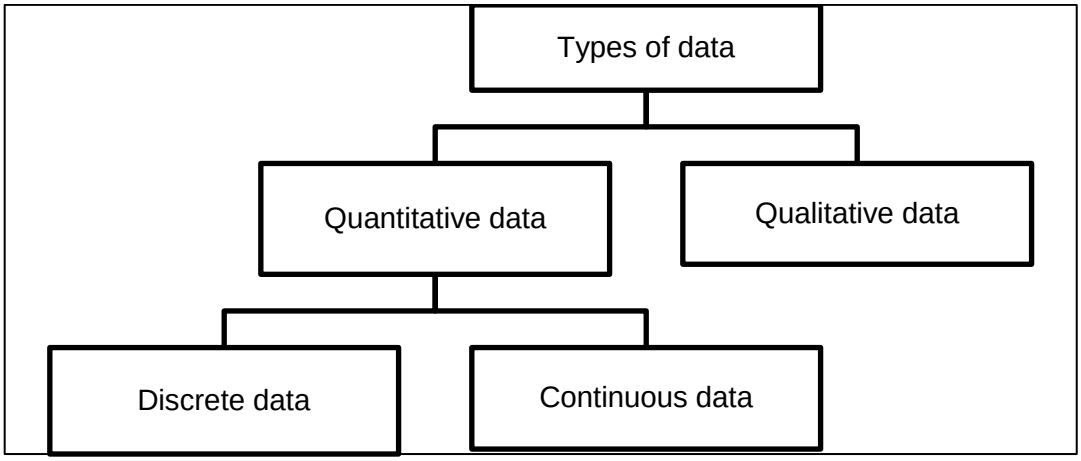


Diagram 2.2: Types of data

5 Measurement Scales

Read pp.10–12 in the textbook.

Measurement is the process used to assign numbers to the observations or elements of a variable. Numbers are used as symbols to represent certain characteristics like age, income, height of the object, person, etc. For example, as a student your student number identifies you.

There are four levels of measurement. Ranked from the weakest to the strongest they are:

- Nominal: Names, labels and categories in no specific order. For example, gender or city of residence;
- Ordinal: Categories into which data are grouped are ranked in some order. For example, clothing size (S/ M/ L);
- Interval: Numerical data, but there is no true zero. (The zero has meaning and does not represent "nothingness".) For example, What is your opinion about the lastest Masterchef series? 1 – Very boring, 2 – Somewhat boring, 3 – Neutral, 4 – Exciting, 5 – Fantastic;
- Ratio: Numerical values that can be used in arithmetic operations. The true zero exists. For example, all real numbers such as Employee ages (Years), Customer Income (Rand), Distance Travelled (km).

6 Sample Design

Read pp.12–17 in the textbook. You must be able to follow a specified method to collect data and if a method is described, you must be able to identify and apply it.

Complete Exercise Questions 1–14, pp.19–24 in the textbook.

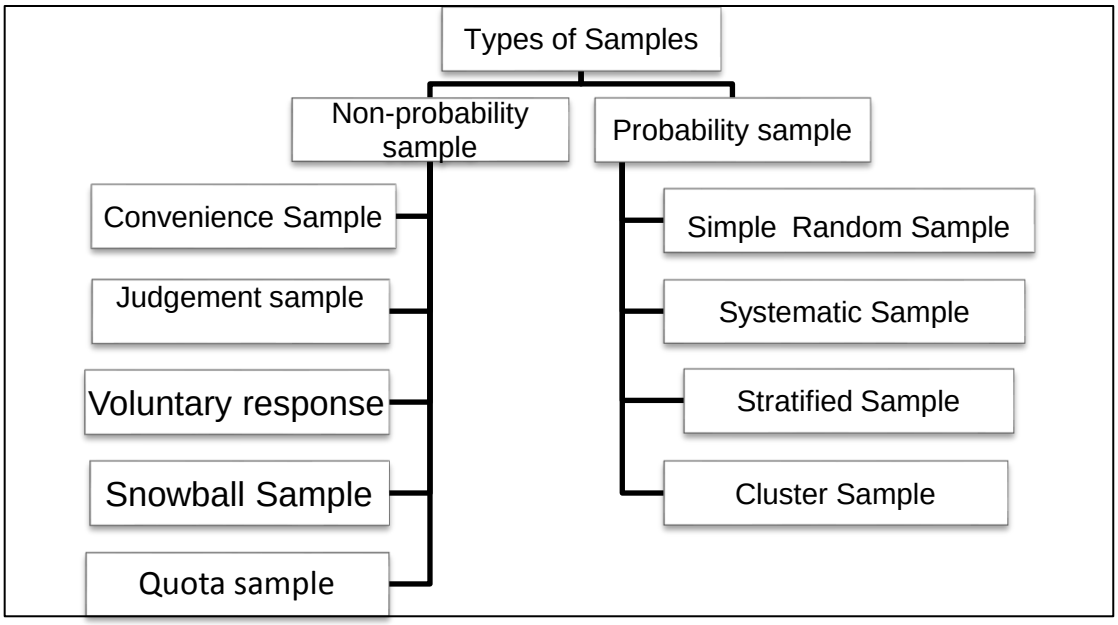


Diagram 2.3: Types of Samples

7 Recommended Additional Reading

You should engage extensively with your Recommended Additional Reading as it is an essential part of the learning experience. It will give you different perspectives and engagement opportunities with different authors and schools of thoughts.

This will give you deeper insight into this discipline and has been designed to add value to the module.

Johnson, RR and Kuby, PJ. 2011. *Elementary statistics*. 11th edition. Pacific Grove, (CA): Duxbury.

Lombaard, C, Van der Merwe, L, Kele, T and Mouton, S. 2010. *Elementary statistics for business and economics*. 1st edition. Cape Town: Heinemann Pearson Publishing.

Sullivan, M. 2010. *Fundamentals of statistics*. 3rd edition. Boston, (MA): Pearson's Education.

The Animated Software Company. 2013 *Internet Glossary of Statistical Terms.* [Online]. Available at: www.animatedsoftware.com/statglos/statglos.htm [Accessed 24 May 2014].

Triola, MF. 2009. *Elementary statistics*. 11th edition. Boston, (MA): Pearson's Education.

Weaver, S. 2012. *STAT 141 - Basic statistical methods.* [Video online]. Available at: http://youtu.be/7Mfa1MbVhcc [Accessed 24 May 2014].

Willemse, I. 2009. *Statistical methods and calculation skills*. 3rd edition. Pretoria: Juta.

8 Recommended Digital Engagement and Activities

Your lecturer will provide you with the blog/ Facebook details for your class from which you can obtain information or give comments. Digital activities will also be available on the blog.

Access the following and familiarise yourself with all the available features on this website:

Triola, MF. 2011 *Elementary statistics.* Pearson Education. [Online]. Available at: http://www.aw.com/triola [Accessed 24 May 2014].

9 Interactive Work Space

9.1 Izimvo Exchange 1

Some examples of the kind of problems that call for statistical thinking are:

- What is the most popular means of transport to class? Is it a bus, car, taxi, train or walking?
- The variations in the death toll on the roads during holiday seasons;
- The influence that changes in the price of petrol can have on the price of taxi or bus fare;
- Before every sport commentary you will hear the broadcaster giving the statistics of what happened to the same teams in similar circumstances over the last couple of years;
- "Every 26th second someone in South Africa becomes a victim of crime;"
- The voter support for the different political parties in the country;
- A well-known golfer is concerned that there are no uniform standards for equipment used during big tournaments and this can cause some players to benefit from the improvement of technology and others not;
- Is placing more police "back on the street" likely to reduce crime?
- Does using unleaded petrol really make a significant contribution to less air pollution?
- Is there any relationship between smoking and lung cancer?

Have a classroom discussion about some more examples where statistics are used. You can also make use of some articles printed in newspapers or magazines to discuss.

In the third example above, two variables can be identified: petrol price and travel fare. Identify the variables in the other examples.

9.2 Izimvo Exchange 2

The following data represent the quantity of soft drinks (in millilitres) in 50 2ℓ litre bottles in the fridge of a convenience store at a service station.

2.109	2.086	2.066	2.075	2.065	2.057	2.052	2.044	2.035	2.038
2.031	2.012	1.994	1.963	2.029	2.012	1.986	1.957	2.025	2.012
1.984	1.951	2.029	2.010	1.981	1.951	2.023	2.005	1.973	1.947
2.020	2.003	1.975	1.941	2.015	1.999	1.971	2.014	1.996	1.941
1.969	1.938	2.013	1.997	1.966	1.908	2.014	1.992	1.967	1.894

1. Use a random number table to select a sample of 10 bottles from this population.

2. Use the systematic sampling method and select a sample of 10 bottles from this population.

3. Calculate the average contents for each of the samples.

4. It is expected that each bottle contains two litres of soft drink. Calculate how much the sample averages deviate from the true population average (2 litres) by subtracting the sample averages from the population average. These deviations are known as the sampling error.

5. Compare your answers with those of your class mates. The answers will most probably differ. This is due to variability in sampling.

A random number table is given on p.555 in the textbook. Follow the steps on p.10 in the textbook to obtain the sample. An acceptable random way of selecting the starting value is to close your eyes and point with a pencil on the table.

9.3 Izimvo Exchange 3

Draft a short questionnaire to collect data on how Grade 10 learners spend their leisure time. Take into account the layout, question ordering and wording. Make use of closed as well as open ended questions.

Note to the Lecturer:
Read pp.25–27 in the textbook on questionnaire design and let the students draw up a questionnaire with only three to four questions. Each question must be of a different type.

9.4 Activity 1

Purpose:
The purpose of this activity is to distinguish between statistical concepts antonyms.

Task:

1. Describe the population and the sample in each of the following situations:

1.1 50 smokers were selected at random to determine what the effectiveness of a televised anti-smoking campaign is on smokers.

1.2 A very popular radio station selected 200 people at random to determine listeners' attitudes towards certain programmes it runs during the day.

1.3 In a recent survey 3 000 South Africans were asked if they read the newspaper daily, 600 said yes.

2. Identify if primary or secondary data are used:

2.1 Information obtained to determine the water levels of the major storage dams in South Africa will be primary/ secondary data. Where can this data be obtained from?

2.2 Describe how you will go about to determine the number of filling stations in your area using both a secondary source and a primary source.

2.3 Identify a problem where you can make use of secondary data.

2.4 Identify a problem where you can make use of primary data.

3. For each of the following situations, would you recommend a sample or a census? Explain your reasoning in each case.

3.1 A jeweller just received a delivery of shock-resistant watches and wants to find out approximately from how high they can be dropped onto a concrete surface without breaking the crystal.

3.2 Tiger Mills wants to learn the age, gender and income characteristics of persons who consume Cheerio breakfast cereal.

3.3 The producers of the Early Bird show want to find out what percentage of television (TV) viewers recognise a photo of host, Joseph Khumalo.

3.4 A researcher wants to determine whether companies that manufacture nuclear submarines might be interested in a new technique for purifying air when such craft are submerged.

4. Distinguish between a parameter and a statistic:

4.1 In surveying political choices of voters, a sample of voters is selected from the population of all eligible voters. Based on the results observed from the sample statistics the analyst proceeds to make inferences on the political choices likely to exist in the population of voters.

4.2 A recent survey of a sample of graduates reported that the average starting salary for a graduate is less than R30 000 per year.

4.3 Starting salaries for 270 graduates increased with 5% from the previous year.

4.4 In a random check of a sample of retail stores, the local health inspector found that 24% of the stores were not storing fish at the proper temperature.

4.5 In 2004, all major league soccer teams spent a total of R1 968 088 on players' salaries. Decide whether the numerical value is from a population or a sample and then specify it as parameter or a statistic.

5. Indicate whether the corresponding sets of observations would be quantitative or qualitative. If quantitative, distinguish between discrete and continuous.

	Quantitative	**Qualitative**	**Discrete**	**Continuous**
Ethnic group				
Age				
ID number				
Net worth (rand)				
Favourite sport				
Temperature				
Home language				
Cooking time for pasta				
Speed of an aeroplane				
Gender				

Commentary related to this activity:
In later chapters these concepts are used to distinguish between which formulae or methods to apply.

9.5 Activity 2

Purpose:
The purpose of this activity is to identify the measurement scale to use for the different variables.

Task:
Specify the measurement scale for each of the following:

1. Whether you are a South African (SA) citizen.

2. The amount you paid to fill up your petrol tank.

3. The time it took you to get to the university this morning.

4. The size of your take-away coffee.

5. Your belt size.

6. Your student number.

7. The occupation of 200 shoppers at a supermarket.

8. The daily temperature measured inside the supermarket.

9. The amount spent by every shopper at a supermarket.

10. Rating a new product as good, average or poor.

11. The recording of the first three digits of the shoppers' cell phone numbers.

Commentary related to this activity:
When dealing with measures of description, only certain descriptive measures can be used with certain measurement scales of data.

9.6 Activity 3

Purpose:
The purpose of this activity is to analyse a case study in terms of the basic statistical concepts.

Task:
The South African government is concerned about the high illiteracy rates amongst adults in South Africa. They wish to estimate the true proportion of adults (over 18 years of age) in South Africa who are illiterate (that is, they cannot read or write in at least one language). A random sample of 10 000 adults were interviewed and 1 107 of them were found to be illiterate.

1. The sample space is _____?

2. The parameter of interest is _____?

3. The statistic is _____?

4. The sample size is _____?

5. The population size is _____?

6. The variable is _____?

7. The measurement scale of the variable is _____?

8. The type of data is _____?

Commentary related to this activity:
Competency in this activity means that you have a good understanding of some of the concept in statistics.

9.7 Revision Exercise 1

1. Classify the following sets of data as qualitative or quantitative. If classified as a quantitative state; is it discreet or continuous?

	Quantitative	Qualitative	Discrete	Continuous
The weight of each member of a soccer team				
Religious affiliation				
Marks obtained in 1st test				
The colours you can identify in a rainbow				
Telephone numbers in a telephone directory				
The number of sit-ups you can do				
The daily temperature at 12h00				
Number of traffic fatalities				
Time required to complete a crossword puzzle.				

2. What is the measurement scale for the given variables?

Religious affiliation	
The number of sit-ups you can do	
The daily temperature at 12h00	
Number of traffic fatalities	
Time required to complete a crossword puzzle	
The ages of the learners in your study group	
The colours you can identify in a rainbow	
Telephone numbers in a telephone directory	
Shoe sizes.	
The three major professional tennis tournaments listed: Australian Open; Wimbledon; US Open	
The amount of weight lost in the past month by a person following a strict diet	
The classification of Boeings as 727, 737 or 747 according to size	

3. A recent study on the average amount of time spent watching TV by a group of students yielded the following results:

Academic Status	Number of Students	Average Time (Hours) Spent Watching TV
1st years	30	6.4
2nd years	20	4.5
3rd years	10	2.8

Choose the correct word in each of the following statements:

The 60 students involved in the study constitute a sample/ population of students.	
The figure 6.4 is known as a parameter/ statistic.	
The estimate of the average amount of time spent watching TV per week by all 1st year students involves a descriptive/ inferential technique.	
The variable of interest is the number of students/ average time watching TV	
The average times calculated for the three groups of students are of a discrete/ continuous nature.	
The academic status of the students can be classified as quantitative/ qualitative data	

9.8 Revision Exercise 2

1. The collection of all possible individuals, objects or measurements is known as:
   - 1.1 A sample;
   - 1.2 A population;
   - 1.3 An inference;
   - 1.4 A statistic.

2. Techniques used to organise, summarise and present the data that have been collected are known as:
   - 2.1 Populations;
   - 2.2 Samples;
   - 2.3 Inferential statistics;
   - 2.4 Descriptive statistics.

3. Techniques used to estimate a population parameter, based on a sample, are known as:
   - 3.1 Populations;
   - 3.2 Samples;
   - 3.3 Inferential statistics;
   - 3.4 Descriptive statistics.

4. In a random sample each item in the population has:
   - 4.1 A chance of being selected;
   - 4.2 The same chance of being selected;
   - 4.3 A 50% chance of being selected;
   - 4.4 No chance of being selected.

5. The sample mean is an example of a _____.
   5.1 Sample statistic;
   5.2 Population parameter;
   5.3 Measurement scale;
   5.4 Discrete variable.

6. Primary data are collected by:
   6.1 Primary school children;
   6.2 People doing it for the first time;
   6.3 The actual people who will be using it;
   6.4 Mainly experienced people.

7. Secondary data are obtained from:
   7.1 Secondary school children;
   7.2 Existing sources;
   7.3 The actual people who will be using it;
   7.4 Mainly experienced people.

8. When conducting a survey, you collect data by:
   8.1 Sampling;
   8.2 Using a secondary source;
   8.3 Asking questions;
   8.4 Using the random number table.

9. To estimate the percentage of defects in a recent manufacturing batch, the quality control manager of Intel Computers selects every 8^{th} chip that comes off the assembly line starting with the 3^{rd} until he obtains a sample of 100. The method he follows to obtain the sample is known as the_____.
   9.1 Simple random sampling method;
   9.2 Systematic sampling method;
   9.3 Stratified sampling method;
   9.4 Snowball sampling method.

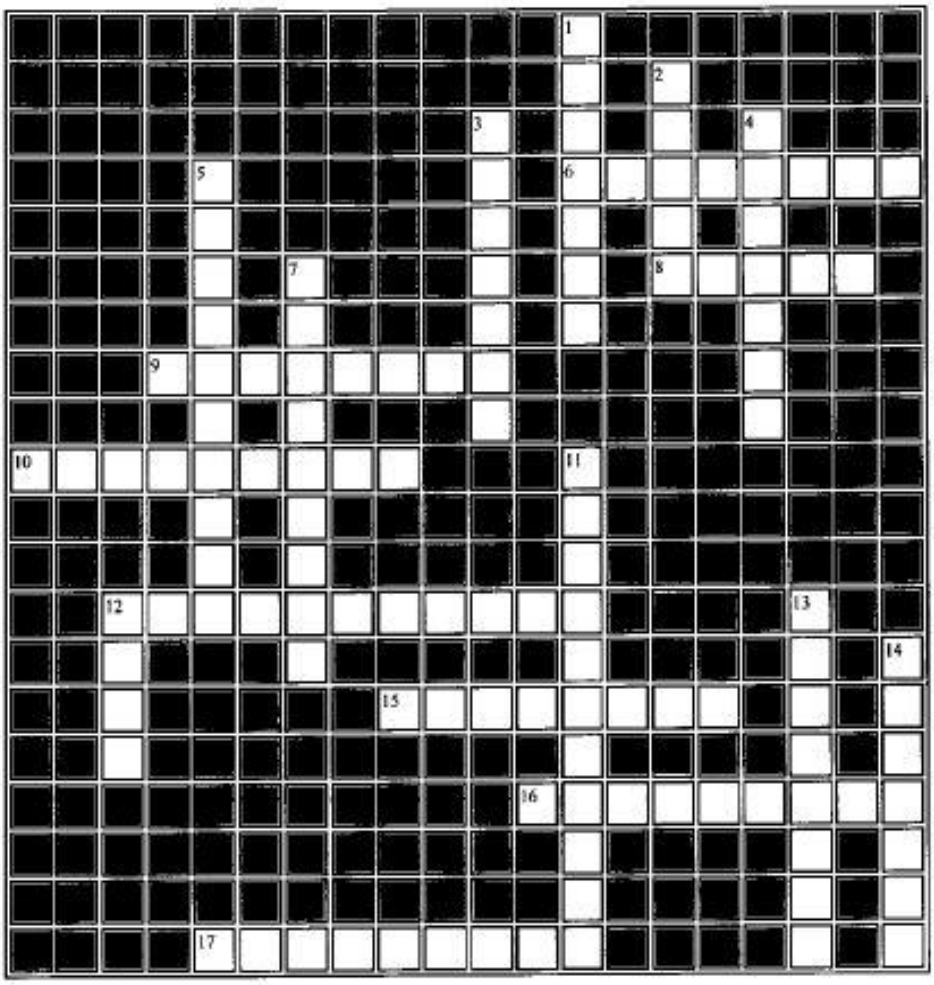
10. Once every hour a random sample of 12 light bulbs is selected from an assembly line delivering this type of light bulb. The number of bulbs in each sample that will not light is divided by 12 to obtain the defective proportion. What is the variable?
    10.1 Type of light bulbs;
    10.2 Sample of 12 light bulbs;
    10.3 Proportion of defective light bulbs;
    10.4 Assembly line.

11. 49, 34, and 46 students are selected from the first year, second year and third year classes with 496, 348 and 481 students respectively. Identify the type of sampling used.
    11.1 Stratified;
    11.2 Cluster;
    11.3 Systematic;
    11.4 Simple random.

12. To avoid working late, a quality control analyst simply inspects the first 100 items produced in a day. This type of sampling used is known as:
    12.1 Cluster sampling;
    12.2 Systematic sampling;
    12.3 Judgmental sampling;
    12.4 Convenient sampling.

9.9 Revision Exercise 3

Complete the following crossword.



Across

6. Data generated from rating scales and used in survey questionnaires
8. This graph represents the "less than cumulative frequency" totals
9. This data consists of whole numbers (eg. no. of cars parked in garage)
10. What does this symbol "fi" represent?
12. Type of Statistics to condense information into tabular and graphical methods (what you are doing at present)
15. What does this symbol "xi" represent?
16. A graph which doesn't have spaces between the bars
17. A quantitative data type where real numbers can be manipulated using arithmetic operations to produce meaningful results

Down

1. Form of data where categories are of equal importance
2. This is the strongest and most desirable data to gather
3. This formula has a name; $c = 1 + 3.3 \log (n)$
4. This form of data implies ranking, where categories are of unequal width
5. This data can consist of decimals (eg. distance travelled to VC)
7. Data collected and processed by others (getting information from the government for a research project you are working on)
11. Refers to data tupes representing categorieds of outcomes (and is qualitative)
12. Actual values recorded on a random variable
13. Data available from outside organisations
14. Data which is captured for the first time with a specific purpose in mind

Solutions to Exercises

Izimvo Exchange 2

Questions:	Model Solutions:
The following data represent the quantity of soft drinks (in millilitres) in 50 2ℓ litre bottles in the fridge of a convenience store at a service station.	
1. Use a random number table to select a sample of 10 bottles from this population.	The answers will vary depending on the starting point in the random number table.
2. Use the systematic sampling method and select a sample of 10 bottles from this population.	Calculate k = $\frac{50}{10} = 5$. That means you want to select every fifth element from the population. Choose the starting point at random from the first five numbers and thereafter every fifth number. The easiest way is to use five pieces of paper of the same size and write the numbers one to five on each one.
3. Calculate the average contents for each of the samples.	Average = $\frac{100.035}{50}$ = 2.0007
4. It is expected that each bottle contains two litres of soft drink. Calculate how much the sample averages deviate from the true population average (2 litres) by subtracting the sample averages from the population average. These deviations are known as the sampling error.	2.0007 – 2.0000 = 0.0007. That is the difference between the expected fill of 2ℓ and the sample average of 2.0007.
5. Compare your answers with those of your class mates. The answers will most probably differ. This is due to variability in sampling.	

Activity 1

Questions:	**Model Solutions:**
1. Describe the population and the sample in each of the following situations: 1.1 50 smokers were selected at random to determine what the effectiveness of a televised anti-smoking campaign is on smokers. 1.2 A very popular radio station selected 200 people at random to determine listeners' attitudes towards certain programmes it runs during the day. 1.3 In a recent survey 3 000 South Africans were asked if they read the newspaper daily, 600 said yes.	1.1 The sample is the 50 smokers selected at random. The population consists of all smokers. 1.2 The sample is the 200 listeners selected at random from the population of all listeners of that radio station. 1.3 The sample is the 3 000 South Africans that were questioned. The population will be all South Africans.
2. Identify if primary or secondary data are used: 2.1 Information obtained to determine the water levels of the major storage dams in South Africa will be primary/ secondary data. Where can this data be obtained from? 2.2 Describe how you will go about to determine the number of filling stations in your area using both a secondary source and a primary source. 2.3 Identify a problem where you can make use of secondary data. 2.4 Identify a problem where you can make use of primary data.	2.1 Secondary data and can be obtained from the Department of Water Affairs. They have collected the information for another purpose. 2.2 Secondary source: The local municipality will have a list of filling stations in your area. Primary data: You will walk or drive every street in your area and record where the filling stations are located. 2.3 Discuss options in class. 2.4 Discuss options in class.
3. For each of the following situations, would you recommend a sample or a census? Explain your reasoning in each case. 3.1 A jeweller just received a delivery of shock-resistant watches and wants to find out approximately from how high they can be dropped onto a concrete surface without breaking the crystal.	3.1 A sample, because of the destructive nature of the experiment.

3.2 Tiger Mills wants to learn the age, gender and income characteristics of persons who consume Cheerio breakfast cereal. 3.3 The producers of the Early Bird show want to find out what percentage of television (TV) viewers recognise a photo of host, Joseph Khumalo. 3.4 A researcher wants to determine whether companies that manufacture nuclear submarines might be interested in a new technique for purifying air when such craft are submerged.	3.2 A sample, because of the widely spread population. 3.3 A sample, because of the widely spread population. 3.4 Population, because it is a very small population.
4. Distinguish between a parameter and a statistic: 4.1 In surveying political choices of voters, a sample of voters is selected from the population of all eligible voters. Based on the results observed from the sample statistics the analyst proceeds to make inferences on the political choices likely to exist in the population of voters. 4.2 A recent survey of a sample of graduates reported that the average starting salary for a graduate is less than R30 000 per year. 4.3 Starting salaries for 270 graduates increased with 5% from the previous year. 4.4 In a random check of a sample of retail stores, the local health inspector found that 24% of the stores were not storing fish at the proper temperature. 4.5 In 2004, all major league soccer teams spent a total of R1 968 088 on players' salaries. Decide whether the numerical value is from a population or a sample and then specify it as parameter or a statistic.	4.1 Sample statistic is the proportion of the political choices of the sample. The inference that is made on the political choices for the population is the parameter. 4.2 The estimate starting salary for all graduates is a parameter. 4.3 Population parameter. 4.4 Sample statistic 4.5 This is a sample statistic.

5. Indicate whether the corresponding sets of observations would be quantitative or qualitative. If quantitative, distinguish between discrete and continuous.

	Quantitative	Qualitative	Discrete	Continuous
Ethnic group		X		
Age	X			X
ID number		X		
Net worth (rand)	X		X	
Favourite sport		X		
Temperature	X			X
Home language		X		
Cooking time for pasta	X			X
Speed of an aeroplane	X			X
Gender		X		

Activity 2

Questions:	Model Solutions:
Specify the measurement scale for each of the following:	
1. Whether you are a South African (SA) citizen.	Nominal
2. The amount you paid to fill up your petrol tank.	Ratio
3. The time it took you to get to the university this morning.	Ratio
4. The size of your take-away coffee.	Ordinal
5. Your belt size.	Ratio
6. Your student number.	Nominal
7. The occupation of 200 shoppers at a supermarket.	Nominal
8. The daily temperature measured inside the supermarket.	Ratio
9. The amount spent by every shopper at a supermarket.	Ratio
10. Rating a new product as good, average or poor.	Ordinal
11. The recording of the first three digits of the shoppers' cell phone numbers.	Nominal

Activity 3

Questions:	Model Solutions:
The South African government is concerned about the high illiteracy rates amongst adults in South Africa. They wish to estimate the true proportion of adults (over 18 years of age) in South Africa who are illiterate (that is, they cannot read or write in at least one language). A random sample of 10 000 adults were interviewed and 1 107 of them were found to be illiterate.	
1. The sample space is _____?	All adults in South Africa.
2. The parameter of interest is _____?	The proportion of adults in South Africa that is illiterate.

Questions:	Model Solutions:
3. The statistic is _____?	The proportion $\frac{1107}{10\,000}$
4. The sample size is _____?	10 000 adults
5. The population size is _____?	Unknown
6. The variable is _____?	Illiteracy rate
7. The measurement scale of the variable is _____?	Ratio scaled
8. The type of data is _____?	Quantitative and discrete

Revision Exercise 1

1. Classify the following sets of data as qualitative or quantitative. If quantitative state if it is discreet or continuous.

	Quantitative	Qualitative	Discrete	Continuous
The weight of each member of the soccer team	**X**			**X**
The religious affiliation		**X**		
Marks obtained in 1st test	**X**		**X**	
The colours you can identify in a rainbow		**X**		
Telephone numbers in a telephone directory		**X**		
The number of sit-ups you do	**X**		**X**	
The daily temperature at 12h00	**X**			**X**
Number of traffic fatalities	**X**		**X**	
Time required completing a crossword puzzle.	**X**			**X**
The ages of the learners in your study group	**X**			**X**

2. What is the measurement scale for the given variables?

Religious affiliation	**Nominal**
The number of sit-ups you do	**Ratio**
The daily temperature at 12h00	**Interval**
Number of traffic fatalities	**Ratio**
Time required to complete a crossword puzzle.	**Interval**
The ages of the learners in your study group	**Ratio**
The colours you can identify in a rainbow	**Nominal**
Telephone numbers in a telephone directory	**Nominal**
Shoe sizes	**Ordinal**
The three major professional tennis tournaments are listed: Australian Open; Wimbledon; US Open	**Nominal**
The amount of weight lost in the past month by a person following a strict diet	**Ratio**
The classification of Boeings as 727, 737 or 747 according to size.	**Ordinal**

3. A recent study on the average amount of time spent watching TV by a group of students yielded the following results:

Academic status	Number of students	Average time spent watching TV
1st years	30	6.4
2nd years	20	4.5
3rd years	10	2.8

Choose the correct word in each of the following statements:

The 60 students involved in the study constitute a sample/population of students.	Sample
The figure 6.4 is known as a parameter/statistic.	Statistic
The estimate of the average amount of time spent watching TV per week by all 1st year students involves a descriptive/ inferential technique.	Inferential
The variable of interest is the number of students/ average time watching TV	Average
The average times calculated for the three groups of students are of a discrete/ continuous nature.	Continuous
The academic status of the students can be classified as quantitative/ qualitative data	Qualitative

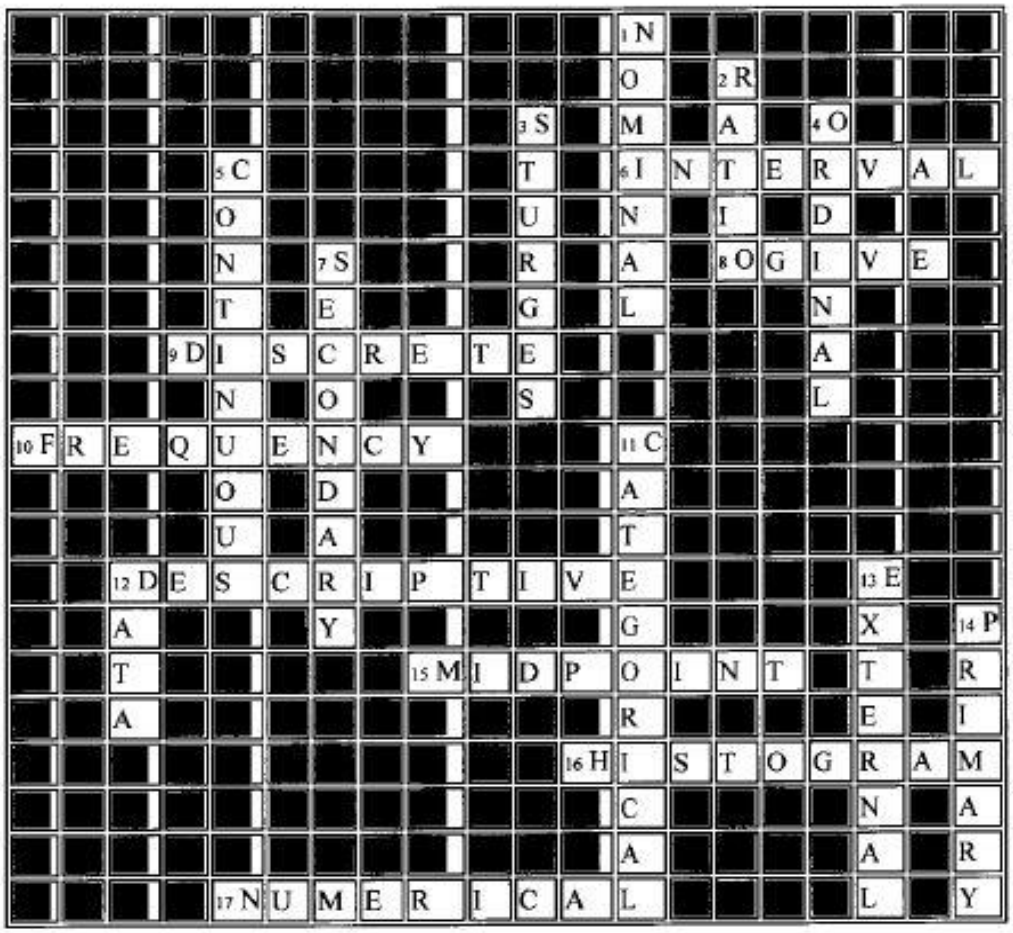
Revision Exercise 2

Questions:	**Model Solutions:**
1. The collection of all possible individuals, objects or measurements is known as: 1.1 A sample; 1.2 A population; 1.3 An inference; 1.4 A statistic.	1.2 A population;
2. Techniques used to organise, summarise and present the data that have been collected are known as: 2.1 Populations; 2.2 Samples; 2.3 Inferential statistics; 2.4 Descriptive statistics.	2.4 Descriptive statistics.
3. Techniques used to estimate a population parameter, based on a sample, are known as: 3.1 Populations; 3.2 Samples; 3.3 Inferential statistics; 3.4 Descriptive statistics.	3.3 Inferential statistics.

Questions:	Model Solutions:
4. In a random sample each item in the population has: 4.1 A chance of being selected; 4.2 The same chance of being selected; 4.3 A 50% chance of being selected; 4.4 No chance of being selected.	4.2 The same chance of being selected.
5. The sample mean is an example of a _____. 5.1 Sample statistic; 5.2 Population parameter; 5.3 Measurement scale; 5.4 Discrete variable.	5.1 Sample statistic.
6. Primary data are collected by: 6.1 Primary school children; 6.2 People doing it for the first time; 6.3 The actual people who will be using it; 6.4 Mainly experienced people.	6.3 The actual people who will be using it.
7. Secondary data are obtained from: 7.1 Secondary school children; 7.2 Existing sources; 7.3 The actual people who will be using it; 7.4 Mainly experienced people.	7.2 Existing sources.
8. When conducting a survey, you collect data by: 8.1 Sampling; 8.2 Using a secondary source; 8.3 Asking questions; 8.4 Using the random number table.	8.3 Asking questions.
9. To estimate the percentage of defects in a recent manufacturing batch, the quality control manager of Intel Computers selects every 8th chip that comes off the assembly line starting with the 3rd until he obtains a sample of 100. The method he follows to obtain the sample is known as the_____. 9.1 Simple random sampling method; 9.2 Systematic sampling method; 9.3 Stratified sampling method; 9.4 Snowball sampling method.	9.2 Systematic sampling method.

Questions:	Model Solutions:
10. Once every hour a random sample of 12 light bulbs is selected from an assembly line delivering this type of light bulb. The number of bulbs in each sample that will not light is divided by 12 to obtain the defective proportion. What is the variable? 10.1 Type of light bulbs; 10.2 Sample of 12 light bulbs; 10.3 Proportion of defective light bulbs; 10.4 Assembly line.	10.3 Proportion of defective light bulbs.
11. 49, 34, and 46 students are selected from the first year, second year and third year classes with 496, 348 and 481 students respectively. Identify the type of sampling used. 11.1 Stratified; 11.2 Cluster; 11.3 Systematic; 11.4 Simple random.	11.1 Stratified.
12. To avoid working late, a quality control analyst simply inspects the first 100 items produced in a day. This type of sampling used is known as: 12.1 Cluster sampling; 12.2 Systematic sampling; 12.3 Judgmental sampling; 12.4 Convenient sampling.	12.2 Systematic sampling.

Revision Exercise 3



Across

6. Data generated from rating scales and used in survey questionnaires
8. This graph represents the "less than cumulative frequency" totals
9. This data consists of whole numbers (eg. no. of cars parked in garage)
10. What does this symbol "fi" represent?
12. Type of Statistics to condense information into tabular and graphical methods (what you are doing at present)
15. What does this symbol "xi" represent?
16. A graph which doesn't have spaces between the bars
17. A quantitative data type where real numbers can be manipulated using arithmetic operations to produce meaningful results

Down

1. Form of data where categories are of equal importance
2. This is the strongest and most desirable data to gather
3. This formula has a name; $c = 1 + 3.3 \log (n)$
4. This form of data implies ranking, where categories are of unequal width
5. This data can consist of decimals (eg. distance travelled to VC)
7. Data collected and processed by others (getting information from the government for a research project you are working on)
11. Refers to data tupes representing categorieds of outcomes (and is qualitative)
12. Actual values recorded on a random variable
13. Data available from outside organisations
14. Data which is captured for the first time with a specific purpose in mind

Learning Unit 3: Ungrouped Data	
Material used for this Learning Unit: • Prescribed text pp.27—34; 63—65; 71—73; 76—82 and 84—86.	**My notes**
How to prepare for this Learning Unit: • Before the first class, be sure that you read pp.27–61 in the textbook. • As you read these sections, see if you can find the answers to the following questions: o Identify the tools you can use to summarise data. o Identify and interpret the graphs you can use for Quantitative Data. o Compute and interpret measures of location and dispersion. o Compute and interpret measures of location and dispersion.	

1 Introduction

The frequency table and ungrouped data graphs will be used in this learning unit to determine numerical summary measures.

2 Why and How Data are Summarised

After collecting data, it must be summarised, as we are interested in the overall picture rather than individual observations. Data are summarised first in a table format and then it can be presented in graphs.

3 Frequency Distribution Tables for Quantitative and Qualitative Data

Read pp.27–34 in the textbook.

- Data organised into a frequency distribution are called grouped data;
- Follow the steps in your textbook to construct the frequency distribution table for quantitative data.
-

A frequency distribution for qualitative data displays the possible categories along with the number of times (or frequency) each category appears in the data set.

The relative frequency for a particular category is the proportion or percentage of the observations within a category.

Graphs for qualitative data: Pie charts, simple bar charts, multiple bar charts and stacked bar charts can be used. Follow the steps in the textbook to construct these graphs and charts.
All graphs must be interpreted.

4 Identify the Features that Describe and Summarise Data Numerically

A complete understanding of the data can be attained by summarising the data numerically using sample statistics. Sample statistics are calculated and used in later units to estimate the population parameters.

Describe data using:

1. Measures of central tendency: arithmetic mean, median and mode;
2. Dispersion: standard deviation, variance and coefficient of variation;
3. Shape: using averages, histogram and Pearson's second coefficient of skewness.

5 Mean, Median and Mode for Ungrouped Data

- Different formulae are needed to do calculations for ungrouped data in a frequency distribution than for raw or ungrouped data;
- The arithmetic mean is the sum of all the values in a data set divided by the total number of observations in the set;
- This is one of the sample statistics that plays an important role in inferential statistics;
- The median of a data set is a position value and occupies the middle value when the original data values are arranged in a numerical order. It divides the bottom 50% of the data from the top 50%;
- The mode is that value which occurs most often in a data set, or the value with the highest frequency in the set.

5.1 Choosing Between the Mean, Median and Mode

The purpose of an average is that it should convey an impression of a distribution in a single figure. All the averages are different ideas with different uses. The factors that play a role in choosing the right average are the following:

- Is the nature of the data numerical or non-numerical?
  - The mode, which is the value that occurs most often, is the only measure of central tendency useful for qualitative (non-numerical) data that cannot be ranked in any way. If qualitative data sets can be ranked, the median can be used. The arithmetic mean involves arithmetic and is appropriate only for quantitative data sets.
- What does each average tell us?
  - Depending on the situation and the problem under investigation, one average may be superior to another, and in some other cases, you can use all three in conjunction;
  - The mode identifies the value that occurs more often than the others do. It may be a good choice if one value occurs much more often than others do. It is possible to have more than mode (multi modal), which will eliminate the mode as an option. Outliers or extreme values do not influence the mode at all;
  - The median indicates the centre of the distribution. The same number of observations lie above and below the median, regardless of how far above or how far below. This means that it is unlikely that outliers at either end of the distribution will affect the median. In a skewed distribution the median will be a better option than the mean;
  - The mean is the most frequently used average because it includes all the values in the data set. This feature makes it the most sensitive to extreme values (outliers).
- What is the shape of the distribution?
  - In a symmetrical distribution, the mean, median and mode will be the same or very close together. Whichever one you choose will give you the same answer;
  - In a skewed distribution, the mode stays at the peak of the distribution because outliers do not influence the mode at all. The median, being dependent on the number of values in the data set rather than on the

size of those values, is less sensitive than the mean. The influence of the outliers is the highest on the arithmetic mean and the arithmetic mean will move in the direction of the outlier. If the mean is very different from the median, the median will be a better option to use;

- In moderately skewed or asymmetrical distributions a very important relationship exists among these three measures of central tendency. In such distributions the distance between the mean and median is about one-third of the distance between the mean and mode. This relationship as can be expressed as:

Symmetrical distribution: Mean = median = mode

Positive skewed distribution: Mode < median < mean

Negatively skewed distribution: Mean < median < mode

6 Range, Variance, Standard Deviation and Coefficient of Variation for a Given Ungrouped Data Set

A measure of dispersion indicates the amount of variation in a data set. It is used to answer a question such as “Are the individual values close to the mean or scattered widely around the mean?”

Closely grouped data have relatively small dispersion values and more widely spread data have larger dispersion values. If all the data values are the same, the measure of dispersion is zero. There is no limit in how spread out data can be; therefore, the measure of dispersion can be very large. Variation in distributions is one of the factors that contribute to risk in business.

The simplest measure of variance or dispersion is the range. The range is determined by arranging the numbers in a data set in an array (highest to lowest or vice versa) and then subtracting the lowest value from the highest. Between these two numbers lie all the other numbers. The further apart the highest and lowest numbers, the wider the range and therefore the more the variation amongst the numbers. The closer the highest and lowest numbers are together, the narrower the range and the less the variation in numbers.

Higher levels of variance usually mean more uncertainty in interpretation, while lower levels reduce uncertainty.

Together with the arithmetic mean, the standard deviation is used extensively in inferential statistics. Most statistical theorems are based on distributions described by their mean and standard deviation because these descriptive measures include all the data in the set and the methods used are based on sound mathematical principles.

The standard deviation of a data set is expressed in the same measurement scale as the original data values.

For a single set of measurements, the mean can be combined with the standard deviation to obtain information about how values in a data set are distributed along the measurement scale. This information depends on the shape of the histogram or the skewness of the distribution.

The coefficient of variation is a refinement of the standard deviation as it eliminates some of the weaknesses of the standard deviation. Consider the following example:

An investor has to choose between two investment options, Option 1 has an estimated return of 15% with a standard deviation of 7% while Option 2 has an estimated return of 13% with a standard deviation of 4%. Which option should the investor choose?

Choosing Option 1 means that the actual return can vary by 15% ± 7% and Option 2 by 13% ± 4%. It is very difficult to make a choice because neither the returns nor standard deviations are equal.

The decision can be made easier by using the coefficient of variation. The coefficient of variation “spreads” the value of the standard deviation across the value of the mean.

With reference to the example above, the investor divides the standard deviation for Option 1 by its expected return, 5%/15% = 0.33. This value means that for every 1% of the expected return of Option 1, the risk (standard deviation) is 0.33%. For Option 2 the coefficient of variation is 0.31%. Option 2 will therefore be a marginally less risky investment because its risk (standard deviation) per each one per cent of return is less than that of Option 1.

7 Measures of Relative Standing: Quartiles and Percentiles for Ungrouped Data

The purpose of measures of relative standing (or measures of position) is to see where a value stands in relation to the other values in a set of data.

- The median, which is a measure of central tendency, is also a measure of relative standing. As you have learned previously, the median divides the data into two equal parts, the bottom 50% and the top 50%;
- It is often necessary to divide a data set into a larger number of parts. Quartiles (derived from the word “quarter”) and percentiles (derived from the “per centum” or “per hundred”) are two of the most popular measures of position;
- Quartiles measure the relative position of the data values by dividing the data set into four equal parts. The first quartile (Q_1) is a value such that 25% of the observations are smaller and the third quartile (Q_3) is a value such that 75% of the values are smaller;
- Percentiles divide the data into 100 equal parts and each percentile (Pj) is a value such that j% of the observations are smaller than Pj. (j can take on a value between 1% and 100%);
- We use these measures to describe the position a specific data value possesses in relation to the rest of the sample.

7.1 Dispersion Using the Interquartile Range

The interquartile range is a measures of dispersion linked to measures of location. It measures the spread of the middle 50% of the data set and is the difference between Q_3 and Q_1. This means that the first 25% and the last 25% of the data are cut off. Large values of this statistic indicate that the first and third quartiles are far apart, indicating a high level of variability. The interquartile range is used to identify outliers which should be discarded before data analysis.

7.2 Box-and-Whisker Plot

To get a quick summary of both centre and spread, use the five-number summary table and the box-and-whisker plot.

The five-number summary table divides the data set into four subsets, with approximately 25% of the observations in each quarter.

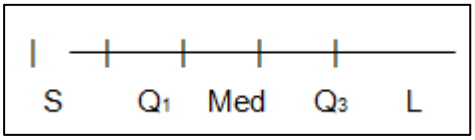


Diagram 4.1 The five-number summary table

Where, the following descriptive statistical measures are used to summarise the data:

- S: The smallest data value;
- Q_1: The lower quartile;
- The median, or Q_2;
- Q_3: The upper quartile;
- L : The largest data value.

The easiest way to develop a five-number summary is to arrange the data in a numerical order, and identify the smallest and largest values, the quartiles and the median.

The box-plot visually shows:

- The range of the data values by indicating the smallest and largest values;
- The inter-quartile range by indicating the first and third quartiles;
- The median to show where the data are centred;
- The degree of symmetry can be identified by inspection;
- The presence of outliers.

8 Recommended Additional Reading

You should engage extensively with your Recommended Additional Reading as it is an essential part of the learning experience. It will give you different perspectives and engagement opportunities with different authors and schools of thoughts. This will give you deeper insight into this discipline and has been designed to add value to the module.

Johnson, RR and Kuby, PJ. 2011. *Elementary statistics*. 11th edition. Pacific Grove, (CA): Duxbury.

Lombaard, C, Van der Merwe, L, Kele, T and Mouton, S. 2010. *Elementary statistics for business and economics*. 1st edition. Cape Town: Heinemann Pearson Publishing.

Sullivan, M. 2010. *Fundamentals of statistics*. 3rd edition. Boston, (MA): Pearson's Education.

Triola, MF. 2009. *Elementary statistics*. 11th edition. Boston, (MA): Pearson's Education.

Weaver, S. 2012. *STAT 141 - Basic statistical methods.* [Video online]. Available at: http://youtu.be/7Mfa1MbVhcc [Accessed 24 May 2014].

Willemse, I. 2009. *Statistical methods and calculation skills*. 3rd edition. Pretoria: Juta.

9 Recommended Digital Engagement and Activities

Your lecturer will provide you with the blog/ Facebook details for your class from which you can obtain information or give comments. Digital activities will also be available on the blog.

Access the following; this Internet project will help you develop skills to summarise data with statistics and also interpret such data.

Triola, MF. 2011 *Elementary statistics.* Pearson Education. [Online]. Available at: http://www.aw.com/triola [Accessed 24 May 2014].

10 Interactive Work Space

10.1 Izimvo Exchange 1

It is important to use the right type of average. All the averages are different ideas with different uses. Discuss which average will be the appropriate one to use in each study:

1. The test marks of five students are as follows:

55	59	66	66	94

2. Calculating the average house prices in a certain suburb in Gauteng.

3. The National Housing Department conducted a survey to estimate the average number of liveable square meters for low cost housing. The reported mean was 14.5 square meters and the median 12.2 square meters. Which measure of central tendency is more appropriate?

4. Number of children in 12 families:

3	2	2	2	3	2	1	2	2	2	2	2

5. Incomes in Rand of 10 families.

8 400	8 300	8 600	7 400	7 300	9 700	8 100	17 100	9 100	9 300

10.2 Izimvo Exchange 2

Bring your financial or scientific calculator with statistics functions to class. If you still have your calculator manual, bring that as well. Calculators are pre-programmed to perform certain statistical functions, such as the mean and standard deviation, to do routine calculations. Managing these skills will save you a lot of time doing your activities, tests and examinations.

The methods that follow; to determine the mean and standard deviation will be explained briefly on a basic Casio and Sharp calculator. Form groups in the class consisting of students with the same calculator and help one another to master the techniques.

1. Enter each data value by keying in the number and pressing the data key. DATA

2. Each time you enter a value, the calculator shows the count of the data entered so far.

3. Find the key on the calculator for working out the standard deviation or the mean. ($\bar{x}$; σ_x or s_x).

4. You will use the 2^{nd} function or colour coded alpha key together with the statistic you want to calculate to get the answer.

5. Once the data have been entered, you can use it to calculate the mean as well.

10.3 Izimvo Exchange 3

Visit your campus bookstore and record the prices of a random sample of 30 different academic text books. Describe the method you used to obtain your data and construct a frequency distribution as well as all the relevant graphs. Compare the graphs obtained with that of your class mates. Look at shape, range, peaks, etc.

10.4 Izimvo Exchange 4

Collect examples of graphs and charts from the printed media that you can use to interpret as you learn about them. Bring your examples to class so that there is a variety of examples that can be discussed in terms of the following:

- Do you think the graphs draws attention?
- Do you understand the purpose of the graphs?
- Is there any intention of misleading the readers?
- What variables and types of data are portrayed?

10.5 Activity 1

Purpose:
The purpose of this activity is to do a complete analysis of a case study given with ungrouped data.

Task:
The following data represent the number of daily and Sunday newspapers published in nine provinces during 2010.

6	18	7	9	24	12	6	19	11

1. Calculate the mean, median and modal number of newspapers. Explain each answer.

2. By comparing the mean, median and modal statistics, what can you conclude about the shape of the distribution?

3. Determine the range, standard deviation, variance and coefficient of variation for the set of newspapers.

4. What proportion of the newspapers will fall within one, two and three standard deviations from the mean?

5. Do a five-number summary table and a box-and-whisker plot. Comment on the skewness of the distribution and the presence of outliers.

6. Determine the inter-quartile range and quartile deviation. Interpret your answer.

7. Determine the middle 70% range.

8. What is the minimum number of newspapers a province must publish to fall within the top 15% of the provinces?

Commentary Related to Activity Design:
This activity is designed to serve as a check to see if you can perform all the possible descriptive numerical summaries for ungrouped data. The most important part of the analysis is whether you can interpret all your results.

10.6 Activity 2

Purpose:
Demonstrate your ability to choose an appropriate graph for the given data and follow the correct steps in constructing the graph.

Task:
Illustrate the following data by means of a bar chart. Analysis of costs (R'00) over four years for temporary workers hired to clean new building sites.

Year	Safety Equipment	Transport per Month	Lunch
2009	120	10	20
2010	140	20	20
2011	100	40	30
2012	110	30	50

Commentary Related to Activity Design:
It is possible to apply different bar graphs to portray the data. It all depends on what information you want from the graph. For example, do you want to compare total costs per year? Do you want to compare the different components?

10.7 Revision Exercise 1

1. A small company pays each of its five cleaners R22 000, two clerks R70 000 each and the manager R270 000. How many employees earn less than the mean salary?

2. Given a negatively skewed distribution with a median of 11 and a mode of 18, which of the following is a possible value for the mean and why?

28	19	10	12

3. For the data 2, 19, 29, 19, 100, 9, 90; which of the mean, median or mode would be changed if the 2 were changed to 29?

4. The mean of a set of five data points is 20. You have three of the data points; 5, 15, 30. The remaining two numbers are also the mode of the distribution, what are the two numbers?

5. For the last 10 days in January this year the Gautrain from Pretoria arrived late in Johannesburg by the following number of minutes. (A negative number means that the train was early by the number of minutes.)

-2	6	4	10	-4	12	2	-1	3	1

Determine the range and the mean of the data set.

6. A sample of students has a mean of 3.2 members in their family. The modal number of family members is two and the median number is 2.1. Based on this information what will the shape of the distribution probably be?

7. Earthquake intensities are measured using a device called a seismograph which is designed to be most sensitive for earthquakes with intensities between 4.0 and 9.00 on the open-ended Richter scale. Measurements of 18 recent earthquakes gave the following readings:

4.5	L	5.5	H	8.7	8.9	6.0	H	5.2
L	4.5	5.2	5.5	6	8.7	8.9	H	H

L indicates that the earthquake had an intensity of below 4.0 and H indicates that the earthquake had intensity above 9.0. Fifty per cent of the earthquake intensities were more than what value?

8. The number of rejects from 50 samples of the same size is as follows:

Number of Rejects in Sample	Number of Samples (Frequency of Rejects)
0	5
1	10
2	10
3	20
4	5

9.1 The arithmetic mean number of rejects per sample is _____?

9.2 Half of the samples have less than _____?

9.3 The modal number of rejects is _____?

9. A financial analyst's sample of six companies' book values (in R'000) were:

25	7	22	33	18	15

If the sample mean is R20 000, what is the sample standard deviation? (Round your answer to the closest Rand.)

10. A company has two regional head offices; in Cape Town and Johannesburg. Workers in Johannesburg claim that their salaries are more variable than the workers in Cape Town. To test their claim the following data was collected for a random sample of 100 workers in each office.

	Cape Town	**Johannesburg**
Mean Salary	R27 000	R25 000
Standard Deviation	R2 000	R2 100

Are the salaries in Johannesburg more variable? Use the coefficient of variation to prove your answer.

11. Random samples of small townhouse selling prices are obtained from ABSA Bank and First National Bank. The results followed normal distributions and are summarized below:

	ABSA	**First National**
Sample Size	50	80
Mean House Price	R150 000	R160 000
Standard Deviation	R20 000	R25 000

a. Which financial institutions' reported prices can be considered more uniform? Use the coefficient of variation to prove your answer.

b. With reference to ABSA, at least 68% of the selling prices were between which two values?

12. In a data set of 120 observations, how many observations lie between the 50^{th} percentile and the 60^{th} percentile?

10.8 Revision Exercise 3

1. A random sample of 40 smokers is classified according to age. The youngest person recorded was 10 years and the oldest 70 years. The median smoking age was 35 years,

25% of the smoking people were older than 45 and 25% were younger than 30. Draw a box-and-whisker plot to summarise the data and determine:

1.1 The shape of the age distribution.

1.2 The interquartile range.

1.3 50% of the smokers were older than what age?

1.4 Smokers younger than what age will be consider outliers?

1.5 The oldest 25% of the smokers are older than what age?

1.6 Are there any outliers present?

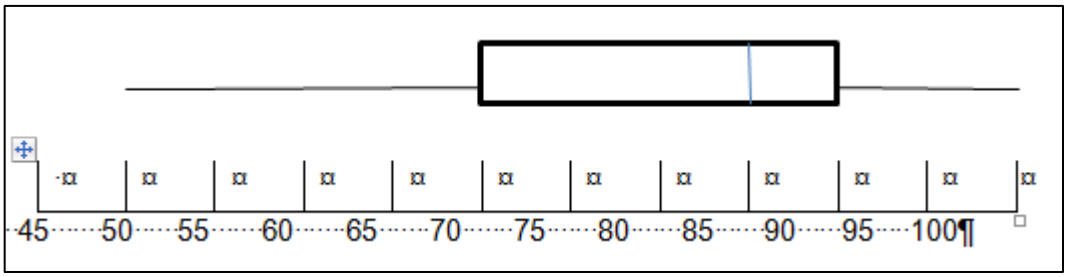
2. The inter quartile range for a data set is; 90–66. Decide which of the following data values would be classified as an outlier. 50; 140; 100?

3. The distances travelled each week (in kilometres) by 12 randomly selected sales representatives of an insurance company are as follows:

100 110 190 200 290 320 380 400 410 580 700 980

What will the five-number summary be?

4. The weights of a sample of female police officers are summarised in the following boxplot:



4.1 What proportion of the female police officers weighs between70kg and 90kg?

4.2 What proportion of the female police officers weigh more than 90kg?

4.3 Determine the median weight of the female police officers.

4.4 Determine the range of the weights.

4.5 What values can the mean possibly be? Why will this be a possibility?

4.6 Values in the original data set that can be considered "outliers" are smaller than or larger than what weights?

10.9 Revision Exercise 4

The following table reveals the income in (R'0 000) of a training facility for disabled persons:

Year	Income from the National Lottery	Subsidy	Total
2009	1.4	2.0	3.4
2010	1.8	2.7	4.5
2011	2.0	0.6	2.6
2012	2.0	1.7	3.7

Construct a multiple bar graph to portray the data. What are the main features you can see from your graphs?

10.10 Revision Exercise 5

The following data were obtained from a questionnaire completed by a sample of 25 people about how they get news: (N=newspaper T=television R=radio M=magazine)

N	T	N	R	N	T	N	R	N
R	T	M	R	M	M	N	M	
M	N	R	T	R	R	T	M	

Summarise the results in a frequency table and construct a pie chart. Interpret your results.

Solutions to Exercises

Izimvo Exchange 1

Questions:	**Model Solutions:**
Discuss which average will be the appropriate one to use in each study:	
1. The test marks of five students are as follows: 55 59 66 66 94	• The arithmetic mean of the marks is 68. This means that the sum of all the marks evenly divided by all the learners will give you 68; • The median value is 66, which means that half of the learners have less than 66 and the other half more than 66; • The mode is 66, which means that more learners obtained 66 than any other mark; • The 94 is an outlier, which distorts the mean, therefore the median is a better option.
2. Calculating the average house prices in a certain suburb in Gauteng.	In calculating the average house prices in a certain suburb in Gauteng, you will most probably make use of the median. This is because the relatively few homes with extremely high or low prices do not affect the median strongly. The median provides a better indication of the "typical" house price.
3. The National Housing Department conducted a survey to estimate the average number of liveable square meters for low cost housing. The reported mean was 14.5 square meters and the median 12.2 square meters. Which measure of central tendency is more appropriate?	Because the mean is higher than the median, we can assume that there is an outlier in the data set, which will make the median a better option to use.
4. Number of children in 12 families: 3 2 2 2 3 2 1 2 2 2 2 2	• Mean = 2.1; • Median = 2; • Mode = 2; • The averages are either the same or very close to one another, therefore any of the averages will be good to use, depending only on the purpose of the calculation.

5. Incomes in Rand of 10 families. 8 400 8 300 8 600 7 400 7 300 9 700 00 100 00 00	• Mean = R9 330; • Median = R8 500; • Mode = No mode; • There is no mode, and an outlier present which causes the distribution to be skewed, therefore use the median.

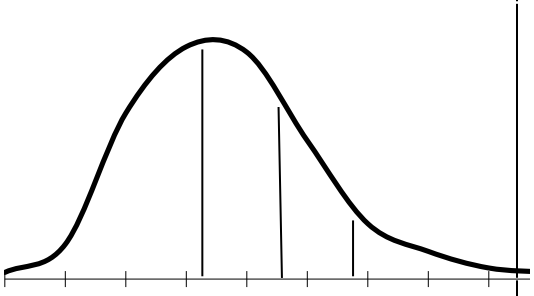
Izimvo Exchange 3

Questions:	**Model Solutions:**
Visit your campus bookstore and record the prices of a random sample of 30 different academic text books. Describe the method you used to obtain your data and construct a frequency distribution as well as all the relevant graphs. Compare the graphs obtained with that of your class mates. Look at shape, range, peaks, etc.	Each student will have a different sample selection. The students would have used different sampling methods. Let them discuss why they have chosen a specific method and how did they go about to select the sample. The class can discuss if the method they followed can be considered a random sampling method.

Activity 1

Questions:	**Model Solutions:**
The following data represent the number of daily and Sunday newspapers published in nine provinces during 2010. 6 18 7 9 24 12 6 19 11	
1. Calculate the mean, median and modal number of newspapers. Explain each answer.	• If you take all the provinces into account, the average number of newspapers published is 12.44; $\bar{x} = \frac{112}{9} = 12.44$ • The median position is $\frac{9+1}{2} = \text{value nr } 5$. Note: The numbers must be in numerical order; • The median = 11. Half of the provinces publish less than 11 different newspapers, the other half more than 11 newspapers; • The mode = 6. More of the provinces publish 6 different newspapers than any other number.

<table>
<tr>
<td>2. By comparing the mean, median and modal statistics, what can you conclude about the shape of the distribution?</td>
<td>Mean = 12.44; Median = 11; Mode = 6

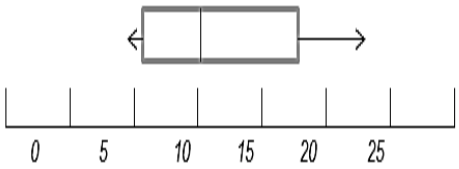
Distribution of Daily and Sunday Newspapers 2010
 
mode< med<mean

The distribution is positively skewed because mode < median < mean.</td>
</tr>
<tr>
<td>3. Determine the range, standard deviation, variance and coefficient of variation for the set of newspapers.</td>
<td>• Range = Largest number – smallest number
= 24 – 6 = 18
• Standard deviation:

<table>
<tr><td>X</td><td>6</td><td>6</td><td>7</td><td>9</td><td>11</td><td>12</td><td>18</td><td>19</td><td>24</td></tr>
<tr><td>$(x - \bar{x})^2$</td><td>41.47</td><td>41.47</td><td>29.59</td><td>11.83</td><td>2.07</td><td>0.19</td><td>30.91</td><td>43.03</td><td>133.63</td></tr>
</table>

$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}} = \sqrt{\frac{322.88}{8}} = \sqrt{40.36} =$ 5.17 newspapers

• Variance (s^2) = 5.17
= 26. 73 newspapers
• Coefficient of variation = 5.17/ 12.44 × 100%
= 41.56%</td>
</tr>
</table>

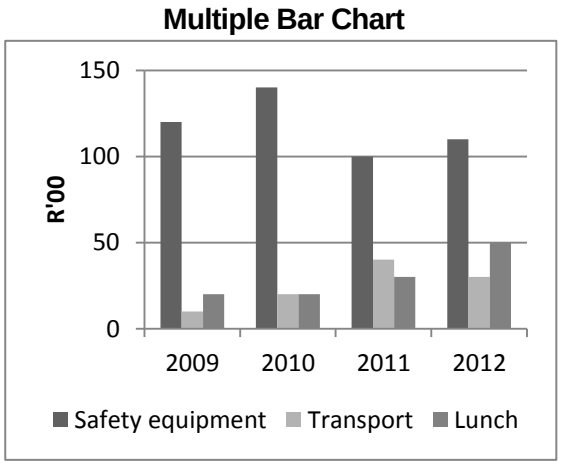
4.	What proportion of the newspapers will fall within one, two and three standard deviations from the mean?	• Within two standard deviations: = 12.44 – 2(5.17) = 2.1 and 12.44 + 2(5.17) = 22.78 At least 75% of the provinces publish between 2.1 and 22.78 newspapers. • Within three standard deviations: = 12.44 – 3(5.17) = -3.07 and 12.44 + 3(5.17) = 27.95 At least 89% of the provinces publish between -3.07 and 27.95 newspapers.
5.	Do a five-number summary table and a box-and-whisker plot. Comment on the skewness of the distribution and the presence of outliers.	• S = 6; • Q_1 = 6.5; • Median = 11; • Q_3 = 18.5; • L = 24.  • Outliers are outside the inner fences; • Inner fence to the left = 6.5 - 1.5(18.5 – 6.5) = -11.5; • Inner fence to the right = 18.5 + 1.5(18.5 – 6.5) = 36.5; • The smallest value as well as the largest value fall within the inner fences, therefore no outliers present.
6.	Determine the inter-quartile range and quartile deviation. Interpret your answer.	• The quartile deviation = $\frac{Q3 - Q1}{2} = \frac{12}{2} = 6$; • The average deviation around the median is 6 newspapers.

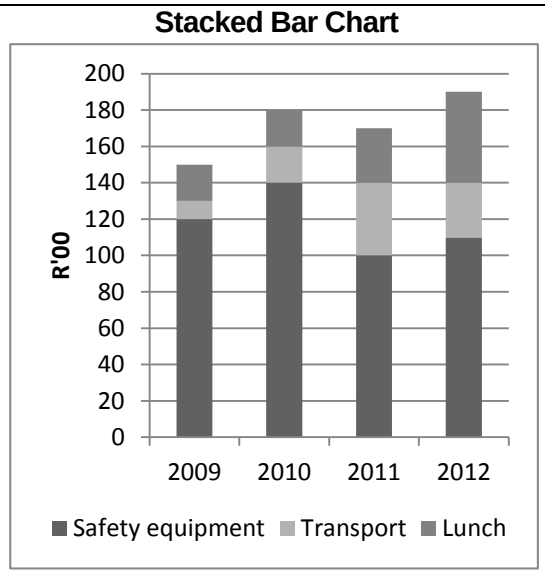
7. Determine the middle 70% range.	Middle 70% range $= P_{85} - P_{15}$ $= 21.5 – 6$ $= 15.5$ Position P_{85} $= 8.5$ P_{85} $= 21.5.$ That means that 85% of the provinces publish less than 21.5 newspapers. Position P_{15} $= 1.5$ P_{15} $= 6.$ That means that 15% of the provinces publish less than 6 newspapers.
8. What is the minimum number of newspapers a province must publish to fall within the top 15% of the provinces?	That will be more than 85% or P_{85}. $P_{85} = 21.5$. That means that the 15% of the provinces that publish the most newspapers publish more than 21.5 newspapers.

Activity 2

Questions:	**Model Solutions:**
Illustrate the following data by means of a bar chart. Analysis of costs (R'00) over four years for temporary workers hired to clean new building sites.	These graphs can easily be constructed using the graph facility in MSWord.
	This is compound information and both a multiple or stacked bar chart will be suitable.

Year	Safety Equipment	Transport per Month	Lunch
2009	120	10	20
2010	140	20	20
2011	100	40	30
2012	110	30	50





Revision Exercise 1

Questions:	Model Solutions:
1. A small company pays each of its five cleaners R22 000, two clerks R70 000 each and the manager R270 000. How many employees earn less than the mean salary?	$\bar{X}$ = (5 × 22 000) + (2 × 70 000) + 270 000 = 520 000 ÷ 8 = 65 000 This distribution will be skewed to the right because of the outlier to the right.
2. Given a negatively skewed distribution with a median of 11 and a mode of 18, which of the following is a possible value for the mean and why? 28 19 10 12	If a distribution is negatively skewed the mean must be smaller than the median, therefore 10.
3. For the data 2, 19, 29, 19, 100, 9, 90; which of the mean, median or mode would be changed if the 2 were changed to 29?	The mean will increase, the median will remain unchanged and there will be two modes; 19 and 29.
4. The mean of a set of five data points is 20. You have three of the data points; 5, 15, 30. The remaining two numbers are also the mode of the distribution, what are the two numbers?	Total = 5 × 20 = 100 5 + 15 + 30 = 50 ∴100 – 50 = 50 The remaining two data points must be the same to obtain a value that occurs more often: Mode = 25.

<table>
<tr>
<td>5. Each prospective employee who applies for a job at a certain bank is given a test. A sample of the length of time it took 42 applicants to write the test was organised into the following frequency distribution table:
<table>
<tr><th>Length of Time (min)</th><th>Number of Applicants</th></tr>
<tr><td>1 -< 4</td><td>4</td></tr>
<tr><td>4 -< 7</td><td>8</td></tr>
<tr><td>7 -< 10</td><td>14</td></tr>
<tr><td>10 -< 13</td><td>9</td></tr>
<tr><td>13-< 16</td><td>5</td></tr>
<tr><td>16-< 19</td><td>2</td></tr>
<tr><td></td><td></td></tr>
</table>
1.1 What is the mean time it took the applicants to write the test?
1.2 Determine median time to write the test.
1.3 Determine the time most applicants took to write the test.</td>
<td>5.1 Mean = $\frac{379}{42} = 9.02 \text{ minutes}$

5.2 Median = 7 + $\frac{(21-12)3}{14}$ = 8.93 minutes. Half of the candidates took shorter than 8.64 minutes to write the test.

5.3 $\bar{x} = \frac{379}{42} = 9.14 \text{ minutes}$</td>
</tr>
<tr>
<td>6. For the last 10 days in January this year the Gautrain from Pretoria arrived late in Johannesburg by the following number of minutes. (A negative number means that the train was early by the number of minutes.)
-2 6 4 10 -4 12 2 -1 3 1
Determine the range and the mean of the data set.</td>
<td>Range = 12 – (-2)
= 14
$\bar{x} = 3.1 \text{ minutes}$ (Use calculator)</td>
</tr>
<tr>
<td>7. A sample of students has a mean of 3.2 members in their family. The modal number of family members is two and the median number is 2.1. Based on this information what will the shape of the distribution probably be?</td>
<td>Positively skewed because mode < median < mean.</td>
</tr>
</table>

8. Earthquake intensities are measured using a device called a seismograph which is designed to be most sensitive for earthquakes with intensities between 4.0 and 9.00 on the open-ended Richter scale. Measurements of 18 recent earthquakes gave the following readings: 4.5 L 5.5 H 8.7 8.9 6.0 H 5.2 L 4.5 5.2 5.5 6 8.7 8.9 H H L indicates that the earthquake had an intensity of below 4.0 and H indicates that the earthquake had intensity above 9.0. Fifty per cent of the earthquake intensities were more than what value?	6 on the Richter scale. That is the median value.
9. The number of rejects from 50 samples of the same size is as follows: Number of Rejects in Sample: 0, 1, 2, 3, 4 Number of Samples (Frequency of Rejects): 5, 10, 10, 20, 5 9.1 The arithmetic mean number of rejects per sample is _____? 9.2 Half of the samples have less than _____? 9.3 The modal number of rejects is _____?	9.1 $\frac{110}{50}$ = 2.2 rejects 9.2 2.5 rejects 9.3 3
10. A financial analyst's sample of six companies' book values (in R'000) were: 25 7 22 33 18 15 If the sample mean is R20 000, what is the sample standard deviation? (Round your answer to the closest Rand.)	$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\frac{396}{6-1}}$ = 8.89 ≈ R8 899

11. A company has two regional head offices; in Cape Town and Johannesburg. Workers in Johannesburg claim that their salaries are more variable than the workers in Cape Town. To test their claim the following data was collected for a random sample of 100 workers in each office.

	Cape Town	**Johannesburg**
Mean Salary	R27 000	R25 000
Standard Deviation	R2 000	R2 100

Are the salaries in Johannesburg more variable? Use the coefficient of variation to prove your answer.

CV(CT) = $\frac{2000}{27000} \times 100 = 7.41\%$

CV(Jhb) = $\frac{2100}{25000} \times 100 = 8.4\%$

Yes, salaries are more variable in Johannesburg because the CV is higher.

12. Random samples of small townhouse selling prices are obtained from ABSA Bank and First National Bank. The results followed normal distributions and are summarised below:

	ABSA	**First National**
Sample Size	50	80
Mean House Price	R150 000	R160 000
Standard Deviation	R20 000	R25 000

12.1 Which financial institutions' reported prices can be considered more uniform? Use the coefficient of variation to prove your answer.

12.2 With reference to ABSA, at least 68% of the selling prices were between which two values?

12.1 CV (ABSA) = $\frac{20000}{150000} \times 100 = 13.33\%$

CV (First National) = $\frac{25000}{160000} \times 100 = 15.62\%$

ABSA's reported prices are more uniform.

12.2 According to the Empirical rule 68% of the prices will fall within 1 standard deviation from the mean; that means between

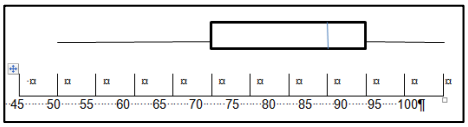
(R150 000 – 1(R25 000) = R125 000

(R150 000 + 1(R25 000) = R175 000

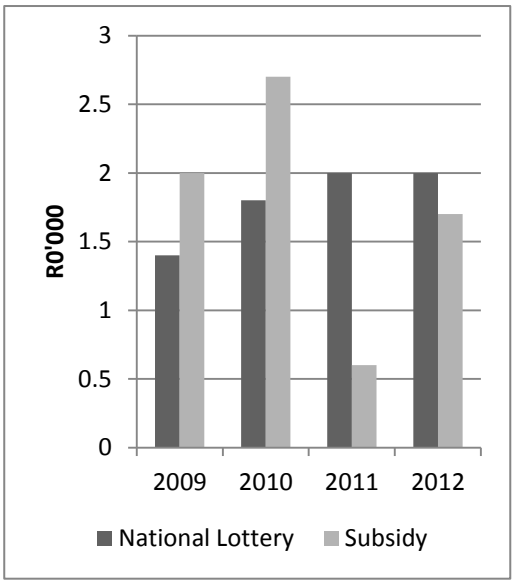
13. In a data set of 120 observations, how many observations lie between the 50th percentile and the 60th percentile?	P_{50} is value number 60.5 P_{60} is value number 72.6 The difference between the positions is 12 values.

Revision Exercise 2

Questions:	**Model Solutions:**
1. A random sample of 40 smokers is classified according to age. The youngest person recorded was 10 years and the oldest 70 years. The median smoking age was 35 years, 25% of the smoking people were older than 45 and 25% were younger than 30. Draw a box-and-whisker plot to summarise the data and determine: 1.1 The shape of the age distribution. 1.2 The interquartile range. 1.3 50% of the smokers were older than what age? 1.4 Smokers younger than what age will be consider outliers? 1.5 The oldest 25% of the smokers are older than what age? 1.6 Are there any outliers present?	1.1 The distribution is positively skewed. 1.2 Inter-quartile range is (45 – 30) = 15 1.3 50% of the smokers are older than 35 year. 1.4 Smokers younger than 7.5 years. 1.5 The oldest 25% are older than 35 years. 1.6 Left inner fence is 30 – 1.5(45 – 30) = 7.5 Right inner fence is 45 + 1.5(45 – 30) = 67.5 No outlier to the left, but the largest value is an outlier.
2. The inter quartile range for a data set is; 90–66. Decide which of the following data values would be classified as an outlier. 50; 140; 100?	Left inner fence = 66 – 1.5(90 – 66) = 30 Right inner fence = 90 + 1.5(90 – 66) = 126 ∴ 149 can be considered an outlier.
3. The distances travelled each week (in kilometres) by 12 randomly selected sales representatives of an insurance company are as follows: 100 110 190 200 290 320 380 400 410 580 700 980 What will the five-number summary be?	S = 100 Q_1 pos. =3.25 Q_1 value = 190 + 0.25(10) =192.5 Median = 350 Q_3 pos = 9.75 Q_3 value = 410 + 0.75(170) = 537.5 L = 980

Questions:	Model Solutions:
4. The weights of a sample of female police officers are summarised in the following boxplot:  4.1 What proportion of the female police officers weighs between70kg and 90kg? 4.2 What proportion of the female police officers weigh more than 90kg? 4.3 Determine the median weight of the female police officers. 4.4 Determine the range of the weights. 4.5 What values can the mean possibly be? Why will this be a possibility? 4.6 Values in the original data set that can be considered "outliers" are smaller than or larger than what weights?	4.1 50% because that is the interquartile range. 4.2 25%, because 90kg is Q_3. 4.3 85kg 4.4 100 – 50 = 50kg 4.5 The mean must be smaller than 85 because the distribution is negatively skewed. 4.6 Left inner fence = 70 – 1.5(20) = 40kg. That means weights lower than 40kg will be considered outliers. Right inner fence = 90 + 1.5(20) = 120kg. That means weights more than 120kg can be considered outliers.

Revision Exercise 3

Questions:	Model Solutions:
The following table reveals the income in (R'0 000) of a training facility for disabled persons: **Year** | **Income from the National Lottery** | **Subsidy** | **Total** 2009 | 1.4 | 2.0 | 3.4 2010 | 1.8 | 2.7 | 4.5 2011 | 2.0 | 0.6 | 2.6 2012 | 2.0 | 1.7 | 3.7 Construct a multiple bar graph to portray the data. What are the main features you can see from your graphs?	**Multiple Bar Graph** 

Revision Exercise 4

Questions:

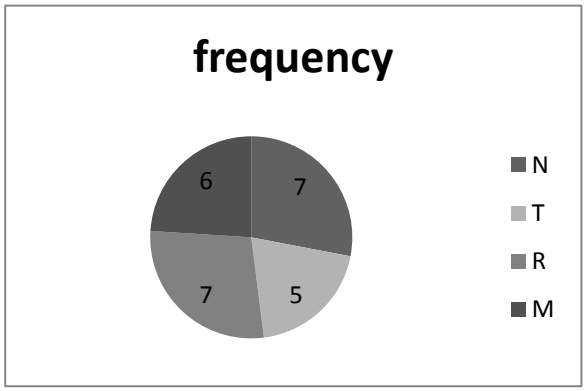
The following data were obtained from a questionnaire completed by a sample of 25 people about how they get news: (N=newspaper T=television R=radio M=magazine)

N	T	N	R	N	T	N	R	N
R	T	M	R	M	M	N	M	
M	N	R	T	R	R	T	M	

Summarise the results in a frequency table and construct a pie chart. Interpret your results.

Model Solutions:

		f
N	~~1111~~ 11	7
T	~~1111~~	5
R	~~1111~~ 11	7
M	~~1111~~ 1	6
		25



Learning Unit 4: Grouped Data

Material used for this Learning Unit:	**My notes**
• Prescribed text pp.35—42; 66—69; 74—82.	
How to prepare for this Learning Unit: • Before the first class, be sure that you read pp.63–97 in the prescribed textbook. • As you read these sections, see if you can find the answers to the following questions: ○ Identify the Features that Describe and summarise Data Numerically. ○ Determine the mean, mode and median for grouped and ungrouped data ○ Identify and determine the variance, range, standard deviation and coefficient of variation for a given ungrouped and grouped data set. ○ What is the spread of values around the mean in a Symmetrical Distribution using the empirical rule? ○ What are the measures of relative standing?	

1 Introduction

The frequency table and ungrouped data graphs will be used in this learning unit to determine numerical summary measures.

2 Why and How Data are Summarised

After collecting data, it must be summarised, as we are interested in the overall picture rather than individual observations. Data are summarised first in a table format and then it can be presented in graphs.

3 Frequency Distribution Tables for Quantitative and Qualitative Data

Read pp.35—42 in the textbook.

A frequency distribution for quantitative data is a summary of numerical data prepared by dividing raw data into several non-overlapping class intervals and then counting how many observations (frequency) of the variable fall into each class.

Working with grouped data inevitably means losing some information of the original data set. You will know how many values fall within each class, but not what the exact values of those observations were. The result of this is that all your summary statistics are estimated values.

- Data organised into a frequency distribution are called grouped data;
- Follow the steps in your textbook to construct the frequency distribution table for quantitative data.

A frequency distribution for qualitative data displays the possible categories along with the number of times (or frequency) each category appears in the data set.

The relative frequency for a particular category is the proportion or percentage of the observations within a category.

4 Frequency Distribution Graphs for Quantitative Data

Read pp.35–42 in your textbook.

The information in a frequency table is more easily understood and informative if captured in a graph. Follow the steps in your textbook to construct a histogram, polygon and ogive.

4.1 Histogram

A histogram is a continuous series of rectangles of equal width but different heights drawn to display the class frequencies. To draw a histogram we use data arranged into a grouped frequency table.

To interpret a histogram look for:

- The overall pattern and obvious deviations from this pattern;
- The overall pattern can be described by its shape, centre and spread;
- Location and number of peaks;
- Presence of gaps and outliers.

4.2 Polygon

A polygon is a line graph that plots the frequency for each class interval against the midpoint of the class interval. It can be used to portray the shape of the distribution in terms of the highs, lows and point of concentration. It provides similar information to the histogram. These graphs allow us to directly compare two or more frequency distributions.

4.3 Ogive

An ogive or cumulative curve is a smooth curve that can be used to estimate how many of the observations lie below or above a given value, rather than merely recording the number of items within intervals. This graph allows you to approximate answers to cumulative type questions to be found for values other than the boundaries of the class intervals.

5 Identify the Features that Describe and Summarise Data Numerically

A complete understanding of the data can be attained by summarising the data numerically using sample statistics. Sample statistics are calculated and used in later units to estimate the population parameters.

Describe data using:

4. Measures of central tendency: arithmetic mean, median and mode;
5. Dispersion: standard deviation, variance and coefficient of variation;
6. Shape: using averages, histogram and Pearson's second coefficient of skewness.

6 Mean, Median and Mode for Grouped Data.

- Different formulae are needed to do calculations for data grouped into a frequency distribution than for raw or ungrouped data;
- The arithmetic mean is the sum of all the values in a data set divided by the total number of observations in the set;
- This is one of the sample statistics that plays an important role in inferential statistics;
- The median of a data set is a position value and occupies the middle value when the original data values are arranged in a numerical order. It divides the bottom 50% of the data from the top 50%;
- The mode is that value which occurs most often in a data set, or the value with the highest frequency in the set.

6.1 Choosing Between the Mean, Median and Mode

The purpose of an average is that it should convey an impression of a distribution in a single figure. All the averages are different ideas with different uses. The factors that play a role in choosing the right average are the following:

- Is the nature of the data numerical or non-numerical?
    - The mode, which is the value that occurs most often, is the only measure of central tendency useful for qualitative (non-numerical) data that cannot be ranked in any way. If qualitative data sets can be ranked, the median can be used. The arithmetic mean involves arithmetic and is appropriate only for quantitative data sets.
- What does each average tell us?
    - Depending on the situation and the problem under investigation, one average may be superior to another, and in some other cases, you can use all three in conjunction;
    - The mode identifies the value that occurs more often than the others do. It may be a good choice if one value occurs much more often than others do. It is possible to have more than mode (multi modal), which will eliminate the mode as an option. Outliers or extreme values do not influence the mode at all;
    - The median indicates the centre of the distribution. The same number of observations lie above and below the median, regardless of how far above or how far below. This means that it is unlikely that outliers at either end of the distribution will affect the median. In a skewed distribution the median will be a better option than the mean;
    - The mean is the most frequently used average because it includes all the values in the data set. This feature makes it the most sensitive to extreme values (outliers).
- What is the shape of the distribution?
    - In a symmetrical distribution, the mean, median and mode will be the same or very close together. Whichever one you choose will give you the same answer;
    - In a skewed distribution, the mode stays at the peak of the distribution because outliers do not influence the mode at all. The median, being dependent on the number of values in the data set rather than on the size of those values, is less sensitive than the mean.

The influence of the outliers is the highest on the arithmetic mean and the arithmetic mean will move in the direction of the outlier. If the mean is very different from the median, the median will be a better option to use;

- In moderately skewed or asymmetrical distributions a very important relationship exists among these three measures of central tendency. In such distributions the distance between the mean and median is about one-third of the distance between the mean and mode. This relationship as can be expressed as:

Symmetrical distribution: Mean = median = mode

Positive skewed distribution: Mode < median < mean

Negatively skewed distribution: Mean < median < mode

7 Range, Variance, Standard Deviation and Coefficient of Variation for a Given Grouped Data Set

A measure of dispersion indicates the amount of variation in a data set. It is used to answer a question such as “Are the individual values close to the mean or scattered widely around the mean?”

Closely grouped data have relatively small dispersion values and more widely spread data have larger dispersion values. If all the data values are the same, the measure of dispersion is zero. There is no limit in how spread out data can be; therefore, the measure of dispersion can be very large. Variation in distributions is one of the factors that contribute to risk in business.

The simplest measure of variance or dispersion is the range. The range is determined by arranging the numbers in a data set in an array (highest to lowest or vice versa) and then subtracting the lowest value from the highest. Between these two numbers lie all the other numbers. The further apart the highest and lowest numbers, the wider the range and therefore the more the variation amongst the numbers. The closer the highest and lowest numbers are together, the narrower the range and the less the variation in numbers.

Higher levels of variance usually mean more uncertainty in interpretation, while lower levels reduce uncertainty.

Together with the arithmetic mean, the standard deviation is used extensively in inferential statistics. Most statistical theorems are based on distributions described by their mean and standard deviation because these descriptive measures include all the data in the set and the methods used are based on sound mathematical principles.

The standard deviation of a data set is expressed in the same measurement scale as the original data values.

For a single set of measurements, the mean can be combined with the standard deviation to obtain information about how values in a data set are distributed along the measurement scale. This information depends on the shape of the histogram or the skewness of the distribution.

The coefficient of variation is a refinement of the standard deviation as it eliminates some of the weaknesses of the standard deviation. Consider the following example:

An investor has to choose between two investment options, Option 1 has an estimated return of 15% with a standard deviation of 7% while Option 2 has an estimated return of 13% with a standard deviation of 4%. Which option should the investor choose?

Choosing Option 1 means that the actual return can vary by 15% ± 7% and Option 2 by 13% ± 4%. It is very difficult to make a choice because neither the returns nor standard deviations are equal.

The decision can be made easier by using the coefficient of variation. The coefficient of variation “spreads” the value of the standard deviation across the value of the mean.

With reference to the example above, the investor divides the standard deviation for Option 1 by its expected return, 5%/15% = 0.33. This value means that for every 1% of the expected return of Option 1, the risk (standard deviation) is 0.33%. For Option 2 the coefficient of variation is 0.31%. Option 2 will therefore be a marginally less risky investment because its risk (standard deviation) per each one per cent of return is less than that of Option 1.

8 Spread of Values around the Mean in a Symmetrical Distribution using the Empirical Rule

If the histogram is symmetrical or normal, you can make a statement about the proportion of data values that fall into various intervals using the Empirical rule:

- Approximately 68,26% of all observations fall within one standard deviation from the mean;
- Approximately 95,44% of all observations fall within two standard deviations from the mean;
- Approximately 99.73% of all observations fall within three standard deviations from the mean.

8.1.1 Shape of the Distribution

You can describe the shape of a distribution by its symmetry or lack of it (skewness). It relates to the shape of the histogram or polygon that you can draw from the data and the locations of the mean, median and mode in the data set.

9 Measures of Relative Standing: Quartiles and Percentiles for Grouped Data

The purpose of measures of relative standing (or measures of position) is to see where a value stands in relation to the other values in a set of data.

- The median, which is a measure of central tendency, is also a measure of relative standing. As you have learned previously, the median divides the data into two equal parts, the bottom 50% and the top 50%;
- It is often necessary to divide a data set into a larger number of parts. Quartiles (derived from the word “quarter”) and percentiles (derived from the “per centum” or “per hundred”) are two of the most popular measures of position;
- Quartiles measure the relative position of the data values by dividing the data set into four equal parts. The first quartile (Q_1) is a value such that 25% of the observations are smaller and the third quartile (Q_3) is a value such that 75% of the values are smaller;
- Percentiles divide the data into 100 equal parts and each percentile (Pj) is a value such that j% of the observations are smaller than Pj. (j can take on a value between 1% and 100%);

- We use these measures to describe the position a specific data value possesses in relation to the rest of the sample.

9.1 Dispersion Using the Interquartile Range

The interquartile range is a measures of dispersion linked to measures of location. It measures the spread of the middle 50% of the data set and is the difference between Q_3 and Q_1. This means that the first 25% and the last 25% of the data are cut off. Large values of this statistic indicate that the first and third quartiles are far apart, indicating a high level of variability. The interquartile range is used to identify outliers which should be discarded before data analysis.

10 Recommended Additional Reading

You should engage extensively with your Recommended Additional Reading as it is an essential part of the learning experience. It will give you different perspectives and engagement opportunities with different authors and schools of thoughts. This will give you deeper insight into this discipline and has been designed to add value to the module.

Johnson, RR and Kuby, PJ. 2011. *Elementary statistics*. 11th edition. Pacific Grove, (CA): Duxbury.

Lombaard, C, Van der Merwe, L, Kele, T and Mouton, S. 2010. *Elementary statistics for business and economics*. 1st edition. Cape Town: Heinemann Pearson Publishing.

Sullivan, M. 2010. *Fundamentals of statistics*. 3rd edition. Boston, (MA): Pearson's Education.

Triola, MF. 2009. *Elementary statistics*. 11th edition. Boston, (MA): Pearson's Education.

Weaver, S. 2012. *STAT 141 - Basic statistical methods.* [Video online]. Available at: http://youtu.be/7Mfa1MbVhcc [Accessed 24 May 2014].

Willemse, I. 2009. *Statistical methods and calculation skills*. 3rd edition. Pretoria: Juta.

11 Recommended Digital Engagement and Activities

Your lecturer will provide you with the blog/ Facebook details for your class from which you can obtain information or give comments. Digital activities will also be available on the blog.

Access the following; this Internet project will help you develop skills to summarise data with statistics and also interpret such data.

Triola, MF. 2011 *Elementary statistics.* Pearson Education. [Online]. Available at: http://www.aw.com/triola [Accessed 24 May 2014].

12 Interactive Work Space

12.1 Izimvo Exchange 1

Visit your campus bookstore and record the prices of a random sample of 30 different academic text books. Describe the method you used to obtain your data and construct a frequency distribution as well as all the relevant graphs. Compare the graphs obtained with that of your class mates. Look at shape, range, peaks, etc.

12.2 Izimvo Exchange 2

Collect examples of graphs and charts from the printed media that you can use to interpret as you learn about them. Bring your examples to class so that there is a variety of examples that can be discussed in terms of the following:

- Do you think the graphs draws attention?
- Do you understand the purpose of the graphs?
- Is there any intention of misleading the readers?
- What variables and types of data are portrayed?

12.3 Activity 1

Purpose:
The purpose of this activity is to manipulate the frequency table to some other formats in order to obtain certain information from it. It also introduces you to basic decision making.

Task:
The following relative frequency distribution resulted from a study of the time (in minutes) employees of an organisation arrive late for work during a specific month:

Minutes Late	% of Employees
$1.00 \leq x < 6$	1
$6.00 \leq x < 11$	3
$11.00 \leq x < 16$	4
$16.00 \leq x < 21$	6
$21.00 \leq x < 26$	7
$26.00 \leq x < 31$	9
$31.00 \leq x < 36$	11
$36.00 \leq x < 41$	19
$41.00 \leq x < 46$	32
$46.00 \leq x < 51$	8

1. Determine the class midpoints for each interval and explain the meaning of the midpoints in two of the intervals.

2. Construct the cumulative frequencies for each interval.

3. If the study covers a sample of 3 000 employees, how many of them fall into each class interval?

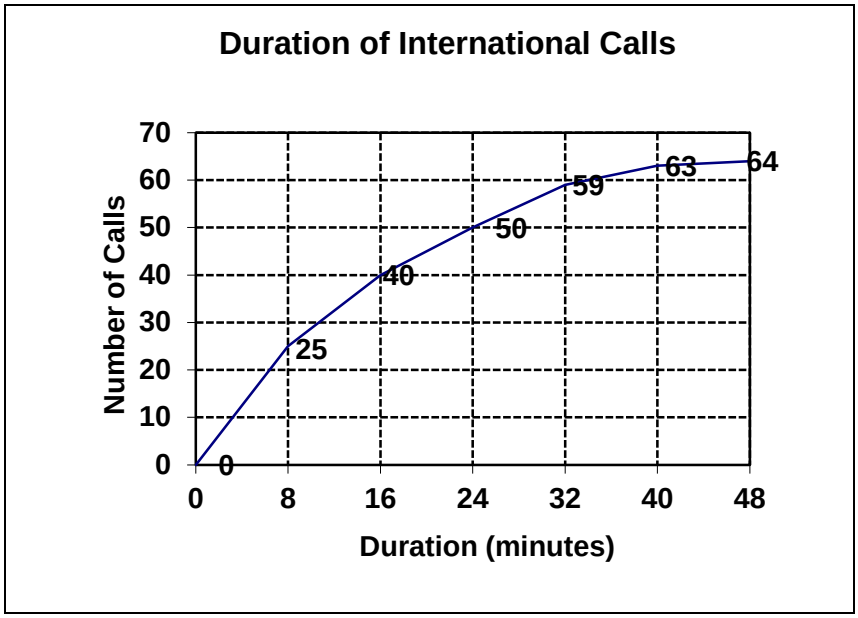
4. Do you think the organisation should be concerned about the results obtained from this study? Give reasons for your answer.

Commentary Related to Activity Design:
This activity introduces you to interpreting data and decision making by making use of the results obtained. If you experience any problems in working with the percentage column, refer back to Learning Unit 1.

12.4 Activity 2

Purpose:
Determine if it is possible to get the same information from a frequency distribution graph and from a frequency table.

Task:
The following graph shows the duration of a sample of international phone calls using a prepaid calling card.



1. What is the sample size?

2. How many intervals are used in the construction of the graph?

3. What is the frequency of the intervals with the longest calls?

4. How long is the duration of most of the calls?

5. Half of the calls are shorter than how many minutes?

6. How many of the calls are longer than 32 minutes?

7. How many of the calls are 10 minutes or less?

12.5 Activity 3

Purpose:
The purpose of this activity is to do a complete analysis of a case study given with grouped data.

Task:
The following data set represents the distribution of annual salaries (R'000) of 50 males who all perform similar jobs in a particular industry.

Salary (R'000)	Frequency	Midpoint (x)	
20 -< 22	1	21	
22 -< 24	3	23	
24 -< 26	14	25	
26 -< 28	20	27	
28 -< 30	10	29	
30 -< 32	2	31	
	50		

1. Calculate the mean, median and modal annual salary for the males in this industry. Explain each answer.

2. By comparing the mean, median and modal statistics, what can you conclude about the shape of the distribution.

3. Determine the range, standard deviation, variance and coefficient of variation for the salaries for the males in the industry.

4. What proportion of the salaries will fall within one, two and three standard deviations from the mean?

5. Determine the inter-quartile range and interpret your answer.

6. Determine the middle 90% range.

7. Which value falls at the 60^{th} percentile? Explain this answer.

8. Plot the histogram and read the value of the mode from the graph. Comment on the skewness of the distribution.

9. Plot the ogive and read the value of the median from the graph.

Commentary Related to Activity Design:
It was required from you to use three different ways to determine skewness. Are the three answers the same? Remember that if you deal with grouped data, your answers are all approximations because the original data are not available.

12.6 Revision Exercise 1

A commercial farmer keeps record of the rainfall figures on his farm. Over the last 50 months the following readings (in ml) were recorded:

55.8	60.9	39.1	40.0	71.4	77.1	37.0	35.5	31.7	65.2
45.9									
83.2									
75.3									
60.7									

Summarise this data in the form of a table and show:

1. A frequency distribution. Determine which interval will show the most popular readings.

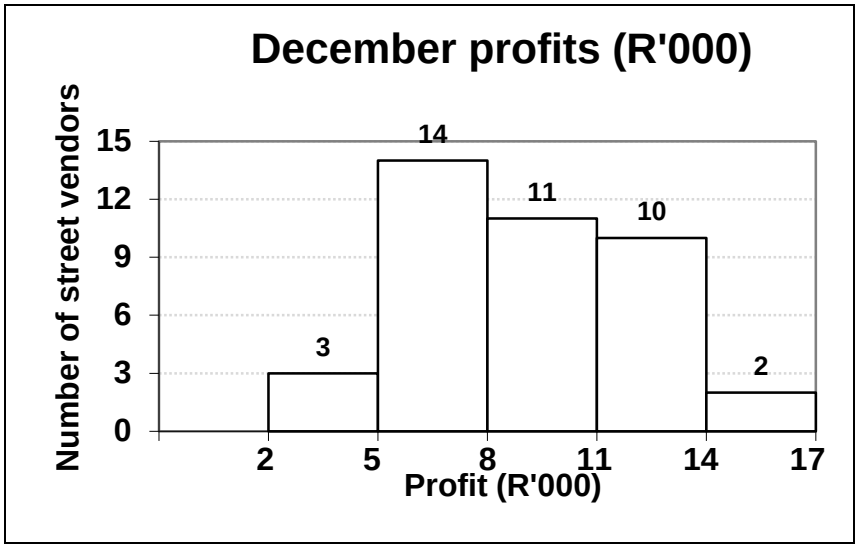
2. A histogram. What will the readings be for the interval with the lowest rainfall?

3. A polygon. What can you conclude about the shape of the polygon? What is the reason for the particular shape?

4. An ogive. Half of the months show a rainfall of less than how many millilitres?

12.7 Revision Exercise 2

A Histogram showing information of a street vendors' association in Johannesburg CBD for December last year.



1. What is summarised in this histogram?

2. How many street vendors were sampled?

3. How much profit is approximately made by the vendors who earned the most?

4. The 17 vendors who earned the least, earned less than how much?

5. What proportion of the vendors earn between R11 000 and R14 000?

6. How many vendors earn less than R8 000?

7. What can you conclude about the shape of the distribution? Explain what it means.

12.8 Revision Exercise 3

13. Each prospective employee who applies for a job at a certain bank is given a test. A sample of the length of time it took 42 applicants to write the test was organised into the following frequency distribution table:

Length of Time (min)	Number of Applicants
1 -< 4	4
4 -< 7	8
7 -< 10	14
10 -< 13	9
13-< 16	5
16-< 19	2

5.1 What is the mean time it took the applicants to write the test?

5.2 Determine median time to write the test.

5.3 Determine the time most applicants took to write the test.

12.9 Revision Exercise 4

1. Packets of biscuits are labelled as weighing 300 grams, but actual weights follow a normal distribution with a mean of 297 grams. If 1.15% of the packets weigh more than 303 grams, what is the standard deviation of the weights?

2. The average amount that 50 customers spent at Macy's is unknown. If approximately 95% of the customers spent between R125 and R375, determine the unknown mean and standard deviation amount spent.

3. The distribution of actual weight of potatoes in a 7kg bag is known as symmetrical with a mean equal to 7kg and a standard deviation equal to 0.25kg. Based on this, between which two limits could we expect 95% of all bags to weigh?

4. Packets of biltong are labelled as weighing 500g. Actual weights follow a normal distribution with a mean of 504g and a standard deviation of 4g. What proportion of packets will weigh between 496g and 512g?

5. In a study of student study debt, 125 undergraduate students selected at random were found to have average debts amounting to R9 500 per student, with a standard deviation of R1 200. The median debt was R8 800. What percentage of students has debts between R7 100 and R11 900?

12.10 Revision Exercise 5

Know the difference between ungrouped data and grouped data

Ungrouped data:
Cuppachino Coffee Shop offers free refills when customers purchase filter coffee. The number of refills taken by 20 customers one Wednesday morning were as follows:

0 1 3 1 2 0 2 2 0 1 0 3 1 0 1 2 1 1 0 2
Arrange these figures in a frequency distribution table.

No of Refills	Tally	Frequency (f_i)

Grouped data:
Data is grouped into categories known as classes when a data set contains a large number of values. A grouped frequency distribution is compiled showing the number, or frequency, of observations in each class.

The number of BBM's received in one day by 22 students were as follows:
50 14 25 8 10 33 52 12 45 15 7 5 98 13 31 52 6 75 17 20 12 64

Arrange these figures in a frequency distribution table, show absolute, relative and cumulativefrequencies.
(Use one of your templates attached to your worksheets given today)

Use the following intervals: 0 – 19, 20 – 39, 40 – 59, 60 – 79, 80 - 99

Solutions to Exercises

Izimvo Exchange 1

Questions:	Model Solutions:
Visit your campus bookstore and record the prices of a random sample of 30 different academic text books. Describe the method you used to obtain your data and construct a frequency distribution as well as all the relevant graphs. Compare the graphs obtained with that of your class mates. Look at shape, range, peaks, etc.	Each student will have a different sample selection. The students would have used different sampling methods. Let them discuss why they have chosen a specific method and how did they go about to select the sample. The class can discuss if the method they followed can be considered a random sampling method.

Activity 1

The following relative frequency distribution resulted from a study of the time (in minutes) employees of an organisation arrive late for work during a specific month:

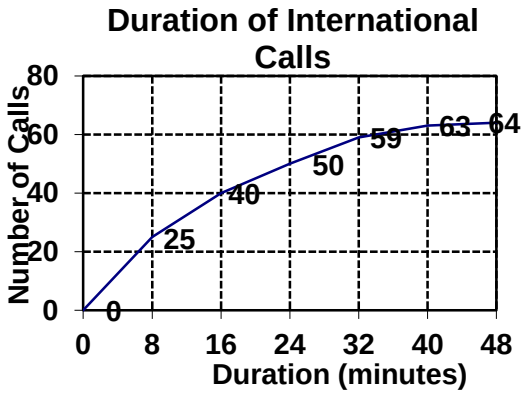
Minutes Late	% of Employees
$1.00 \leq x < 6$	1
$6.00 \leq x < 11$	3
$11.00 \leq x < 16$	4
$16.00 \leq x < 21$	6
$21.00 \leq x < 26$	7
$26.00 \leq x < 31$	9
$31.00 \leq x < 36$	11
$36.00 \leq x < 41$	19
$41.00 \leq x < 46$	32
$46.00 \leq x < 51$	8

Minutes Late	% of Employees	X	C<f	f
$1.00 \leq x < 6$	1	3.5	1	30
$6.00 \leq x < 11$	3	8.5	4	90
$11.00 \leq x < 16$	4	13.5	8	120
$16.00 \leq x < 21$	6	18.5	14	180
$21.00 \leq x < 26$	7	23.5	21	210
$26.00 \leq x < 31$	9	28.5	30	270
$31.00 \leq x < 36$	11	33.5	41	330
$36.00 \leq x < 41$	19	38.5	60	570
$41.00 \leq x < 46$	32	43.5	92	960
$46.00 \leq x < 51$	8	48.5	100	240
	100			**3 000**

Questions:	**Model Solutions:**
1. Determine the class midpoints for each interval and explain the meaning of the midpoints in two of the intervals.	• Interval 2: The value that represents this interval is 8.5; 3% of the employees are approximately 8.5 minutes late; • Interval 7: The value that represents this interval is 33.5; 11% of the employees are approximately 33.5 minutes late.

Questions:	Model Solutions:
2. Construct the cumulative frequencies for each interval.	
3. If the study covers a sample of 3 000 employees, how many of them fall into each class interval?	
4. Do you think the organisation should be concerned about the results obtained from this study? Give reasons for your answer.	Yes, the company should be concerned because 70% of the employees are late for half an hour or longer.

Activity 2

Questions:	Model Solutions:
The following graph shows the duration of a sample of international phone calls using a prepaid calling card.  	
1. What is the sample size?	The last point is the sample size = 64.
2. How many intervals are used in the construction of the graph?	Count the number of upper boundaries = 6.
3. What is the frequency of the intervals with the longest calls?	The steepest increase is in the interval between 8 and 16. The frequency of that interval is 40 – 25 = 15.
4. How long is the duration of most of the calls?	Between 8 and 16 minutes.

Questions:	Model Solutions:
5. Half of the calls are shorter than how many minutes?	Half of the calls are 32. If you read 32 from the y-axis to the graph and drop down to the x-axis, the duration is approximately 11 minutes.
6. How many of the calls are longer than 32 minutes?	59 calls are shorter than 32 minutes. There are 64 calls in the sample; therefore 5 calls were longer than 32 minutes.
7. How many of the calls are 10 minutes or less?	Approximately 29 calls.

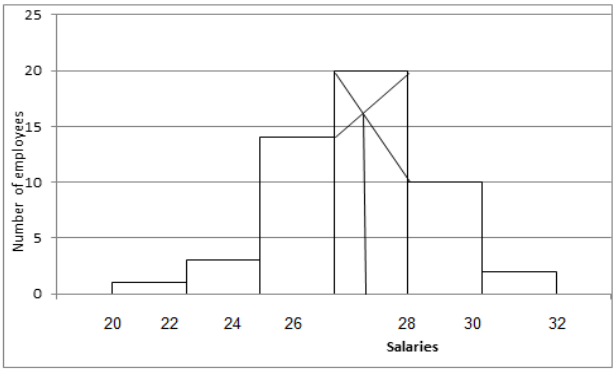
Activity 3

The following data set represents the distribution of annual salaries (R'000) of 50 males who all perform similar jobs in a particular industry.

Salary	Frequency	x	Xf	C<f	X^2f
20 -< 22	1	21	21	1	441
22 -< 24	3	23	69	4	1587
24 -< 26	14	25	350	18	8750
26 -< 28	20	27	540	38	14580
28 -< 30	10	29	290	48	8410
30 -< 32	2	31	62	50	1922
	50		1332		35690

Questions:	Model Solutions:
1. Calculate the mean, median and modal annual salary for the males in this industry. Explain each answer.	• The average of all 50 males is approximately R26 000: $\bar{x} = \frac{1332}{50} = 26.64$ • Median is value number 25. This value will fall in interval 4 between 26 and 28: Median = $26 + \left(\frac{25-18}{20}\right)2 = 26.7$ This means that half of the males in similar jobs will approximately be paid less than R26 700; • Mode will fall in the interval with the highest frequency. That is interval 4 between 26 and 28: Mode = $26 + \left(\frac{10}{10+14}\right)2 = 26.75$ This means that more of the males get approximately R26 830 than any other salary.

2.	By comparing the mean, median and modal statistics, what can you conclude about the shape of the distribution.	$\bar{x} = 26.64$ Median = 26.7 Mode = 26.75 The distribution is slightly negatively skewed because mode > median > mean. These statistics are very close together; therefore we can say the distribution is approximately normal.
3.	Determine the range, standard deviation, variance and coefficient of variation for the salaries for the males in the industry.	• Range = Upper boundary of last interval – lower boundary of first interval = 32 – 20 = 12 (approximately) • Standard deviation = $s = \sqrt{\frac{\sum fx^2 - \frac{1}{n}(\sum fx)^2}{n-1}} = \sqrt{\frac{35690 - \frac{1}{50}(1332)^2}{50-1}} = 2.05$ ≈ R2 050 (Students can use their calculators' pre-programmed function to calculate this answer. The formula from the formula sheet would give the same answer.) • Variance (s^2) = 2.05^2 = 4.20 ≈ R4 200. Coefficient of variation $= \frac{2.05}{26.64} \times 100 = 7.70\%$
4.	What proportion of the salaries will fall within one, two and three standard deviations from the mean?	• Within two standard deviations: = 26.64 – 2(2.05) = 22.54 = 26.64 + 2(2.05) = 30.74 According to Chebycheff's theorem at least 75% of the salaries will be between R22 540 and R30 740; • Within three standard deviations: = 26.64 – 3(2.05) = 20.49 = 26.64 + 3(2.05) = 32.79 According to Chebycheff's theorem at least 89% of the salaries will be between R20 490 and R32 790. • Within four standard deviations: = 26.64 – 4(2.05) = 18.44 = 26.64 + 4(2.05) = 34.84 According to Chebycheff's theorem at least 94% of the salaries will be between R18 440 and R34 840.

5.	Determine the inter-quartile range and interpret your answer.	Inter-quartile range = 27.95 – 25.21. That means the middle 50% of the values fall between 25.21 and 27.95.
6.	Determine the middle 90% range.	Middle 90% range= $P_{95} - P_5$ P_{95} = $28 + \frac{(47.5-38)2}{10} = 29.9$ P_5 = $22 + \frac{(2.5-1)2}{3} = 22.17$
7.	Which value falls at the 60th percentile? Explain this answer.	$P_{60} = 26 + \frac{(30-18)2}{20} = 27.2$
8.	Plot the histogram and read the value of the mode from the graph. Comment on the skewness of the distribution.	**Histogram of Annual Salaries**  The mode is just smaller than R26 750. The distribution is negatively skewed but close to normal.
9.	Plot the ogive and read the value of the median from the graph.	**Ogive for Annual Salaries**  The median is just lower than R27 000.

Revision Exercise 1

A commercial farmer keeps record of the rainfall figures on his farm. Over the last 50 months the following readings (in ml) were recorded:

55.8	60.9	39.1	40.0	71.4	77.1	37.0	35.5	31.7	65.2
45.9	59.1	91.3	56.0	36.7	52.6	49.5	65.8	44.6	62.3
83.2	58.2	69.3	42.3	71.7	47.3	48.0	69.8	33.8	61.2
75.3	94.6	61.8	64.9	60.6	61.5	56.3	78.8	27.1	76.0
60.7	47.2	30.0	39.8	87.1	69.0	74.5	68.2	65.0	66.3

Questions:

Summarise this data in the form of a table and show:

1. A frequency distribution. Determine which interval will show the most popular readings.

Model Solutions:

Range = 94.6 - 27.1 = 67.5

Number of intervals (K) = 1 + 3.3log50 = 6.6 ≈ 7

Width (c) = $\frac{67.5}{7} = \mathbf{9.64} \approx \mathbf{10}$

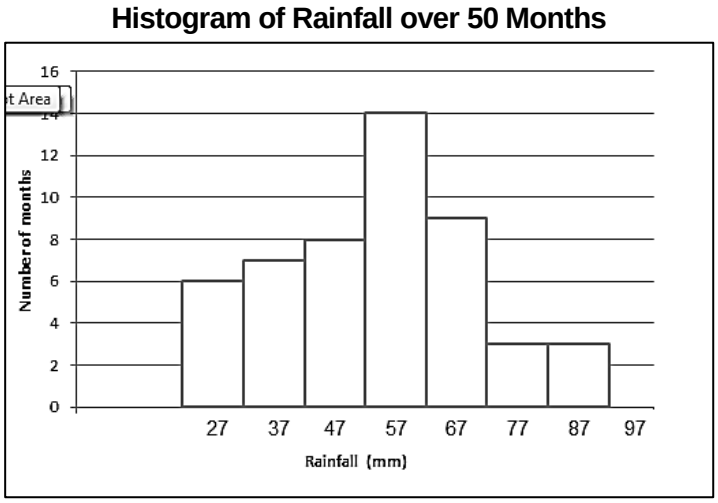
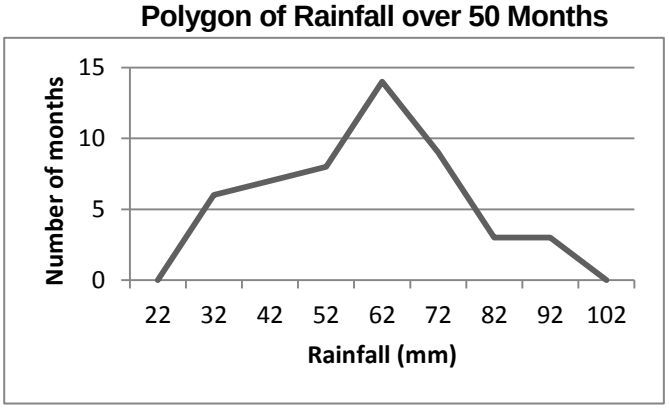
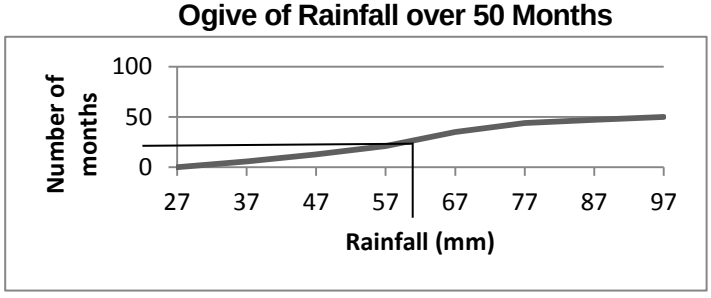
Test: K × c > R

7 × 10 > 67.5 Thus the frequency table will work out.

Start with an integer just smaller than the smallest value in the data set.

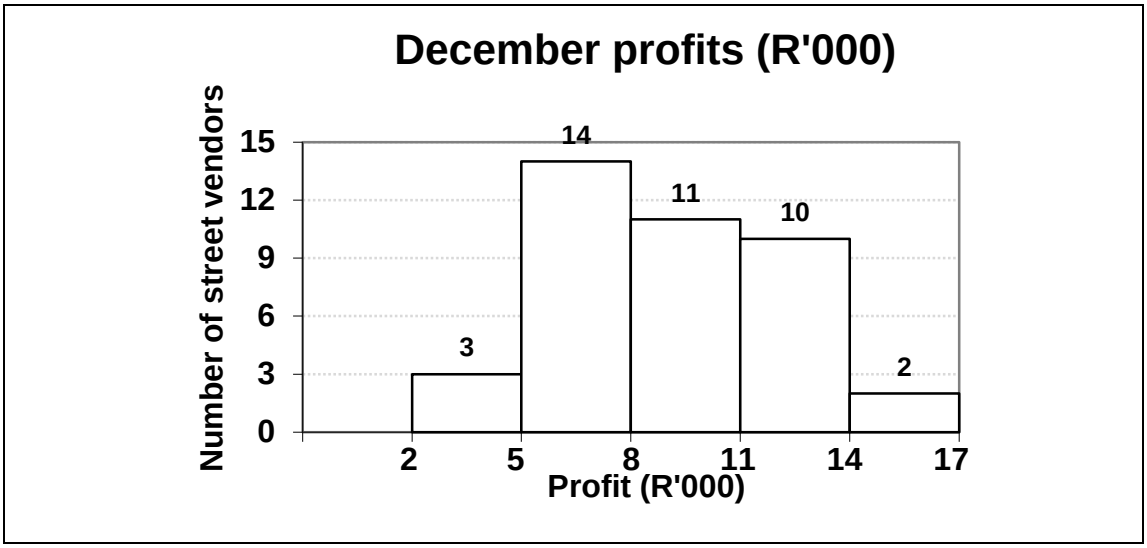
Class intervals	Tally	Fre-quency	%f	C<f
27–36	~~1111~~ 1	6	12	6
37–46	~~1111~~ 11	7	14	13
47–56	~~1111~~ 111	8	16	21
57–66	~~1111~~ ~~1111~~ 1111	14	28	35
67–76	~~1111~~ 1111	9	18	44
77–86	111	3	6	47
87–96	111	3	6	50
		50	100	

The most popular rainfall reading is between 57mm and 67mm.

2.	A histogram. What will the readings be for the interval with the lowest rainfall?	Between 27mm and 37mm. 
3.	A polygon. What can you conclude about the shape of the polygon? What is the reason for the particular shape?	The shape is approximately normal. The data are evenly distributed. 
4.	An ogive. Half of the months show a rainfall of less than how many millilitres?	Half of 50 is 25. Draw a straight line from 25 to the graph and drop a straight line down to the x-axis. Read the answer from there. It is approximately 60mm. 

Revision Exercise 4

A Histogram showing information of a street vendors' association in Johannesburg CBD for December last year.



Questions:		Model Solutions:
1.	What is summarised in this histogram?	Profits made during December by a sample of street vendors.
2.	How many street vendors were sampled?	3 + 14 + 11 + 10 + 2 = 40
3.	How much profit is approximately made by the vendors who earned the most?	The ones that earned the most will fall in the last interval: therefore they earned between R14 000 and R17 000.
4.	The 17 vendors who earned the least, earned less than how much?	Less than R8 000.
5.	What proportion of the vendors earn between R11 000 and R14 000?	Between R8 000 and R11 000. $\frac{10}{40} = 25\%$
6.	How many vendors earn less than R8 000?	3 + 14 = 17
7.	What can you conclude about the shape of the distribution? Explain what it means.	The distribution is positively skewed which means that more of the values are concentrated to the left or smaller side of the distribution.

Revision Exercise 3

<table>
<tr><th>Questions:</th><th>Model Solutions:</th></tr>
<tr><td>
1. Each prospective employee who applies for a job at a certain bank is given a test. A sample of the length of time it took 42 applicants to write the test was organised into the following frequency distribution table:
<table>
<tr><th>Length of Time (min)</th><th>Number of Applicants</th></tr>
<tr><td>1 -< 4</td><td>4</td></tr>
<tr><td>4 -< 7</td><td>8</td></tr>
<tr><td>7 -< 10</td><td>14</td></tr>
<tr><td>10 -< 13</td><td>9</td></tr>
<tr><td>13-< 16</td><td>5</td></tr>
<tr><td>16-< 19</td><td>2</td></tr>
<tr><td></td><td></td></tr>
</table>
1.1 What is the mean time it took the applicants to write the test?
1.2 Determine median time to write the test.
1.3 Determine the time most applicants took to write the test.
</td><td>
1.1 Mean = $\frac{379}{42} = 9.02 \text{ minutes}$

1.2 Median = 7 + $\frac{(21-12)3}{14}$ = 8.93 minutes. Half of the candidates took shorter than 8.64 minutes to write the test.

1.3 $\bar{x} = \frac{379}{42} = 9.14 \text{ minutes}$
</td></tr>
<tr><td>
2. For the last 10 days in January this year the Gautrain from Pretoria arrived late in Johannesburg by the following number of minutes. (A negative number means that the train was early by the number of minutes.)

-2 6 4 10 -4 12 2 -1 3 1

Determine the range and the mean of the data set.
</td><td>
Range = 12 – (-2)
= 14
$\bar{x} = 3.1 \text{ minutes}$ (Use calculator)
</td></tr>
<tr><td>
3. A sample of students has a mean of 3.2 members in their family. The modal number of family members is two and the median number is 2.1. Based on this information what will the shape of the distribution probably be?
</td><td>
Positively skewed because mode < median < mean.
</td></tr>
</table>

4. Earthquake intensities are measured using a device called a seismograph which is designed to be most sensitive for earthquakes with intensities between 4.0 and 9.00 on the open-ended Richter scale. Measurements of 18 recent earthquakes gave the following readings:

4.5	L	5.5	H	8.7	8.9	6.0	H	5.2
L	4.5	5.2	5.5	6	8.7	8.9	H	H

L indicates that the earthquake had an intensity of below 4.0 and H indicates that the earthquake had intensity above 9.0. Fifty per cent of the earthquake intensities were more than what value?

6 on the Richter scale. That is the median value.

5. The number of rejects from 50 samples of the same size is as follows:

Number of Rejects in Sample	Number of Samples (Frequency of Rejects)
0	5
1	10
2	10
3	20
4	5

5.1 The arithmetic mean number of rejects per sample is _____?

5.2 Half of the samples have less than _____?

5.3 The modal number of rejects is _____?

5.1 $\frac{110}{50}$ = 2.2 rejects

5.2 2.5 rejects

5.3 3

6. A financial analyst's sample of six companies' book values (in R'000) were:

25 7 22 33 18 15

If the sample mean is R20 000, what is the sample standard deviation? (Round your answer to the closest Rand.)

$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\frac{396}{6-1}} = 8.89 \approx$ R8 899

<table>
<tr><th>Questions:</th><th>Model Solutions:</th></tr>
<tr><td>7. A company has two regional head offices; in Cape Town and Johannesburg. Workers in Johannesburg claim that their salaries are more variable than the workers in Cape Town. To test their claim the following data was collected for a random sample of 100 workers in each office.

<table>
<tr><th></th><th>Cape Town</th><th>Johannesburg</th></tr>
<tr><td>Mean Salary</td><td>R27 000</td><td>R25 000</td></tr>
<tr><td>Standard Deviation</td><td>R2 000</td><td>R2 100</td></tr>
</table>

Are the salaries in Johannesburg more variable? Use the coefficient of variation to prove your answer.</td><td>CV(CT) = $\frac{2000}{27000} \times 100 = 7.41\%$
CV(Jhb) = $\frac{2100}{25000} \times 100 = 8.4\%$

Yes, salaries are more variable in Johannesburg because the CV is higher.</td></tr>
<tr><td>8. Random samples of small townhouse selling prices are obtained from ABSA Bank and First National Bank. The results followed normal distributions and are summarized below:

<table>
<tr><th></th><th>ABSA</th><th>First National</th></tr>
<tr><td>Sample Size</td><td>50</td><td>80</td></tr>
<tr><td>Mean House Price</td><td>R150 000</td><td>R160 000</td></tr>
<tr><td>Standard Deviation</td><td>R20 000</td><td>R25 000</td></tr>
</table>

8.1 Which financial institutions' reported prices can be considered more uniform? Use the coefficient of variation to prove your answer.
8.2 With reference to ABSA, at least 68% of the selling prices were between which two values?</td><td>8.1 CV (ABSA) = $\frac{20000}{150000} \times 100 = 13.33\%$
CV (First National) = $\frac{25000}{160000} \times 100 = 15.62\%$
ABSA's reported prices are more uniform.

8.2 According to the Empirical rule 68% of the prices will fall within 1 standard deviation from the mean; that means between
(R150 000 – 1(R25 000) = R125 000
(R150 000 + 1(R25 000) = R175 000</td></tr>
<tr><td>9. In a data set of 120 observations, how many observations lie between the 50th percentile and the 60th percentile?</td><td>P_{50} is value number 60.5
P_{60} is value number 72.6
The difference between the positions is 12 values.</td></tr>
</table>

Revision Exercise 4

Questions:	Model Solutions:
1. Packets of biscuits are labelled as weighing 300 grams, but actual weights follow a normal distribution with a mean of 297 grams. If 1.15% of the packets weigh more than 303 grams, what is the standard deviation of the weights?	Within 3 standard deviations you can find 97.7% of the observations. That means in the upper tail you will find 1.15% of the values. R297 + 3(s) = 303 $\therefore$3s = 6 $\therefore$1s = 2
2. The average amount that 50 customers spent at Macy's is unknown. If approximately 95% of the customers spent between R125 and R375, determine the unknown mean and standard deviation amount spent.	95% of the observations fall within 2 standard deviations from the mean (Empirical rule). 4s = 375 – 250 = 125 S = $\frac{125}{4} = \mathbf{31.25}$ Mean = $\frac{125}{2}$ = 62.5
3. The distribution of actual weight of potatoes in a 7kg bag is known as symmetrical with a mean equal to 7kg and a standard deviation equal to 0.25kg. Based on this, between which two limits could we expect 95% of all bags to weigh?	According to the Empirical rule 95% of the observations will fall within 2 standard deviation from the mean. Lower limit: 7kg – 2(0.25) = 6.5 kg Upper limit: 7 kg + 2(0.25) = 7.5 kg
4. Packets of biltong are labelled as weighing 500g. Actual weights follow a normal distribution with a mean of 504g and a standard deviation of 4g. What proportion of packets will weigh between 496g and 512g?	504 -496 = 8 $\frac{8}{4} = \mathbf{2}$s. According to the Empirical rule 95% of the packets will weigh between 496g and 512g.
5. In a study of student study debt, 125 undergraduate students selected at random were found to have average debts amounting to R9 500 per student, with a standard deviation of R1 200. The median debt was R8 800. What percentage of students has debts between R7 100 and R11 900?	The distribution is skewed, therefore use Chebyshef's rule. R11 900 – R7 100 = R4800 $\frac{4800}{1200} = \mathbf{4}\,\boldsymbol{s}$ $\therefore$2 s to the left and 2 s to the right. That means that at least 95% of the students between R7 100 and R11 900.

Learning Unit 5: Linear Regression and Correlation Analysis

Material used for this Learning Unit:	My notes
• Prescribed text pp.299—321.	
How to prepare for this Learning Unit:	
• Before the first class, be sure that you read pp.299—321 in the prescribed textbook.As you read these sections, see if you can find the answers to the following questions: o What is linear regression? o Why is it important? o Why do you need to do Correlation analysis?	

1 Introduction

Most statistical studies examine data on more than one variable because we are interested in the relationship, if any, between variables. You may want to find out how the value of one variable changes when the value of another variable changes, that is, finding out if one variable relates to another. The way we express this interest is through simple correlation and regression analysis.

This analysis is used to determine

- Whether there is a relationship between variables;
- The strength of the relationship;
- How the relationship is used to forecast.

In this course you will study linear relationship only.

1.1 Dependent Variable and the Independent Variable

The first step in studying regression and correlation is to label the two variables. The label “Y” goes to the variable which we are trying to predict or cannot control. This is the dependent or response variable. The response variable is explained by the explanatory (or independent) variable, labelled “X.”

1.2 Scatter Diagram

The first step in identifying whether there is a relationship and if it is linear, positive or negative, is to plot a scatter diagram. This graphically plots the relationship between X and Y. The value of the independent X is plotted in respect of the horizontal axis, and the value of the dependent Y is plotted in respect of the vertical

axis. The points are not connected. The strength of the relationship may be concluded tentatively.

Possible shapes of the scatter plot are shown on p.377 in textbook.

1.3 The Coefficient of Correlation (r)

This coefficient measures the strength of association between the variables. The value of *r* can be negative or positive and indicates the direction of the relationship between X and Y. When the slope of the scattered points is negative, r is negative and when it is positive, r is positive.

If a negative relationship exists, that is, if Y decreases as X increases, then r will fall between -1 and 0. If there is a positive relationship, that is, if Y increases as X increases, then r will fall between 0 and +1. If there is no relationship between X and Y, then $r = 0$. This measure enables us to make statements such as "the correlation is strong, weak, etc." The strength of the correlation is not dependent on direction. $r = 0.95$ and $r = -0.95$ are equal in strength.

1.4 The Coefficient of Determination (r^2)

This measure enables you to calculate the total variation in Y explained by the association between X and Y. The value of this measure is the square of the correlation coefficient multiplied by 100. The answer will always be positive within the range 0 – 100%.
Coefficient of determination = $r^2 \cdot 100$

1.5 Regression Analysis

The general process of predicting dependent variable Y based on independent variable X, is known as regression analysis. If it is known that a relationship exists between two variables, you develop an estimating equation that will give the best prediction of Y. "Best" refer to how close the predictions of Y are to the actual values of Y. A linear relationship will result in a straight line if plotted on a graph. A linear function involving two variables can be expressed in the form:

$\hat{y} = a + bx$ where:

$\hat{y}$ = estimated Y-value for a given X-value.

a = intercept on the Y-axis.

b = the slope of the linear line.

1.6 Interpret the Y-Intercept and the Slope.

The *y*-intercept or "*a*" value is the point where the regression line crosses the *y*-axis. This is the *y*-value if the *x*-value is zero.

The slope or "b" value is the average change in Y for each one unit change in X. A positive slope indicates a positive relationship and a negative slope indicates a negative relationship. For a horizontal line, b=0, which means there is no relationship between the two variables.

1.7 Predict the Value of the Dependent Variable for a Given Independent Variable.

The purpose of regression analysis is to use the regression equation to estimate values for *y* from known x values. Estimates of y are found by substitution a given x value into the regression equation.

Note: *(i)* Estimated y-values are meaningful only for x-values in (or close to) the range of the given data. Estimating y within the given range of the x-values, use interpolation. Predicting a y-value for an x-value outside the given range, extrapolation. You cannot be certain of the behaviour of the data for which you have no observations!

2 Recommended Additional Reading

You should engage extensively with your Recommended Additional Reading as it is an essential part of the learning experience. It will give you different perspectives and engagement opportunities with different authors and schools of thoughts.

This will give you deeper insight into this discipline and has been designed to add value to the module.

Johnson RR, Kuby PJ. 2011. *Elementary Statistics*. 11 ed. Duxbury.

Sullivan M. 2010. *Fundamentals of Statistics*. 3 ed. Pearson's Education.

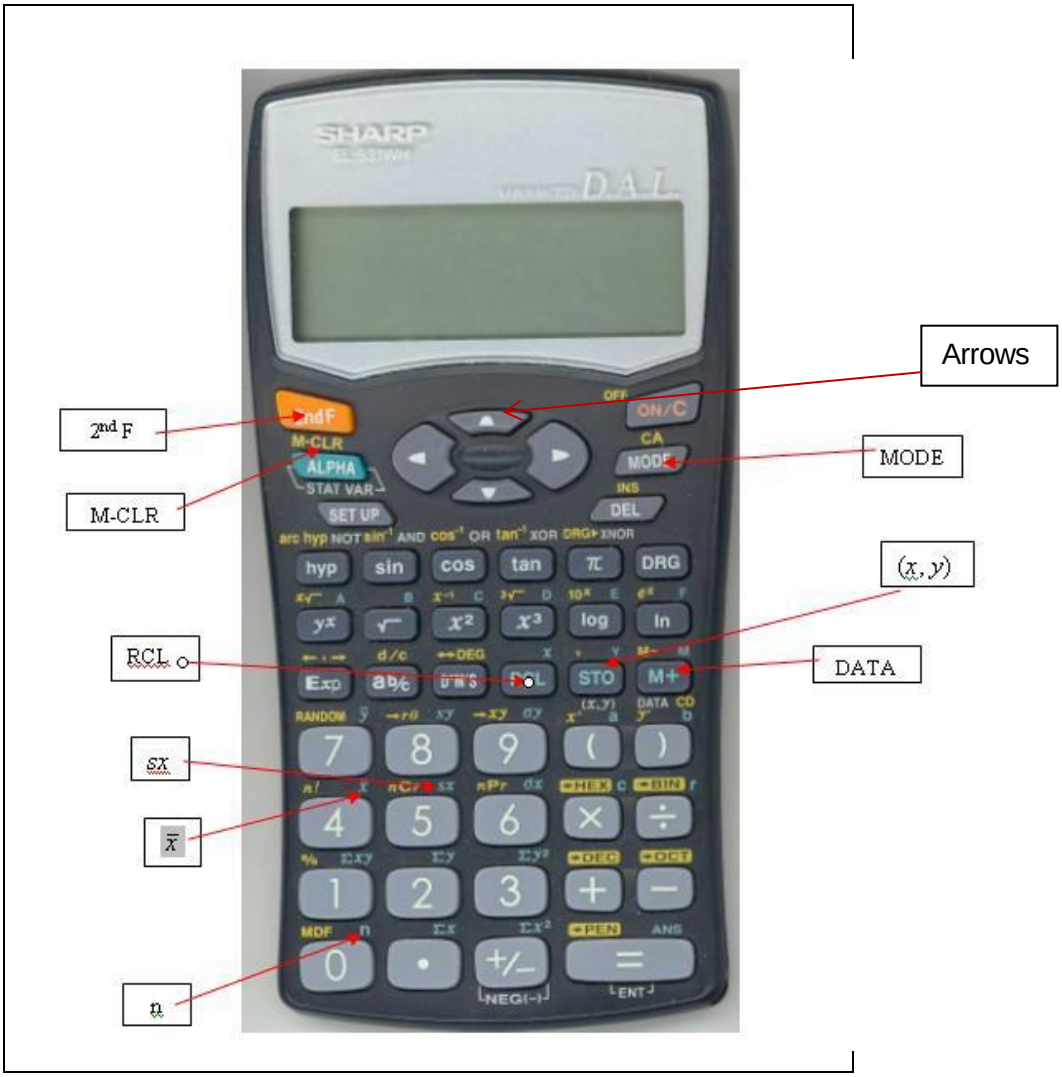
Triola MF. 2009. *Elementary Statistics*. 11 ed. Pearson's Education.

Willemse I. 2009. *Statistical Methods and Calculation Skills*. 3 ed. Juta.
http://youtu.be/7Mfa1MbVhcc

Your lecturer will provide you with the blog/Facebook details for your class from which you can obtain information or give comments. Digital activities will also be available on the blog.

Go to http://www.aw.com/triola and find the Internet project on linear regression.
Correlation and Regression Analysis using a calculator

3 Statistics on the Sharp EL-531WH calculator



Your calculator has two different MODES. One is for normal calculations and the other for statistical calculations.

1. Select Statistical Mode by pressing MODE followed by 1
2. Select 1 again (LINE) for Linear Regression.
3. <STAT 1> will appear on display.
4. Clear the statistical memory before you start with the data input.
   Select 2ndF MODE This CA function clears all MODE memory.
5. Enter the *x*-value of the first data point and select 

   .
   Enter the *y*-value of the first data point and select 

Note: When in linear regression mode the 2ndF key is not required as the key now assumes the (x, y) function shown below the key.

6. <DATA SET = 1> will appear on the screen.

7. Continue in the same way until all the data values have been entered.
   Example:

X value	Y value
3	86
4	92
5	95
4	83
2	78
3	82

Calculator steps:
Select 2ndF MODE to clear memory.

3 STO 86 M+
4 STO 92 M+
5 STO 95 M+
4 STO 83 M+
2 STO 78 M+
3 STO 82 M+

8. Once all the values have been entered, clear the screen by selecting <On/C>.

9. If you have made a mistake when entering the values, delete the incorrect data point by scrolling up or down to the data point of concern using the arrows top middle of calculator:
   - Select **2nd F** **M+** to delete the incorrect *x*- and *y*-value. Enter the correct *x*- and *y*-values.
   - It does not matter when this data point was entered.

10. Retrieving Totals

    All of the totals previously calculated manually can now be obtained as follows:
    n: RCL 0
    Σy: RCL 2
    Σx: RCL ·

Σxy: RCL ·
Σx^2: RCL +
Σy^2: RCL 3

For the example: n = 6 Σy = 516 Σx = 21 Σxy = 1835 Σx^2= 79 Σy^2= 44582

11. Regression and correlation output:

11.1 To obtain Pearson's correlation coefficient *r*, select **RCL** ÷
11.2 To obtain *a* (the *y*-intercept of the regression line), select **RCL** (
11.3 To obtain *b* (the slope of the regression line), select **RCL**)

For the example: *r* = 0,86, *a* = 67,55 *b* = 5,27

12. After completing the regression question it is important to go back to the NORMAL mode before doing any other calculations.

4 Statistics on the Casio FX82MS calculator

1. Set the calculator to Statistical Mode by selecting MODE followed by 3 for the REG (regression) option. Select 1 again (LINE) for Linear Regression.
<REG> will appear on display.

2. Clear the statistical memory before you start with the data input.
SHIFT CLR 1 =

5 Entering x and y data values

1. Enter the x-value of the first data point and select ,

2. Enter the y-value of the first data point and select M+

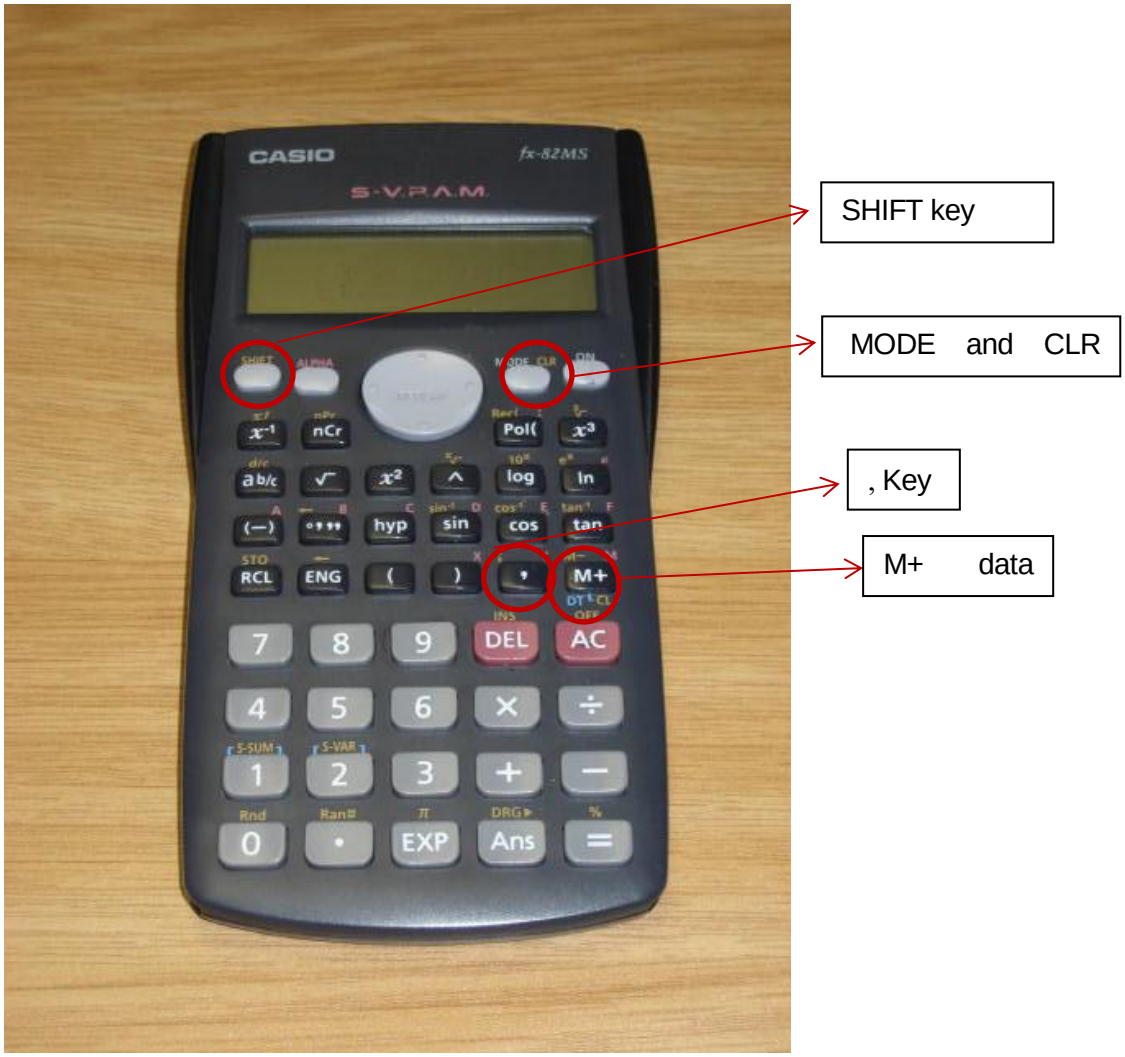
3. It should say n = 1 on the screen.

4. Enter the x-value of the second data point and select , Enter the y-value of the second data point and select M+.

5. n = 2 will appear on the screen.

6. Continue in the same way until all the data values have been entered.

7. Once all the values have been entered, clear the screen by selecting AC.



8. If you have made a mistake when entering the values, delete the incorrect data point by scrolling up or down to the data value of concern using the arrows located on the large round button.

9. Delete the incorrect value, by selecting SHIFT M+. This will delete the applicable x- and y-value.

10. Enter the correct values. Once you have finished editing the data, selecting AC to exit the data edit mode.

Example:

X value	Y value
3	86
4	92
5	95
4	83
2	78
3	82

Calculator steps:

SHIFT CLR 1 = to clear memory.

3 , 86 M+
4 , 92 M+
5 , 95 M+
4 , 83 M+
2 , 78 M+
3 , 82 M+

11. All of the totals required to calculate r, b and a , which were previously calculated manually in class, can now be obtained from the [S-SUM] menu as follows:

n: **SHIFT 1 3**
Σx: **SHIFT 1 2**
Σx^2: **SHIFT 1 1**
Σy: **SHIFT 1 SCROLL RIGHT WITH ARROW 2**
Σxy: **SHIFT 1 SCROLL RIGHT WITH ARROW 3**
Σy^2: SHIFT 1 SCROLL RIGHT WITH ARROW 1

12. Retrieving r, a and b:

To obtain Pearson's correlation coefficient r, select
SHIFT 2 < **SCROLL RIGHT WITH ARROW TWICE** 3 =
To obtain ***a*** - the y-intercept value, select
SHIFT 2 < **SCROLL RIGHT WITH ARROW TWICE** 1 =
To obtain b – the slope, select
SHIFT 2 < **SCROLL RIGHT WITH ARROW TWICE** 2 =

13. Before attempting a new question, one has to clear the data set that is currently in the calculator's memory.

14. After completing the regression question, go back to normal calculation MODE.

6 Interactive Work Space

6.1 Izimvo Exchange 1

For each of the following pairs of variables, indicate which variable is the independent variable and which one is the dependent variable. What type of relationship do you expect in each case?

1. The age of second hand cars and their prices.

2. The weight of new cars and their fuel consumption in kilometres per litre.

3. The heights and the weights of adult men.

4. The heights and the IQ scores of adult men.

5. Family income level and percentage of income donated to charities.

6. Number of days absent and test marks.

7. Shoe size and height of person.

8. Years of education and annual income.

9. Self-esteem and body weight.

10. Estimated study hours and subsequent test marks.

11. Densely populated areas and crime rates.

12. Students who often watch TV and performance in examinations.

6.2 Izimvo Exchange 2

Visit bookstore that has a display of “best-sellers” books. Record the prices and number of pages of at least 10 books. Use a random sampling method to select the books. Use the data to develop a regression equation with price the dependent variable. Are there any apparent outliers in your data?

6.3 Activity 1

Purpose:
This activity is designed to test all aspects of regression and correlation analysis as well as interpretations.

Task:
A travel agency is interested in knowing how airline fares are related to the length of the flight in kilometres. The agency hypothesised that the longer the flight, the more the airfare. The following sample data were collected:

Kilometres	2375	1400	1250	2325	985	2025
Airfare (R)	1330	810	750	1266	621	1110

1. Identify the dependent and independent variables?

2. Plot a scatter diagram of the data.

3. Referring to diagram, do you think that the travel agency's hypothesis is correct? Why or why not?

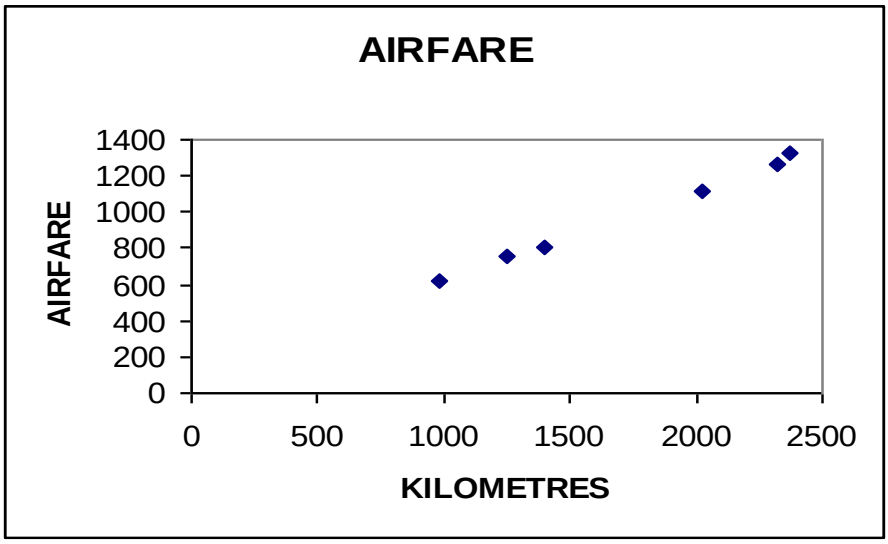
4. Calculate the correlation coefficient and interpret the answer.

5. Determine the coefficient of determination and interpret the answer.

6. If you are going to travel 2 400 kilometres, what do you expect the airfare to be?

Solution:

Airfare depends on the distance the aeroplane flies and therefore airfare is the *Y*-variable and kilometres the *X*-variable.



The agency's hypothesis was correct. The scatter plot shows a positive relationship which means that the longer the flight, the more expensive the airfare.

Commentary Related to Activity Design:
Observational data allow us to state that two variables might be related, but one cannot claim causation.

6.4 Activity 2

The maximum temperature (in °C) and coffee sales (in Rands) for a coffee shop for eight randomly selected days were recorded:

°C	16	19.5	25.5	30	32.5	36	39	40.5
Sales (R)	262	248	197	200	133	139	114	112

1. Draw a scatter diagram of coffee sales and daily temperature. Comment on the results.

2. Calculate the correlation coefficient and interpret the answer.

3. Determine the coefficient of determination and interpret the answer.

4. If the forecasted temperature is 18°C, how many cups of coffee do you expect to sell?

5. How many cups of coffee do you expect to sell at the y-intercept point? Why will this not be a realistic estimate?
6. Interpret the slope of the distribution of cups of coffee sold.

Commentary Related to Activity Design:
Observational data allow us to state that two variables might be related, but you cannot claim causation. The slope of a distribution is also known as the marginal value.

6.5 Revision Exercise 1

Bulk advertising is often used to influence buyer's response. For example a product is advertised at "2 for R20" to convince people they are getting a bargain. To test this theory a Fruit and Veg store advertises an item for equal periods of time at five bulk rates and records the quantities sold:

Number of items in bulk sale	Quantity sold
1	35
2	50
3	45
4	75
5	62

1. Plot a scatter diagram to indicate if a relationship exists.

2. Calculate and interpret the correlation coefficient.

3. Calculate and interpret the coefficient of determination.
4. Determine the regression equation and interpret the a, and b, values in terms of number of items and quantities sold.

5. Estimate the quantity that can be sold if six items are advertised at a certain price. Can you rely on this estimate? Give a reason for your answer.

6.6 Exercise 2

The following table shows the weeks employed of six people working at a car manufacturing inspection station and the number of cars each of them checked between 8h00 and 12h00 on a given day.

Weeks employed	**Cars checked**
5	16
1	15
7	19
9	23
2	14
12	21

The following totals are also provided:

$$\sum x = 36 \quad \sum y = 108 \quad \sum xy = 715 \quad \sum x^2 = 304$$
$$\sum y^2 = 2008$$

Answer the next four questions:

1. Name the independent variable? Name the dependent variable?

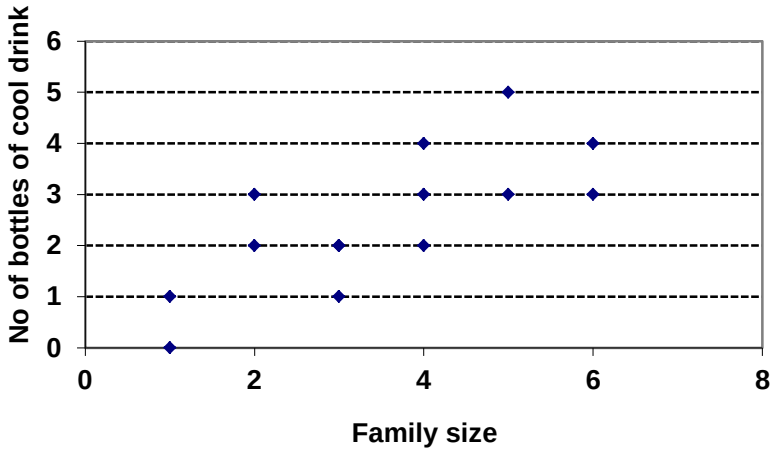
2. Prove that the number of weeks employed is a good measure to use to predict the number of cars checked, making use of the coefficient of determination.

3. Estimate the number of cars a person employed for six weeks can check between 8h00 and 12h00 on a given day.

4. Calculate the slop of the regression line.

6.7 Revision Exercise 3

Below is a scatter diagram illustrating the number of bottles of cold drinks purchased and the size of families? The sample was selected from the customers who visited the Do-Little Supermarket during a specific day.



1. What is the sample size?

2. Do you expect a positive or negative relationship between the two variables?

3. What is the smallest value reported for family size?

4. What is the largest value reported for number of bottles of cold drink?

Solutions to exercises

Izimvo Exchange 1

For each of the following pairs of variables, indicate which variable is the independent variable and which one is the dependent variable. What type of relationship do you expect in each case?

Questions:	Model Solutions:
1. The age of second hand cars and their prices.	Independent: Age Dependent: Price
2. The weight of new cars and their fuel consumption in kilometres per litre.	Independent: Weight Dependent: Fuel consumption
3. The heights and the weights of adult men.	Independent: Height Dependent: Weight
4. The heights and the IQ scores of adult men.	Not possible to determine because you cannot expect a relationship.
5. Family income level and percentage of income donated to charities.	Independent: Income Dependent: Charity donation
6. Number of days absent and test marks.	Independent: Days absent Dependent: Test mark
7. Shoe size and height of person.	Independent: Height Dependent: Shoe size
8. Years of education and annual income.	Independent: Years of education Dependent: Income
9. Self-esteem and body weight.	Independent: Body weight Dependent: Self-esteem
10. Estimated study hours and subsequent test marks.	Independent: Hours of study Dependent: Test marks
11. Densely populated areas and crime rates.	Independent: Population density Dependent: Crime rate
12. Students who often watch TV and performance in examinations.	Independent: Time watching TV Dependent: Exam performance

Izimvo Exchange 2

Visit bookstore that has a display of "best-sellers" books. Record the prices and number of pages of at least ten books. Use a random sampling method to select the books. Use the data to develop a regression equation with price the dependent variable. Are there any apparent outliers in your data?

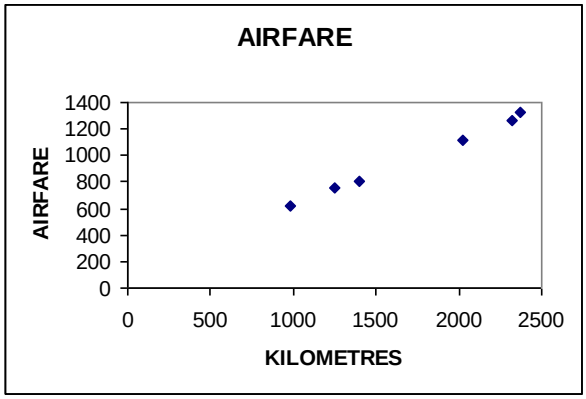
(The students will have different answers. Let them discuss their experiment and how they have collected the data.)

Activity 1

Purpose:
This activity is designed to test all aspects of regression and correlation analysis as well as interpretations.

Task:
A travel agency is interested in knowing how airline fares are related to the length of the flight in kilometres. The agency hypothesised that the longer the flight, the more the airfare. The following sample data were collected:

Kilometres	2375	1400	1250	2325	985	2025
Airfare (R)	1330	810	750	1266	621	1110

Questions:	**Model Solutions:**
1. Identify the dependent and independent variables?	Airfare depends on the distance the aeroplane flies and therefore airfare is the *Y*-variable and kilometres the *X*-variable.
2. Plot a scatter diagram of the data.	
3. Referring to diagram, do you think that the travel agency's hypothesis is correct? Why or why not?	The agency's hypothesis was correct. The scatter plot shows a positive relationship which means that the longer the flight, the more expensive the airfare.

4. Calculate the correlation coefficient and interpret the answer.

Distance (X)	Airfare(Y)	XY	X^2	Y^2
2375	1330	3158750	5640625	1768900
1400	810	1134000	1960000	656100
1250	750	937500	1562500	562500
2325	1266	2943450	5405625	1602756
985	621	611685	970225	385641
2025	1110	2247750	4100625	1232100
10360	**5887**	**11033135**	**19639600**	**6207997**

$\bar{x} = \frac{10360}{6} = 1726.67 \quad \bar{y} = \frac{5887}{6} = 981.17$

$S_{XX} = \sum x^2 - \frac{1}{n}(\sum x)^2$ = 19639600 - $\frac{1}{6}(10360)^2$ = 1751333.33

$S_{XY} = \sum xy - \frac{1}{n}\sum x \sum y$ = 11033135 - $\frac{1}{6}$(10360)(5887)= 868248.33

$S_{YY} = \sum y^2 - \frac{1}{n}(\sum y)^2 = 6207997 - \frac{1}{6}(5887)^2 = 431868.83$

$r = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}} = \frac{868248.33}{\sqrt{(1751333.33 \times 431868.83}}$ = 0.998

The correlation is positive and very good.

5. Determine the coefficient of determination and interpret the answer.

$r^2 \times 100 = 0.9982 \times 100 = 99.6\%$

99.6% of the variation in airfare is due to variation in distance.

6. If you are going to travel 2 400 kilometres, what do you expect the airfare to be?

$\hat{y}$= a + bx $b = \frac{S_{XY}}{S_{XX}} = \frac{868248.33}{1751333.33} = 0.4958$

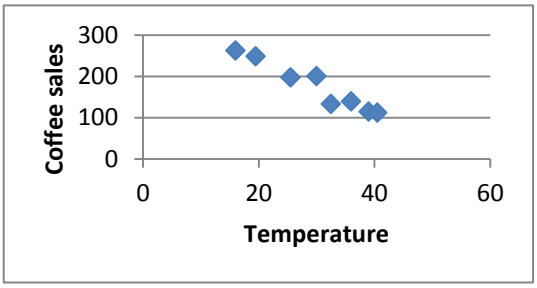
=125.09 + 0.4958x $a = \bar{y} - b\bar{x}$ = 981.17 – 0.4958(1726.67) =125.09

$\hat{y}_{(x=2400)}$=125.09 + 0.4958(2400) = R1315.01

Activity 2

The maximum temperature (in °C) and coffee sales (in Rands) for a coffee shop for eight randomly selected days were recorded:

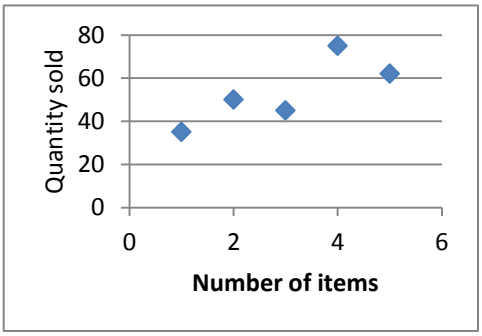
°C	16	19.5	25.5	30	32.5	36	39	40.5
Sales (R)	262	248	197	200	133	139	114	112

Questions:	**Model Solutions:**
1. Draw a scatter diagram of coffee sales and daily temperature. Comment on the results.	 The scatter plot shows a negative relationship between temperature and coffee sales. the higher the temperature, the lower the coffee sales.
2. Calculate the correlation coefficient and interpret the answer.	r = -0.97 The correlation is negative and very good.
3. Determine the coefficient of determination and interpret the answer.	$r^2 \times 100 = (-0.97)^2 = 94\%$ of the variation in coffee sales is caused by a variation in temperature
4. If the forecasted temperature is 18°C, how many cups of coffee do you expect to sell?	$\hat{y}$= 368.49 -6.46x $\hat{y}$= 368.49 -6.46(18) = R252.21
5. How many cups of coffee do you expect to sell at the y-intercept point? Why will this not be a realistic estimate?	R368.49. It is not realistic because temperature is interval scale data. There is no true zero.
6. Interpret the slope of the distribution of cups of coffee sold.	For every I°C increase in temperature there is a decrease of 6.46 cups of coffee sold.

Revision Exercise 1

Bulk advertising is often used to influence buyer's response. For example a product is advertised at "2 for R20" to convince people they are getting a bargain. To test this theory a Fruit and Veg store advertises an item for equal periods of time at five (5) bulk rates and records the quantities sold:

Number of items in bulk sale	Quantity sold
1	35
2	50
3	45
4	75
5	62

Questions:	**Model Solutions:**
1. Plot a scatter diagram to indicate if a relationship exists.	 There is a positive relationship, the more items in bulk sale, the more quantities sold.
2. Calculate and interpret the correlation coefficient.	r = 0.81 There is a good positive relationship between bulk rate and sales.
3. Calculate and interpret the coefficient of determination.	$r^2 \times 100 = (0.81)^2 = 66\%$ of the variation in sales can be explained by the variation in the number of items in bulk rate.
4. Determine the regression equation and interpret the a, and b, values in terms of number of items and quantities sold.	$\hat{y} = 29.7 + 7.9x$ If there are no items in the bulk rate, sales will be 29.7 items. For every one more item in the bulk rate the sales will increase with 7.9 items.
5. Estimate the quantity that can be sold if six items are advertised at a certain price. Can you rely on this estimate? Give a reason for your answer.	$\hat{y} = 29.7 + 7.9(6) = 77.1$ items

Revision Exercise 2

The following table shows the weeks employed of six people working at a car manufacturing inspection station and the number of cars each of them checked between 8h00 and 12h00 on a given day.

Weeks employed	**Cars checked**
5	16
1	15
7	19
9	23
2	14
12	21

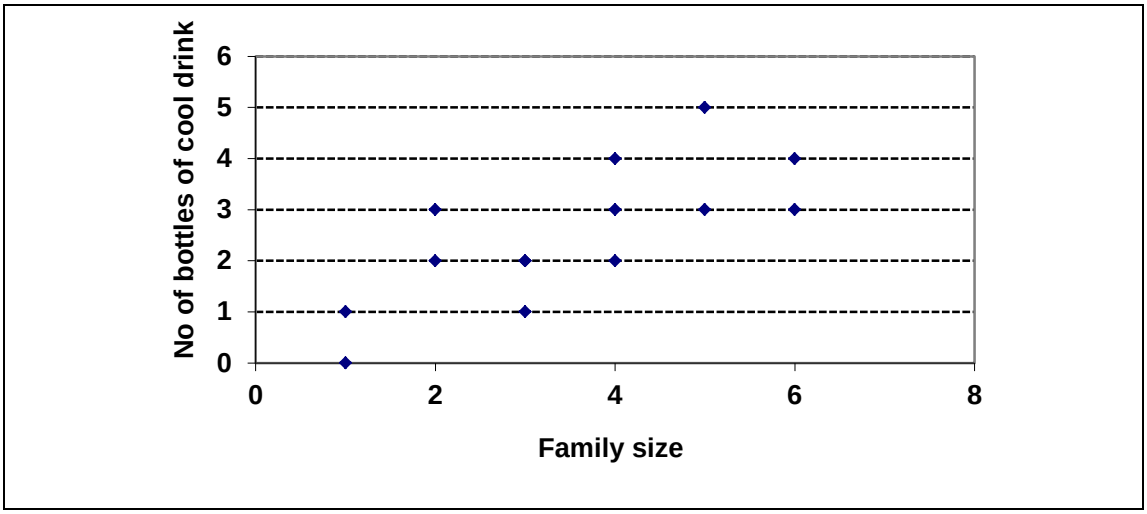
The following totals are also provided:

$$\sum x = 36 \quad \sum y = 108 \quad \sum xy = 715 \quad \sum x^2 = 304 \quad \sum y^2 = 2008$$

Questions:	**Model Solutions:**
1. Name the independent variable? Name the dependent variable?	Independent variable: Weeks employed Dependent variable: Cars checked
2. Prove that the number of weeks employed is a good measure to use to predict the number of cars checked, making use of the coefficient of determination.	R = 0.89 $r^2 \times 100 = (0.89)^2 = 79.21\%$ of the variation in number of cars checked can be explained by the variation in the number of weeks employed.
3. Estimate the number of cars a person employed for six weeks can check between 8h00 and 12h00 on a given day.	08h00 and 12h00 on a given day. $\hat{y}$= 13.44 +0.76(6) =18 cars
4. Calculate the slop of the regression line.	For every one more week employed, a person can check 0.76 cars more.

5.6 Revision Exercise 3

Below is a scatter diagram illustrating the number of bottles of cold drinks purchased and the size of families? The sample was selected from the customers who visited the Do-Little Supermarket during a specific day.



Questions:		Model Solutions:
1.	What is the sample size?	n = 13
2.	Do you expect a positive or negative relationship between the two variables?	Positive relationship.
3.	What is the smallest value reported for family size?	One
4.	What is the largest value reported for number of bottles of cold drink?	Five

Learning Unit 6: Index Numbers	
Material used for this Learning Unit: • Prescribed text pp.323—358.	**My notes**
How to prepare for this Learning Unit: • Before the first class, be sure that you read pp.323—358 in the prescribed textbook. • As you read these sections, see if you can find the answers to the following questions: o What are index numbers? o Why do you need them? o When do you use them?	

1 The Concept "Index Numbers"

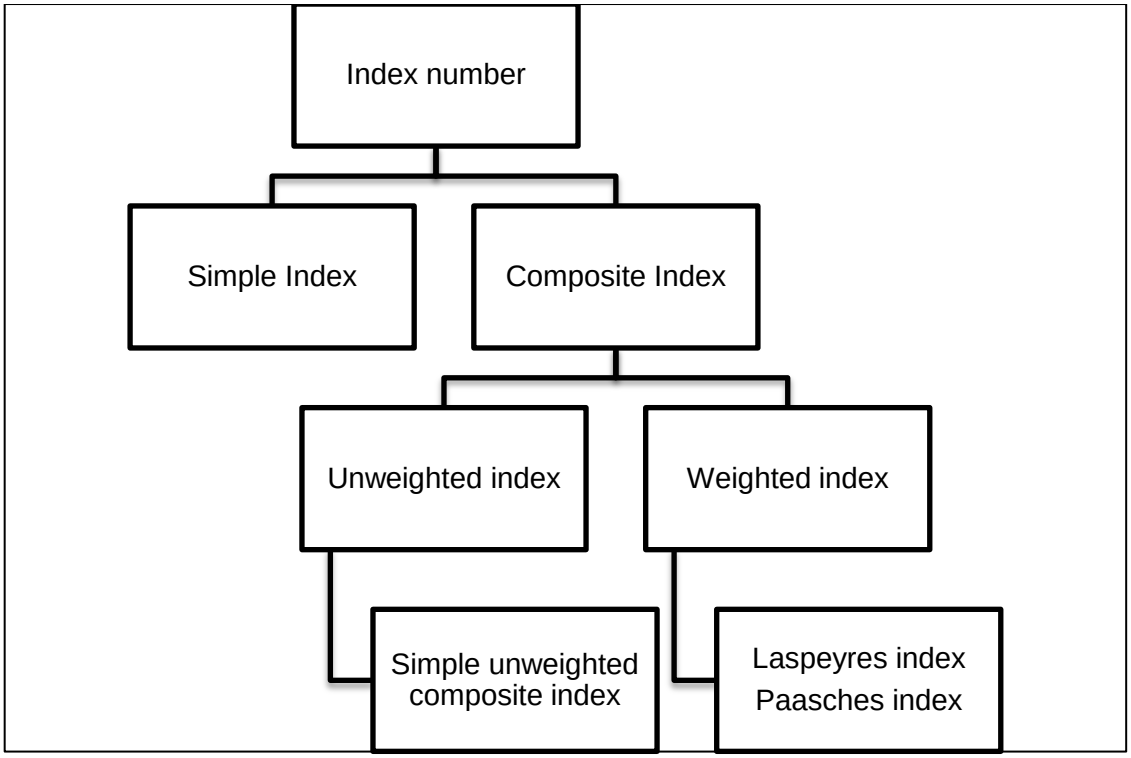
- It is a relative figure expressed as a percentage and measures the change in a set of measurements over time**;**
- The reason why we calculate index numbers is because we base business forecasts on what has happened in the past;
- To calculate an index number, the current value (numerator) is divided by a *base* value (denominator) and the result is multiplied by 100;
- The base period is the point in time to which a comparison is made.

2 Categories of Index Numbers

Measurements compared by an index number can be concerned with price, quantity or value.

- A price index (Ip) compares changes in prices from one period to another. A simple or composite index can be calculated;
- A quantity or volume index (Iq) measures how much the quantity of a variable changes over time. Simple or composite quantity indexes can be calculated.

The flow chart shows the different categories of index numbers followed by the explanation of each category.



3 The Difference between Simple and Composite Index Numbers

A simple index number is used to measure the rate of change between time periods for a single item. For example; the changes in the price of the Sunday newspaper from last January 2011 to January 2012. A composite or aggregate index number is used to measure the rate of change between time periods for a group of two or more items in a basket.

4 The Difference between Unweighted and Weighted Index Numbers

- An unweighted Index numbers deals with only price *or* quantity. It means each item in a basket is of equal importance;
- A weighted index number deals with both price and quantity. It means each item is not of equal importance. A system of weights is applied so the index reflects the relative importance of its components. The price of an item is generally weighted by the quantity sold during a time period.

5 Interpretation of Index Numbers

The base year index is defined as equal to 100 and each subsequent year will be above or below 100 depending on whether there has been an increase or decrease in the data compared to the base year. For example; an index number of 114 means an increase of 14% above the base year value and an index number of 91 means a decrease of 9% below the base year value. If the calculated index is 100, it means there was no change in the price or quantity compared to the base year value.

Price index numbers may be used to remove the effects of price changes. This is known as “deflating data.”

Index numbers are useful to measure historical performance of commodities, bur have limited forecasting abilities. It is incorporated into time series analysis to track past performances and time series tools are then used to predict turning points in economic activity.

6 Calculating an Index Number.

Step 1: Obtain the prices or quantities for the commodity during the time period of interest.
Step 2: Select the period to be used as the base period and label the different columns with P_o and Q_o in the base period and P_n and Q_n in the current period.
Identify the appropriate formula for the question and substitute the values into the formula.
Step 3: Divide the denominator by the numerator and multiply this ratio by 100.
Step 4: Interpret the answer.

7 Examples of Important Indices

- JSE all share index;
- JSE Gold Index;
- CPI – Consumer’s Price Index;
- Inflation Rate;
- PPI – Production Price Index;
- Business Confidence Index;
- New Car Sales Index.

The South African Statistical Services is an important source of economic indexes and publishes regular reports based on indexes, e.g. CPI.

8 The Consumer Price Index (CPI) and Inflation Rate

The CPI is a very important economic indicator and is used to determine the inflation rate and cost of living. The monthly CPI publication by Stats SA is a matter of great interest for business, government and ordinary citizens. The formula used to determining the CPI in South Africa is that of Laspeyres.

9 Recommended Additional Reading

You should engage extensively with your Recommended Additional Reading as it is an essential part of the learning experience. It will give you different perspectives and engagement opportunities with different authors and schools of thoughts. This will give you deeper insight into this discipline and has been designed to add value to the module.

Willemse, I. 2009. *Statistical methods and calculation skills*. 3rd edition. Pretoria: Juta.

10 Recommended Digital Engagement and Activities

Your lecturer will provide you with the blog/Facebook details for your class from which you can obtain information or give comments. Digital activities will also be available on the blog.

On the CD Rom enclosed with your textbook: Refer to the section on “Reading statistical tables” for detailed instructions on how to read the required critical values associated with a level of confidence from the standard normal distribution tables.

Access the following Internet Project Page. The Internet project provides simulations, demonstrations or activities that enhance the topics covered in this learning unit.

11 Interactive Work Space

Consult the web p. http://www.stassa.gov.za for detailed information on economic indicators.

11.1 Izimvo Exchange 1

Collect some articles containing index numbers, either in the business sections of the newspapers or in financial and business magazines, to bring to class. Find out how your examples are calculated and what they are used for. The different types of indexes can be discussed in class.

11.2 Izimvo Exchange 2

Discuss the concept questions on p. 418.

11.3 Activity 1

Purpose:
The purpose is to use summary values to measure how prices and quantities change over time periods. The value of an item in the current period is expressed as a ratio of the value in a base period. A step-by-step approach will be used in this activity to explain all concepts.

Task:
The table below shows the prices (R) and quantities (kg) of food items bought during 2010 and 2011.

	2010		**2011**					
	Price Po	**Quantity Qo**	**Price P_1**	**Quantity Q_1**	**P_1Qo**	**PoQo**	**Q_1Po**	**P_1Q_1**
Rice	7	80	6	70	480	560	490	420
Meat	30	50	35	60	1 750	1 500	1 800	2 100
Potatoes	3	100	3	100	300	300	300	300
	40	230	44	230	2 530	2 360	2 590	2 820

Step 1:
The base period is given as the year 2010.

Step 2:
Label 2010 price and quantity columns as P_o and Q_o and for 2012 as P_1 and Q_1.

1. With 2010 as the base period, calculate the 2011 simple price and quantity index for rice.

2. Which food item shows the largest change in quantity over the two year period?

3. Use 2010 as the base period and calculate the simple unweighted composite price and quantity indexes for 2011.

4. Use 2010 as the base period and calculate the change in price and quantity by making use of Laspeyres' method for 2011.
5. Use 2010 as the base period and calculate the change in quantity and price by making use of Paasshe's method for 2011.

Commentary Related to Activity Design:
Although an index is calculated as a percentage, the percentage symbol (%) is dropped and only used in interpretations. Laspeyres Index measures price changes over time, keeping quantities constant. This permits price changes to be monitored without the confounding effect of simultaneous quantity changes. In calculating Paashe's index, current period quantities are used to reflect more recent consumption patterns.

11.4 Activity 2

The following table shows the 2008 and 2009 prices and registrations of four makes of cars by a dealer in Johannesburg. The number of motor vehicles per make of car registered in 2008 and 2009 were used as respective weights.

	Prices (R'000)		**Number of registrations**			
Make	**2008 Po**	**2009 P_1**	**2008 Qo**	**2009 Q_1**	**Po Q_1**	**P_1Q_1**
Audi	170	228	57	68	9 690	12 996
BMW	226	286	526	492	118 876	150 436
Toyota	102	136	1703	1803	173 706	231 608
Volkswagen	77	99	1343	1229	103 411	132 957
Total	**575**	**749**	**3 629**	**3 592**	**40 5683**	**527 997**

Answer the next three questions.

1. With April 2008 as the base period, calculate the simple quantity index for BMW for April 2009.

2. With April 2008 as the base period, calculate the unweighted composite price index for April 2009.

3. With April 2008 as the base period, calculate the change in value for April 2009 by making use of Laspeyres' and Paashe's methods.

11.5 Revision Exercise 1

1. The Consumer Price Index for 2000 was 126 and in 2012 it was 209 with the same base year. If the majority of new graduates in the year 2000 could expect to get a job earning R24 000 a year, what must the starting salary be for the majority of new graduates in 2012 in order for them to be at the same standard of living as their counterparts in 2000? Is the typical 2012 graduate better or worse off than his/ her 2000 counterpart?

2. Your income in the base period was R30 000 and is now R40 000. During the same period the Consumer Price Index rose from 100 to 140. What is your real (deflated) income now taking the Consumer Price Index into account? Did your income keep pace with this increase?

3. Mr Makhubedu is offered a job in Cape Town with a salary of R123 500 per year. The cost of living index is 132. If he currently earns R100 000 per year in Johannesburg with a cost of living index of 120, will he be financially better off accepting the job offer?

11.6 Revision Exercise 2

The prices and quantities of bread were compared over a two year period. Use the data in the table below to answer the next three questions.

	Years					
Commodity	**2010**	**2011**	**2010**	**2011**	P_1Q_0	P_0Q_0
White bread	12.50	15.00	8	2	120.00	100
Brown bread	16.00	22.50	5	7	112.50	80
Rye bread	27.50	20.00	2	5	40.00	55
Total	56.00	57.50			272.50	235

1. Which type of bread shows the smallest relative change in price over the two years?

2. Calculate the simple composite quantity index for 2011 for all the commodities.

3. Calculate Laspeyres price index for 2011.

4. Calculate Paashe's price index for 2011. Compare the two weighted indexes you have computed.

11.7 Revision Exercise 3

The table below shows the prices and annual consumption of the raw materials used in Gauteng Breweries in 2010 and 2011

	Prices		**Unit Quantities**	
	2010	**2011**	**2010**	**2011**
Malt	49	46	10 874	15 116
Hops	512	724	732	696
Sugar	46	51	1 865	2 486
Wheat flour	31	27	873	1 093

1. Which raw material shows the largest relative change in price over the two years?

2. Which raw material shows the largest relative change in quantity over the two years?

3. Calculate the unweighted composite price and quantity indexes for 2011 and interpret the answer.

4. Calculate Laspeyres price index for 2011.

5. Calculate Paashe's price index for 2011.

6. Compare the two weighted indexes you have computed.

Solutions to Exercises

Izimvo Exchange 1

Collect some articles containing index numbers, either in the business sections of the newspapers or in financial and business magazines, to bring to class. Find out how your examples are calculated and what they are used for. The different types of indexes can be discussed in class.

Izimvo Exchange 2

Discuss the concept questions on page 418.

(Answers are given on the CD accompanying the textbook.)

Activity 1

Purpose:
The purpose is to use summary values to measure how prices and quantities change over time periods. The value of an item in the current period is expressed as a ratio of the value in a base period. A step-by-step approach will be used in this activity to explain all concepts.

Task:
The table below shows the prices (R) and quantities (kg) of food items bought during 2010 and 2011.

	2010		**2011**					
	Price Po	**Quantity Qo**	**Price P_1**	**Quantity Q_1**	**P_1Qo**	**$PoQo$**	**Q_1Po**	**P_1Q_1**
Rice	7	80	6	70	480	560	490	420
Meat	30	50	35	60	1 750	1 500	1 800	2 100
Potatoes	3	100	3	100	300	300	300	300
	40	230	44	230	2 530	2 360	2 590	2 820

Step 1:
The base period is given as the year 2010.

Step 2:
Label 2010 price and quantity columns as P_o and Q_o and for 2012 as P_1 and Q_1.

Questions:	**Model Solutions:**
1. With 2010 as the base period, calculate the 2011 simple price and quantity index for rice.	Step 3: The simple price index formula is $I_p = \frac{P_1}{P_o} \times 100$ Step 4: Insert rice prices in the formula: $I_{p(rice)} = \frac{6}{7} \times 100 = 85.71$ Step 5: Interpret the answer: There was a decrease of 100% -85.71% = 14.29% in the price of rice from 2010 to 2011.
2. Which food item shows the largest change in quantity over the two year period?	Do a simple quantity index for all three products and compare the answers. $I_q = \frac{Q_1}{Q_b} \times 100$ $I_{q(rice)} = \frac{70}{80} \times 100 = 87.5$ There is a decrease of 12.5% in quantity over the period. $I_{q(meat)} = \frac{60}{50} \times 100 = 120$ There is an increase of 20% in quantity over the period. $I_{q(potatoes)} = \frac{100}{100} \times 100 = 100$ There is no change in quantity consumed over the period. Rice shows a decrease of 12.5% and meat an increase of 20%. The largest change is then in the quantity meat consumed.
3. Use 2010 as the base period and calculate the simple unweighted composite price and quantity indexes for 2011.	Step 1: Sum the prices of all the items in the given period (*P*). Step 2: Sum the prices of all the items in the base period (P_o). Step 3: Substitute totals into formula and interpret the results. $I_p = \frac{\sum P_1}{\sum P_o} \times 100 = \frac{44}{40} \times 100 = 110$ There is an average increase of 10% in price for the three commodities over the time period. $I_q = \frac{\sum Q_1}{\sum Q_o} \times 100 = \frac{230}{230} \times 100 = 100$ There is no change in the quantity consumed for the three commodities over the time period. Use 2010 as the base period and calculate the change in price and quantity by making

	use of Laspeyres' method for 2011. $Ip = \frac{\sum P_1 Q_o}{\sum P_o Q_o} \times 100$ $= \frac{2\,530}{2\,360} \times 100 = 107.20$ There is an average increase of 8.9% in price since 2010 with the quantities held constant at 2010. $Ip = \frac{\sum Q_1 P_o}{\sum Q_o P_o} \times 100$ $= \frac{2\,590}{2\,360} \times 100 = 109.75$ There is an average increase of 11.46% in quantity since 2010 with the prices held constant at 2010.
4. Use 2010 as the base period and calculate the change in quantity and price by making use of Paasshe's method for 2011.	$Ip = \frac{\sum P_1 Q_1}{\sum P_o Q_1} \times 100$ $= \frac{2\,820}{2590} \times 100 = 108.88$ There is an average increase of 7.2% in price since 2010 with the quantities held constant at 2010. $Ip = \frac{\sum Q_n P_n}{\sum Q_o P_n} \times 100$ $= \frac{2820}{2\,530} \times 100 = 111.46$ There is an average increase of 9.75% in quantity since 2010 with the prices held constant at 2010.
5. Use 2010 as the base period and calculate the change in price and quantity by making use of Irving Fisher's method for 2011.	Ip (Irving Fisher) = $\sqrt{I_P \times I_L}$ =$\sqrt{107.2 \times 109.75}$ = 110.60 There is an increase of 10.16% since 2010.

Activity 2

The following table shows the 2008 and 2009 prices and registrations of four makes of cars by a dealer in Johannesburg. The number of motor vehicles per make of car registered in 2008 and 2009 were used as respective weights.

	Prices (R'000)		**Number of registrations**			
Make	**2008 Po**	**2009 P_1**	**2008 Qo**	**2009 Q_1**	**PoQ_1**	**P_1Qo**
Audi	170	228	57	68	9 690	12 996
BMW	226	286	526	492	118 876	150 436
Toyota	102	136	1 703	1 803	173 706	231 608
Volkswagen	77	99	1 343	1 229	103 411	132 957
Total	**575**	**749**	**3 629**	**3 592**	**405 683**	**527 997**

Questions:	**Model Solutions:**
1. With April 2008 as the base period, calculate the simple quantity index for BMW for April 2009.	$I_q = \frac{Q_1}{Q_b} \times 100 = \frac{492}{526} \times 100 = 93.54$ There was a decrease of 6.46% in the quantities registered of BMW's.
2. With April 2008 as the base period, calculate the unweighted composite price index for April 2009.	$I_p = \frac{\sum P_1}{\sum P_o} \times 100 = \frac{749}{575} \times 100 = 130.26$ There was a 30.26% increase in the price of the vehicles registered from 2008 to 2009.
3. With April 2008 as the base period, calculate the change in value for April 2009 by making use of Laspeyres' and Paashe's methods.	$Ip(L) = \frac{\sum P_1 Q_o}{\sum P_o Q_o} \times 100 = \frac{527997}{405683} \times 100 = 130.15$ An increase of 30.15% in price. $Ip(P) = \frac{\sum P_1 Q_1}{\sum P_o Q_1} \times 100 = \frac{523095}{401291} \times 100 = 130.35$ (Increase of 30.35%)

Revision Exercise 1

Questions:	**Model Solutions:**
1. The Consumer Price Index for 2 000 was 126 and in 2012 it was 209 with the same base year. If the majority of new graduates in the year 2000 could expect to get a job earning R24 000 a year, what must the starting salary be for the majority of new graduates in 2012 in order for them to be at the same standard of living as their counterparts in 2 000? Is the typical 2012 graduate better or worse off than his/ her 2 000 counterpart?	$\frac{24\,000}{126} \times 100 = 19\,047.62$ For 2012: $19\,047.62 \times \frac{209}{100} = R39\,809.53/$ If that is the salary in 2012 they will neither be better nor worse off.
2. Your income in the base period was R30 000 and is now R40 000. During the same period the Consumer Price Index rose from 100 to 140. What is your real (Deflated) income now taking the Consumer Price Index into account? Did your income keep pace with this increase?	$\frac{40\,000}{140} \times 100 = \mathrm{R}28\,571.43$ Your income did not keep pace with inflation.
3. Mr Makhubedu is offered a job in Cape Town with a salary of R123 500 per year. The cost of living index is 132. If he currently earns R100 000 per year in Johannesburg with a cost of living index of 120, will he be financially better off accepting the job offer?	$\frac{12\,3500}{132} \times 100 = \mathrm{R}93\,560.61$ $\frac{100\,000}{120} \times 100 = \mathrm{R}83\,333.33$ He will be better off in Cape Town

Revision Exercise 2

The prices and quantities of bread were compared over a two (2) year period. Use the data in the table below to answer the next three questions.

	Years					
Commodity	**2010**	**2011**	**2010**	**2011**		
White bread	12.50	15.00	8	2	120.00	100
Brown bread	16.00	22.50	5	7	112.50	80
Rye bread	27.50	20.00	2	5	40.00	55
Total	56.00	57.50			272.50	235

Questions:	**Model Solutions:**
1. Which type of bread shows the smallest relative change in price over the two years?	White: 20% increase Brown: 40.6% increase Rye: 27.27% decrease The smallest relative change is in the price of white bread.
2. Calculate the simple composite quantity index for 2011 for all the commodities.	$I_q = \frac{\sum Q_1}{\sum Q_o} \times 100 = \frac{14}{15} \times 100 = 93.3$ Index shows a decrease of 6.7% in overall quantities.
3. Calculate Laspeyres price index for 2011.	$Ip(L) = \frac{\sum P_1 Q_o}{\sum P_o Q_o} \times 100 = \frac{272.50}{235} \times 100$ $= 115.74$ An increase of 15.74% in price.
4. Calculate Paashe's price index for 2011. Compare the two weighted indexes you have computed.	$Ip(P) = \frac{\sum P_n Q_n}{\sum P_o Q_n} \times 100 = \frac{287.5}{274.5} \times 100$ $= 104.74$ (Increase of 4.74%)

Revision Exercise 3

The table below shows the prices and annual consumption of the raw materials used in Gauteng Breweries in 2010 and 2011

	Prices		Unit Quantities	
	2010	**2011**	**2010**	**2011**
Malt	49	46	10874	15 116
Hops	512	724	732	696
Sugar	46	51	1865	2 486
Wheat flour	31	27	873	1 093

Questions:	**Model Solutions:**
1. Which raw material shows the largest relative change in price over the two years?	Malt: Decrease of 6.12% Hops: Increase of 41.41% Sugar: Increase of 10.87% Wheat: Decrease of 12.9%
2. Which raw material shows the largest relative change in quantity over the two years?	Malt: Increase of 39.01% Hops: Decrease of 4.92% Sugar: Increase of 33.3% Wheat: Increase of 25.2%
3. Calculate the unweighted composite price and quantity indexes for 2011 and interpret the answer.	$I_p = \frac{\sum P_1}{\sum P_o} \times 100 = \frac{848}{638} \times 100 = 132.92$ (Increase of 32.82%) $I_q = \frac{\sum Q_1}{\sum Q_o} \times 100 = \frac{19391}{14344} \times 100 = 135.19$ (Increase of 35.19%)
4. Calculate Laspeyres price index for 2011.	$Ip(L) = \frac{\sum P_1 Q_o}{\sum P_o Q_o} \times 100 = \frac{1148858}{1020463} \times 100 = 112.58$ An increase of 12.58% in price.
5. Calculate Paashe's price index for 2011.	$Ip(P) = \frac{\sum P_1 Q_1}{\sum P_o Q_1} \times 100 = \frac{1355537}{1245275} \times 100 = 108.85$ (Increase of 8.85%)
6. Compare the two weighted indexes you have computed.	Due to the weight factor kept at current quantities, the changes in overall prices are lower for Paashe than Laspeyres.

Learning Unit 7: Financial Mathematics	
Material used for this Learning Unit: • Prescribed text pp.394—421.	**My notes**
How to prepare for this Learning Unit: • Before the first class, be sure that you read pp.394—421 in the prescribed textbook. • As you read these sections, see if you can find the answers to the following questions: o What is the purpose of calculating interest? o When do you calculate compound interest? o When do you calculate simple interest? o What is an annuity? o Why do you need to calculate this	

1 Introduction

Interest is the cost of money.

- If you borrow money, you pay interest.
- If you invest money you receive interest.
- The fact is “money earns money”.

1.1 Concepts used in interest calculations.

- Interest (I) is the money paid for borrowed capital or earned on capital invested.
- Principal or present value (P) is the capital on which interest is calculated at the beginning of the period.
- Rate of interest (r) is the percentage on the principle amount paid *per year*.
- Time period (t) is the period for which the money is borrowed/ invested *in years*.
- Amount or future value (A) is the amount at the end of the period (P + I).

1.2 Principles and notation of simple interest

- Simple interest is calculated once only per time period on the principal sum;
- Interest is not available before the end of the term and the interest is not added to the principle to earn interest on interest;
- The standard formulae for calculating simple interest are:

  $I = Prt$

  $A = P(1+rt)$

  $A = P + I$

Note: Exact interest is calculated on a basis of 365 days per year or 366 in a leap year. Ordinary interest is calculated on a basis of 360 days per year or 30 days per month.

1.3 Calculate all possible components in simple interest.

You must be able to calculate any of the four components of interest, namely present values, the future values, the interest rate and the term of investment or loan.

1.4 Principles of compound interest

When interest is not paid out, but continuously added to the principal, the principal is continuously increasing and the interest is compounded.

Interest can be compounded

- Yearly, this is once a year;
- Semi-annually. This is twice per year;
- Quarterly. This which means four times per year;
- Monthly. This is 12 times a year;
- Daily. This means 365 times per year.

The given yearly rate should be adjusted to a period rate (i).
If the interest rate is 12% per year then
i = 12% per year.
i= 1% per month.
i = 3% per quarter.
i = 6% semi-annually.

The time period (t), which is normally quoted in years, should be adjusted to the number of interest periods (n) per transaction.

n = number of years × number of compounding periods per year

- Interest compounded quarterly for n = 5 years = 5×4 = 20 interest compounding periods
- Interest calculated monthly for n =3 years = 3 × 12 = 36 compounding periods.

Formula for calculating compound interest is:

$$A = P(1 + i)^n$$

1.5 Calculate unknown components in compound interest.

You must be able to calculate any of the four components of interest, namely the present value, the future value, the interest rate or the term of investment or loan.
The interest rate and term must coincide with the compounding period

1.6 Nominal and Effective rate

Read page 403 and take note of the following formulae:

$$r = \left(1 + \frac{i}{m}\right)^{m} - 1$$

1.7 Concepts of annuities

An annuity is a sequence of equal payments made at equal time intervals such as instalment payments, pensions, insurance premiums, home loan payments, rent, etc.

- The time between successive payments (R) is the payment interval and the time between the first payment and the last payment is called the term of the annuity.
- The payment interval and the interest period always coincide which mean if the interest is compounded monthly, the payments will be monthly.

Annuities are classified into two main types depending on the commencement of payment within the first period.

- Ordinary annuities certain refer to annuities where the regular payments are made at the end of each payment interval.
- Ordinary annuities due refer to annuities where the periodic payment (R) falls at the beginning of each payment interval.
- Deferred annuities refer to annuities where the first of the regular payments will be made at some future date and not in the first time period.

1.8 Determine the values of different types of annuities.

Annuity calculations require finding one of three values, namely the present value (P), the future value (F) and the regular payment (R).

To calculate the future value (F) of an ordinary annuity certain, apply the following formula: $F_v = R\frac{(1+i)^n-1}{i}$

To calculate the present value or principle (P) of an ordinary annuity certain, apply the following formula: $P_v = R\frac{1-(1+i)^{-n}}{i}$

To calculate the future value (F) of an ordinary annuity due, apply the following formula: $F_v = R\frac{[(1+i)^n-1](1+i)}{i}$

To calculate the present value (P) of an ordinary annuity due, apply the following formula

$$P_v = R\frac{[1-(1+i)^{-n}](1+i)}{i}$$

To calculate the present value (P) of a deferred annuity, apply the following formula:

$$P_v = R(\frac{1-(1+i)^{-(m+n)}}{i} - \frac{1-(1+i)^{-m}}{i})$$

with m = the number of deferred periods.

1.9 Net Present Value Method

Read pp. 414 – 415 in your prescribed text.

2 Recommended Additional Reading

You should engage extensively with your Recommended Additional Reading as it is an essential part of the learning experience. It will give you different perspectives and engagement opportunities with different authors and schools of thoughts.

Willemse I. 2009. *Statistical Methods and Calculation Skills*. 3 ed. Juta. *http://youtu.be/RmroXgbRX2Y*

3 Recommended Digital Engagement and Activities

Your lecturer will provide you with the blog/Facebook details for your class from which you can obtain information or give comments. Digital activities will also be available on the blog.

4 Interactive Work Space

4.1 Izimvo Exchange 1

Complete and discuss the questions 33—37 on pp.420—421

4.2 Izimvo Exchange 2

You have just received an inheritance of R50 000 and decide to invest the money rather than spending it on luxury items. Do an investigation on how and where you will invest your money and what you will have in five years' time. Compare with your classmates to see whose money will grow the most. Do not take any income tax implications into account.

4.3 Activity 1

Purpose:
The following activities use the simple interest concepts. You must be able to calculate any of the formula components.

Task:

1. A student receives a R9 000 loan from a wealthy aunt in order to finance his four year study program. The terms are that the student repays his aunt in full at the end of 10 years at a simple interest rate of 5% per year. Determine the interest which must be paid on the loan.

2. Your Great-aunt Agatha dies and leaves you an inheritance of R60 000 which is to be paid to you in 10 months' time. The money is currently invested at 10% simple interest. How much will you receive ten months from today and how much interest is included in the sum?

3. A company that charges a customer 6% simple interest on invoices that are not paid on time receives an interest cheque for R15.65 from a customer whose account was 75 days overdue. What was the outstanding amount?

4. You borrow R325 000 from your father to start your own business. Your father charges simple interest on the loan which must be repaid in 18 months' time. If you repay your father R350 000 at the end of the loan period, what rate of simple interest is charged on the loan.

5. A sum of R1000 earns interest at a rate of 8% per year simple interest. How long will it take for the investment to grow to R2000?

Commentary Related to Activity Design:
The questions will most of the time state that simple interest must be used.

4.4 Activity 2

Purpose:
The purpose of this activity is to practise how to convert yearly interest and the time period to coincide with the compounded period. You will also use the compound methods to calculate interest, present values and future values.

Task:

1. If consumer prices are increasing at a rate of 9% per year compounded semi-annually, how much will an item costing R2.25 today cost in three years from today?

2. Thandi Sithole foresees cash flow problems ahead. She borrowed R10 000 18 months ago from Maxwell Moneyless at 15% per annum, compounded semi-annually. How much must she repay Maxwell today?

3. Andrew has just withdrawn all his money R721.81 from his bank account. If the bank paid 8% interest compounded quarterly, how much did Andrew deposit 12 years ago?

Commentary Related to Activity Design:
In compound interest calculations remember that the interest rate and term must coincide with the compounding period.

4.5 Activity 3

Purpose
The purpose of this activity is to practise if you can distinguish between the three types of annuities.

Task:

1. Mr Kubeka is planning to buy a new car in two years' time. He estimates it will cost him R140 000. His trade-in value on his current car will be R60 000 in two years' time. How much should Mr Kubeka save monthly in an account that pays 9% p.a. compounded monthly?

2. Beauty purchases a used car for R25 000. She pays a deposit of R5 000 and agrees to pay the balance back in 12 equal quarterly payments. Mr Mandla charges her 16% interest compounded quarterly. How much will she have paid back at the end of every quarter?

3. How much money must be deposited at the beginning of every 6-month period if the objective is to accumulate R5 000 within four years at an interest rate of 12% compounded semi-annually?

4. You have won R20 000 in the Lotto. You decide to invest the money in a retirement annuity that will pay you a quarterly income in advance at 8% interest for 10 years. How much will you receive quarterly for the next 10 years?

5. In 10 years' time the ABC Company will need R12 000 to replace machinery. What semi-annual deposits must be made at the end of every six months for the next 10 years if interest is calculated at 6% compounded semi-annually?

Commentary Related to Activity Design:
If there is a regular payment involved, you use the annuity methods to answer the questions. If the payment in the first interval falls at the end of that interval, use the ordinary annuities certain method. If the payment falls at the beginning of the first interval, and all intervals thereafter, then use the ordinary annuities due methods. If the payments start at a time period later than today, the annuity is deferred.

4.6 Revision Exercise 1

1. Julie bought make-up on her account to the value of R430. If interest is charged at a simple rate of 26% per annum, how much interest will be charged on her account at the end of the first month?

2. Tom borrows R8 900 at 24% compounded monthly, how much does he owe at the end of two years?
3. Your child will be ready to go to university in seven years' time and you estimate that you will need about R20 000 to cover educational expenses. If you start an instalment plan at 12% interest compounded monthly, what amount must you deposit at the end of each month in order to achieve your goal?

4.7 Revision Exercise 3

An example should clarify the point.

1. Calculate the effective rates of interest if the nominal rate is 15% per annum and interest is calculated:
   a. Yearly
   b. Half-Yearly
   c. Quarterly
   d. Monthly
   e. Daily

   What did you notice?

2. Calculate the effective interest rate if the nominal rate is 22% and the interest is calculated:
   a. Half-yearly
   b. Quarterly
   c. Monthly

3. Determine the nominal rate per year, compounded monthly for an effective rate of 19.56%

Solutions to Exercises

Izimvo Exchange 1

Complete and discuss the concept question on 420 - 421.

(The answers are given in the CD accompanying the textbook.)

Izimvo Exchange 2

You have just received an inheritance of R50 000 and decide to invest the money rather than spending it on luxury items. Do an investigation on how and where you will invest your money and what you will have in five (5) years' time. Compare with your classmates to see whose money will grow the most. Do not take any income tax implications into account.

(The students can do an internet search for a suitable investment or visit banks and investments houses for their current investment options.)

Activity 1

Questions:	**Model Solutions:**
1. A student receives a R9 000 loan from a wealthy aunt in order to finance his four year study program. The terms are that the student repays his aunt in full at the end of 10 years at a simple interest rate of 5% per year. Determine the interest which must be paid on the loan.	I = Prt = 9000(0.05)(10) = R4 500
2. Your Great-aunt Agatha dies and leaves you an inheritance of R60 000 which is to be paid to you in 10 months' time. The money is currently invested at 10% simple interest. How much will you receive ten months from today and how much interest is included in the sum?	I = Prt $= 60\,000(0.10)\left(\frac{10}{12}\right) = R5\,000$ A = 60 000 + 5 000 = R65 000
3. A company that charges a customer 6% <u>simple interest</u> on invoices that are not paid on time receives an interest cheque for R15.65 from a customer whose account was 75 days overdue. What was the outstanding amount?	I = Prt $15.65 = P\,(0.06)\left(\frac{75}{365}\right)$ P = R1 269.39

4.	You borrow R325 000 from your father to start your own business. Your father charges simple interest on the loan which must be repaid in 18 months' time. If you repay your father R350 000 at the end of the loan period, what rate of simple interest is charged on the loan.	I = Prt 25 000 = 325 000· r $\frac{18}{12}$ r = 5.13%
5.	A sum of R1 000 earns interest at a rate of 8% per year simple interest. How long will it take for the investment to grow to R2 000?	= Prt 1000 = 1000 (0.08) t t = 12.5 year or 12 years and 6 months

Activity 2

Questions:		**Model Solutions:**
1.	If consumer prices are increasing at a rate of 9% per year compounded semi-annually, how much will an item costing R2.25 today cost in three years from today?	A = P(1 + i)^n = 2.25(1 + $\frac{9}{200}$)^6 = R2.93
2.	Thandi Sithole foresees cash flow problems ahead. She borrowed R10 000 18 months ago from Maxwell Moneyless at 15% per annum, compounded semi-annually. How much must she repay Maxwell today?	= P(1 + i)^n = 10 000(1 + $\frac{15}{200}$)^3 = R12 422.97
3.	Andrew has just withdrawn all his money R721.81 from his bank account. If the bank paid 8% interest compounded quarterly, how much did Andrew deposit 12 years ago?	21.81 = P(1 + $\frac{8}{400}$)^{48} P = 279.01 rand

Activity 3

Questions:		Model Solutions:
1.	Mr Kubeka is planning to buy a new car in two years' time. He estimates it will cost him R140 000. His trade-in value on his current car will be R60 000 in two years' time. How much should Mr Kubeka save monthly in an account that pays 9% p.a. compounded monthly?	$80\,000 = R\dfrac{(1+\frac{9}{1200})^{24}-1}{\frac{9}{1200}}$ R = 3 054.78
2.	Beauty purchases a used car for R25 000. She pays a deposit of R5 000 and agrees to pay the balance back in 12 equal quarterly payments. Mr Mandla charges her 16% interest compounded quarterly. How much will she have paid back at the end of every quarter?	$P = R\left[\dfrac{1-(1+i)^{-n}}{i}\right]$ $20\,000 = R\left[\dfrac{1-\left(1+\frac{16}{400}\right)^{-12}}{\frac{16}{400}}\right]$ R = 2131.04
3.	How much money must be deposited at the beginning of every 6-month period if the objective is to accumulate R5 000 within four years at an interest rate of 12% compounded semi-annually?	$S = R\left[\dfrac{[(1+i)^n-1][1+i]}{i}\right]$ $5000 = R\left[\dfrac{[(1+0.06)^8-1][1+0.06]}{0.06}\right]$ R = 479.58
4.	You have won R20 000 in the Lotto. You decide to invest the money in a retirement annuity that will pay you a quarterly income in advance at 8% interest for 10 years. How much will you receive quarterly for the next 10 years?	$P = R\left(\dfrac{[1-(1+i)^{-n}][1+i]}{i}\right)$ $20\,000 = R\left(\frac{[1-(1+0.02)^{-40}][1+0.02]}{0.02}\right)$ R = 716.78
5.	In 10 years' time the ABC Company will need R12 000 to replace machinery. What semi-annual deposits must be made <u>at the end</u> of every six months for the next 10 years if interest is calculated at 6% compounded semi-annually?	$S = R\dfrac{(1+i)^n-1}{i}$ $12\,000 = R\frac{(1+0.03)^{20}-1}{0.03}$ R = 446.59

Revision Exercise 1

Questions:	Model Solutions:
1. Julie bought make-up on her account to the value of R430. If interest is charged at a simple rate of 26% per annum, how much interest will be charged on her account at the end of the first month?	= Prt = 430·(0.26) $\frac{1}{12}$ = R9.32
2. Tom borrows R8 900 at 24% compounded monthly, how much does he owe at the end of two years?	A = P(1 + i)^n A = 8 900(1 +0.02)^{24} = R14 315.09
3. Your child will be ready to go to university in seven years' time and you estimate that you will need about R20 000 to cover educational expenses. If you start an instalment plan at 12% interest compounded monthly, what amount must you deposit at the end of each month in order to achieve your goal?	$S = R\frac{(1+i)^n - 1}{i}$ $20\,000 = R\frac{(1+0.01)^{84} - 1}{0.01}$ R = R153.05

Revision Exercise 2

Questions:	Model Solutions:
A couple estimate that they can afford a mortgage payment of R325 at the end of every month. They can obtain a 20-year mortgage at an interest rate of 9% compounded monthly. What is the largest mortgage loan they can afford?	$P = R\left[\frac{1-(1+i)^{-n}}{i}\right]$ $P = 325\left[\frac{1-(1+0.0075)^{-240}}{0.0075}\right]$ P = R36 122.11

Revision Exercise 3

Questions:	**Model Solutions:**
1. Calculate the effective rates of interest if the nominal rate is 15% per annum and interest is calculated: 1.1 Yearly 1.2 Half-Yearly 1.3 Quarterly 1.4 Monthly 1.5 Daily What did you notice?	a. I no difference b. 15,56% p.a c. 15,87% p.a d. 16,08% p.a e. 16,18% p.a Effective rate increases
2. Calculate the effective interest rate if the nominal rate is 22% and the interest is calculated: 1.1 Half-yearly 1.2 Quarterly 1.3 Monthly	a. 23.21% b. 23.88% *c.* 24.31%
3. Determine the nominal rate per year, compounded monthly for an effective rate of 19.56%	738: 12 (x,y) 19,56 2nd F APR 17,998% 733: 12 2nd F APR 19,56 2nd F EFF

Formula sheet Quantitative Techniques A and B (QUAT5112 & QUAT6211)

1. Measures of central tendancy

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{n}$

Median = $O_{me} + \frac{c\left(\frac{n}{2} - (f<)\right)}{f_{me}}$

Mode = $O_{mo} + \frac{c(f_m - f_{m-1})}{2f_m - f_{m-1} - f_{m+1}}$

2. Measures of Dispersion

R = $x_{max} - x_{min}$; R = upper class boundary of highest class – lower class boundary of lowest class

$s^2 = \frac{\sum x^2 - n\bar{x}^2}{n-1}$

$s^2 = \frac{\sum f x^2 - n\bar{x}^2}{n-1}$, and $s = \sqrt{s^2}$

Interquartile range = $Q_3 - Q_1$

Quartile Deviation = $\frac{Q_3 - Q_1}{2}$

cv = $\frac{s}{\bar{x}}$ x 100%

$Sk_p = \frac{3(\text{Mean} - \text{median})}{standard\ deviation}$; or $Sk_p = \frac{(\text{Mean} - \text{Mode})}{Standard\ Deviation}$

$Q_1 = O_{q1} + \frac{c[\frac{n}{4} - f(<)]}{f_{q1}}$ and $Q_3 = O_{q3} + \frac{c[\frac{3n}{4} - f(<)]}{f_{q3}}$

3. Probability distributions

P (r) = ${}_nC_r p^r q^{(n-r)}$

Mean: μ = np

Standard deviation: $\sigma = \sqrt{npq}$ where q = 1 – p

P (x) = $\frac{e^{-a}a^x}{x!}$ For x = 0, 1, 2, 3,

Mean = a

Standard deviation = $\sqrt{a}$

4. Confidence Interval

$$\bar{x} - z\frac{\sigma_x}{\sqrt{n}} \leq \mu \leq \bar{x} + z\frac{\sigma_x}{\sqrt{n}}$$

$$\bar{x} - t_{(n-1)}\frac{s_x}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{(n-1)}\frac{s_x}{\sqrt{n}}$$

$$p - z\sqrt{\frac{pq}{n}} \leq \pi \leq p + z\sqrt{\frac{pq}{n}}$$

$$(\overline{x_1} - \overline{x_2}) - z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\overline{x_1} - \overline{x_2}) + z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(\overline{x_1} - \overline{x_2}) - t\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\overline{x_1} - \overline{x_2}) + t\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(p_1 - p_2) - z\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}} \leq \pi_1 - \pi_2 \leq (p_1 - p_2) + z\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$$

5. Regression

$\hat{y} = a + bx$, where $b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$

And $a = \frac{n\sum y - b\sum x}{n}$

6. Hypothesis testing

$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ or $z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ or $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

$z = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$, or $z = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

$t = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$, or $s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}}$$

$$z = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2}}}$$

$z = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n_1} + \frac{\hat{\pi}(1-\hat{\pi})}{n_2}}}$ where $\hat{\pi} = \frac{x_1 + x_2}{n_1 + n_2}$

$X^2 = \sum \frac{(f_o - f_e)^2}{f_e}$

7. Index numbers

Price relative $= \frac{p_1}{p_0} x\ 100\%$

Laspeyres price index $= \frac{\sum p_1 q_0}{\sum p_0 q_0}$ x100%

Paasche price index $= \frac{\sum p_1 q_1}{\sum p_0 q_1}$ x100%

8. Time series

$\hat{y} = a + bx$, where $a = \frac{\sum y}{n}$ and $b = \frac{\sum xy}{\sum x^2}$

9. Linear Correlation

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

$$R_s = 1 - \frac{\partial \sum d^2}{n(n^2-1)}$$

10. Financial Mathematics

$$F_v = P_v(1 + in)$$

$$F_v = P_v(1 + i)^n$$

$F_v = P_v\left(1 + \frac{i}{k}\right)^{nk}$ where k = number of compounding periods in a year.

$$r = (1 + \frac{i}{m})^m - 1$$

$$F_v = R\frac{(1+i)^n - 1}{i}$$

$$P_v = R\frac{1 - (1+i)^{-n}}{i}$$

$$F_v = R\frac{[(1+i)^n - 1](1+i)}{i}$$

$$P_v = R\frac{[1 - (1+i)^{-n}](1+i)}{i}$$

$$P_v = R(\frac{1 - (1+i)^{-(m+n)}}{i} - \frac{1 - (1+i)^{-m}}{i})$$

Bibliography

Johnson, RR and Kuby, PJ. 2011. *Elementary statistics*. 11th edition. Pacific Grove, (CA): Duxbury.

Lombaard, C, Van der Merwe, L, Kele, T and Mouton, S. 2010. *Elementary statistics for business and economics*. 1st edition. Cape Town: Heinemann Pearson Publishing.

Ratio, Rate and Proportion 1. 2009. [Video online]. Available at: http://youtu.be/avjovNP60Nk [Accessed 24 May 2014].

Sullivan, M. 2010. *Fundamentals of statistics*. 3rd edition. Boston, (MA): Pearson's Education.

Triola, MF. 2009. *Elementary statistics*. 11th edition. Boston, (MA): Pearson's Education.

Weaver, S. 2012. *STAT 141 - Basic statistical methods.* [Video online]. Available at: http://youtu.be/7Mfa1MbVhcc [Accessed 24 May 2014].

Wegner, T. 2012. *Applied business statistics methods and Excel-based application*. 3rd edition. Pretoria: JUTA. ISBN: 9780702177743

Willemse, I. 2009. *Statistical methods and calculation skills*. 3rd edition. Pretoria: Juta.

Intellectual Property

Plagiarism is any use of the words, ideas or images of another person without acknowledging the source using the required conventions. Below is a description of plagiarism and referencing. Please make sure that you are familiar with this information before attempting your assignment.

Introduction to Referencing and Plagiarism

What is 'Plagiarism'?

'Plagiarism' is the act of taking someone's words or ideas and presenting them as your own.

What is 'Referencing'?

'Referencing' is the act of referring to or consulting. A 'reference' is a publication or passage from a publication that is referred to.

Referencing is the acknowledgment of any work that is not your own, but is used by you in an academic document. It is simply a way of giving credit to and acknowledging the ideas and words of others.

When writing assignments, students are required to acknowledge the work, words or ideas of others, through the technique of referencing. Referencing occurs in the text at the place where the work of others is being cited, and at the end of the document, in the bibliography.

Cumming (2007) describes the bibliography as a list of all the work (published and unpublished) that a writer has read in the course of preparing a piece of writing. This includes items that are not directly cited in the work.

A reference is required when you:

- Quote directly: when you use the exact words as they appear in the source;
- Copy directly: when you copy data, figures, tables, images, music, videos or frameworks;
- Summarise: when you write a short account of what is in the source;
- Paraphrase: when you state the work, words and ideas of someone else in your own words.

It is standard practice in the academic world to recognise and respect the ownership of ideas through good referencing techniques. However, there are other reasons why referencing is useful.

Good Reasons for Referencing

It is good academic practice to reference because:

- It enhances the quality of your writing;
- It demonstrates the scope, depth and breadth of your research;
- It gives structure and strength to the aims of your article or paper;
- It endorses your arguments;
- It allows readers to access source documents relating to your work, quickly and easily (Neville, 2007, p.7).

Sources

The following would count as 'sources':

- Books,
- Chapters from books,
- Encyclopaedia,
- Articles,
- Journals,
- Magazines,
- Periodicals,
- Newspaper articles,
- Items from the Internet (images, videos, etc.),
- Pictures,
- Unpublished notes, articles, papers, books, manuscripts, dissertations, theses, etc.,
- Diagrams,
- Videos,
- Films,
- Music,
- Works of fiction (novels, short stories or poetry).

What You Need to Document from the Hard Copy Source You are Using

(Not every detail will be applicable in every case. However, the following lists provide a guide to what information is needed.)

You need to acknowledge:

- The words or work of the author(s),
- The author(s)'s or editor(s)'s full names,
- If your source is a group/ organisation/ body, you need all the details,
- Name of the journal, periodical, magazine, book, etc.,
- Edition,
- Publisher's name,
- Place of publication (i.e. the city of publication),
- Year of publication,
- Volume number,
- Issue number,
- Page numbers.

What You Need to Document if you are Citing Electronic Sources

- Author(s)'s/ editor(s)'s name,
- Title of the page,
- Title of the site,
- Copyright date, or the date that the page was last updated,
- Full Internet address of page(s),
- Date you accessed/ viewed the source,
- Any other relevant information pertaining to the web page or website.

Referencing Systems

There are a number of referencing systems in use and each has its own consistent rules. While these may differ from system-to-system, the referencing system followed needs to be used consistently, throughout the text. Different referencing systems cannot be mixed in the same piece of work!

A detailed guide to referencing, entitled Referencing and Plagiarism Guide is available from your library. Please refer to it if you require further assistance.

When is Referencing Not Necessary?

This is a difficult question to answer – usually when something is 'common knowledge'. However, it is not always clear what 'common knowledge' is.

Examples of 'common knowledge' are:

- Nelson Mandela was released from prison in 1990;
- The world's largest diamond was found in South Africa;
- South Africa is divided into nine (9) provinces;
- The lion is also known as 'The King of the Jungle'.
- $E = mc^2$
- Jan Van Riebeeck was the first person to settle in the Southern Cape.

Usually, all of the above examples would not be referenced. The equation $E = mc^2$ is Einstein's famous equation for calculations of total energy and has become so familiar that it is not referenced to Einstein.

Sometimes what we think is 'common knowledge', is not. E.g. the above statement about Van Riebeeck is only partly true – he was the first European to settle in the Cape. It was, however, not an 'uninhabited' area when he got there. The Khoisan, the original inhabitants of the Cape, had been living in the area for some time. It is not entirely accurate then to claim that Van Riebeeck was the first inhabitant. (Crampton, 2004, p.57)

It is thus generally safer to always check your facts and try to find a reputable source for your claim.

Important Plagiarism Reminders

The IIE respects the intellectual property of other people and requires its students to be familiar with the necessary referencing conventions. Please ensure that you seek assistance in this regard before submitting work if you are uncertain.

If you fail to acknowledge the work or ideas of others or do so inadequately this will be handled in terms of the Plagiarism Policy (available in the library) and/ or the Student Code of Conduct – dependent on whether or not plagiarism and/ or cheating (passing off the work of other people as your own by copying the work of other students or copying off the Internet or from another source) is suspected.

This campus offers individual and group training on referencing conventions – please speak to your librarian or ADC/ Campus Co-Navigator in this regard.

<u>Reiteration of the Declaration you have signed:</u>

1. I have been informed about the seriousness of acts of plagiarism.
2. I understand what plagiarism is.
3. I am aware that The Independent Institute of Education (IIE) has a policy regarding plagiarism and that it does not accept acts of plagiarism.
4. I am aware that the Plagiarism Policy and the Student Code of Conduct prescribe the consequences of plagiarism.
5. I am aware that referencing guides are available in my student handbook or equivalent and in the library and that following them is a requirement for successful completion of my programme.
6. I am aware that should I require support or assistance in using referencing guides to avoid plagiarism I may speak to the lecturers, the librarian or the campus ADC/ Campus Co-Navigator.
7. I am aware of the consequences of plagiarism.

Please ask for assistance prior to submitting work if you are at all unsure.