Kingdom of Saudi Arabia

مدينة الملك عبدالعزيز
للعلوم والتقنية KACST

Business Research Techniques with Statistical Software Usage and Interpretation

Dr. Abdulmonem H. Alzalabani
Dr. G. Sundaran Sagaran
Mr. Manuel G. Cadongonan

Riyadh
2014 AD - 1435 H

Kingdom of Saudi Arabia

مدينة الملك عبدالعزيز
للعلوم والتقنية KACST

Business Research Techniques with Statistical Software Usage and Interpretation

Dr. Abdulmonem H. Alzalabani
Dr. G. Sundaran Sagaran
Mr. Manuel G. Cadongonan

Riyadh
2014 AD - 1435 H

King Fahd National Library Cataloging-in-Publication Data

Alzalabani, Abdulmonem
business research techniques with statistlcal
software usage and interpretation. / Abdulmonem
Alzalabani .- Riyadh, 2012

161p : 25cm

ISBN: 978-603-8049-54-9

1-Business - Computer programs
2- Business - Methodology I-Title

650.028 dc 1434/793

L.D.no. 1434/793
ISBN: 978-603-8049-54-9

King Abdulaziz City for Science and Technology
P.O. Box 6086 Riyadh 11442
Kingdom of Saudi Arabia
Tel. 011 4883444 - 011 4883555 Fax: 011 4883756
Website: www.kacst.edu.sa
E. Library: kacst.edu.sa/ar/about/publications
Email: awareness@kacst.edu.sa


بسم الله الرحمن الرحيم

Table of Contents

Introduction

This era is characterized by the tremendous and rapid science and technology advancement in all areas of knowledge as well as by the ever-growing number of discoveries and inventions in different fields and applications. This scientific leaps and breakthroughs, witnessed by the human civilization, have radically changed all aspects of human life today.

Because of the significant accumulation of these different types of sciences and their applications, some concepts, like scientific awareness, scientific enlightenment, and scientific education, have surfaced in a context targeting the scientific growth of individuals for the sake of their own development. These new concepts intend to raise public awareness and understanding of science and its products as well as knowledge and its developments. They also shed light on the advancement of technological products and their usage.

For these reasons and others, taking an interest in what is known as "scientific culture" has noticeably grown in importance. This concept has emerged on the public cultural scene as an urgent necessity to configure awareness citizens of the scientific developments around them, especially after the massive knowledge explosion that changed many human behavioral and intellectual patterns, after the presence of science with its theories and techniques in almost all human daily activities.

The National Plan for Science, Technology and Innovation (NPSTI) in the Kingdom of Saudi Arabia (KSA) takes into account the importance of spreading scientific awareness and culture to connect the society as a whole to the advancement of science and its basic concepts, and then build a scientific culture that encourages scientific research and technological development in KSA.

Since its establishment, King Abdulaziz City for Science and Technology (KACST) has taken an active interest in raising scientific awareness and spreading the culture of science among all members of society. This fact reflects KACST continuous keenness to produce a wide array of scientific publications, such as journals, reports, scientific books, booklets and leaflets targeting the general readers and beneficiaries of these types of publications. KACST also organizes many activities and events including the Science and Technology Week, as well as hosting scientific lectures, seminars, and conferences to contribute in educating the society members and the advancement of their scientific knowledge as well as the enrichment of the Arab world libraries and the Arabic digital content.

This book is another contribution by KACST to libraries and specialists with the aim of enriching human knowledge and raising public awareness of science. As it has always been, KACST will continue to support authors to publish their scientific ideas and thoughts to promote knowledge production as well as appreciation of scientific achievements of others.

President of King Abdulaziz City for Science and Technology

Dr. Mohammed ibn Ibrahim Al-Suwaiyel

Preface

Business management is multifaceted in nature. It has different functional areas and there is a necessity that students should get a feel of real-life business situations for them to be equipped with skills to embark upon a career in business.

Business management programs give emphasis to both practice and theory. Business students need to expose themselves to different practical aspects of actual business like doing critical analyses, conducting researches, preparing reports, doing presentations etc.

Hence, this guide on business research will greatly support research students to undertake business research surveys confidently. Our years of experience with business research have proved that not many simple-to-use research books are at hand for a quick reference. Many of the books that are available are cumbersome to read when looking for needed information as everything is discussed at length. This actually led us to think of coming up with a straightaway reference in business research, with a simple limited framework in mind. But as we progressed, the frame further expanded, and here we are with a final outcome. With much pleasure, we present to you this book, which is a culmination of our years of teaching experience with business students.

About the Book

There are many books in the market that deal with business research. These books mostly focus on building the theoretical foundation before embarking on the research aspect. Furthermore, these books are cumbersome for first timers who prefer a simple book that takes them step by step towards their endeavor. This is where *Business Research Techniques with Statistical Software Usage and Interpretation* helps; it can be your instructive and practical companion to overcome the rigorous task of writing a business research.

The topics are systematized to jibe with the process of doing research, thus, you go through the steps smoothly. It is easy to understand the fundamental principles since they are coupled with illustrations. But what really stands out in this book are the presentations of how data can be analyzed and interpreted using various tools. There are even case examples which are examined point-by-point and they may be duplicated as studies situated in other contexts.

Chapter 1 introduces you to the rudiments of research writing and assists you in citations and construction of a questionnaire. *Chapter 2* presents the core of research, i.e., organizing data employing the appropriate statistical tools using Microsoft Excel. Various models treated with the SPSS (Statistical Package for the Social Sciences) are provided in *Chapter 3*.

You can try the cases themselves and check your facility in data analysis. There are more than enough samples for you to explore. Besides, the authors made an effort to avoid a purely technical discussion by employing straightforward language.

This textbook is designed to cater to all types of research enthusiasts but with special focus on issues and trends in the business market today.

Arm yourself with advanced skills in business research with this *book*. Once you acquire the most important strategies, you will have the most useful results.

Acknowledgements

Producing this kind of book takes the cooperation of many. We are thankful to all those who have directly or indirectly contributed to the birth of this book, especially students and staff members of Industrial Management Technology Department of Yanbu Industrial College who sparked the idea of this creation in our minds.

A very special thank you goes to the reviewers who generously gave of their time, either in reading the drafts or in giving constructive suggestions throughout the preparation of this book.

Ultimately, any errors of omission or commission are ours, and we bear the responsibility for them. Likewise, we leave a room for enrichment of the book's content that is why we solicit comments and recommendations from the readers.

1 HOW TO DO RESEARCH

By the end of this part you should be able:

- *To generate ideas that will help in the selection of a suitable research topic;*
- *To draft a research proposal;*
- *To search for information;*
- *To organize dissertation;*
- *To reference the literature accurately;*
- *To use right tool to collect research data;*
- *To prepare research questionnaire;*
- *To use right distribution techniques.*

"You see things; and you say, 'Why?' But I dream things that never were; and I say, 'Why not?"

-George Bernard Shaw

1.1 Getting Started

Students of business will be exposed to some sort of research writing alongside the theory being introduced before they graduate. Be it a minor research essay or mini thesis, a research writing task is a good learning ground for students to think independently. Conducting a research is not that easy especially for first timers, as it has its own methodology that need to be followed. Below is a brief outline of how to go about doing a research.

Selection of a Tentative Topic for Research

Usually at the undergraduate level, the lecturers, to relieve the students from the burden of searching an appropriate topic, will provide the research topics themselves. If no readymade topic is available or prescribed by your lecturers, take considerable care in selecting your topic. Go for a topic that is of interest to you and within the scope of the concerned course. To sustain your efforts, it is critical that you have a strong interest in the selected topic. Research requires a good amount of enthusiasm to be transformed into effort and this will be possible only if the topic is of interest to you and challenges you. At the same time, the topic should be manageable also. If you have a topic with a very wide scope, in your mind, narrow it down to something specific by seeking the assistance of knowledgeable persons, use mind mapping concept or relevance tree approach. However, don't make it too narrow, as it will not guarantee you with enough sources of data.

Search for Information / Literature Review

Read Widely

- If your topic is related to a particular course, read the relevant part of the course material/text book first to have a good idea about the theoretical background of the topic. For example, if you are planning to do a research on 'managerial motivation in public sector enterprises', you should have adequate knowledge of the basic concepts and different theories of motivation, in order to do the research effectively.
- Search for recent literature on the topic and closely related topics to your choice. Try searching databases in your library, based on keywords. If you are consulting business journals, check the references list that will lead to additional resources related to your topic.

Check Out Print Materials Like:

- Previous research related to your topic
- Business journals and publications
- Newspapers and magazines
- Government publications
- Other recognized agency reports, bulletins and publications
- Statistical information provided by various agencies
  [e.g. Chambers of Commerce, Trade Associations, Statistical Bureaus etc.]
- Educational publications related to your topic

Note: Refer to Appendix VIII, page 192 on how to identifying key words and how they are formed, that may help the readers to do the search and finding on related studies.

Search the World Wide Web

World Wide Web is a boon to a researcher nowadays. It provides you with an ocean of data, but be very careful when you pick out information from the web, as anybody can put any information, which may not be authentic and dependable, in the web. Ensure the genuineness of the information before you rely on it. Look at the organization that has hosted the site, the name and other pertinent details of the person who has posted the information. Better rely only on established websites.

Use of Search Engines

Use *general search engines* like Yahoo, Google, Alta Vista and Bing (MSN) to look for key words.

Remember the use of words like 'AND', 'OR' and 'NOT when using these search engines. They help you with union, intersection and exclusion of search words.

For Example '*work place stress* AND *violence'* searches articles with both *work place stress* and *violence* together. The use of AND has narrowed the search.

But the use of '*work place stress* OR *violence*' searches all articles with combination of work *place, violence and both*. The use of OR has widened the search.

For '*work place stress* NOT *violence'*, the use of NOT will exclude the term *violence* from search.

At times *'truncation'* of keys words can help in search for articles with different words. Like if you type Marke*, the search engine will select articles with market, marketing and marketers etc. You can also pick up words with different spellings with the use of '?' like in '*Organi?ation'* where both words with spelling '*Organisation'* and '*Organization*' are selected.

Take Stock of What You Have Collected

It is quite possible that an idea from literature review may very well give you a revised or sometimes a new perspective. Keep in mind that your research should not be influenced substantially or led by the literature you have reviewed. *YOUR RESEARCH SHOULD REPRESENT YOUR ORIGINAL THINKING*. Seek advice from the research guide / teacher / advisor whenever needed to ensure the feasibility of the research.

BY NOW, YOU MIGHT HAVE TAKEN A ROUGH IDEA OF WHAT IS TO BE DONE AND YOU MAY PROCEED TO STATE YOUR HYPOTHESIS or your RESEARCH QUESTION/s.

In very simple words, hypothesis means a tentative conclusion about the research. It is the researcher's duty to test the hypothesis in order to accept or reject it. To test your hypothesis, you have to collect data and analyze it. Data collection techniques and statistical tools for data analysis are discussed in detail in Chapter 2 and Chapter 3.

The end product of the research process will be a research report. Preparation of a research report requires much care and attention, as it is the medium for communicating research results to the outside world.

A good research report should achieve the following objectives:

- Effectively communicate the findings of the research project
- Provide logical interpretations of those findings.
- Illustrate the credibility of the research project

1.2 Dissertation Proposal

Prior to any research, there is a necessity at times to come out with a proposal. A proposal is a ***blueprint*** of what you are planning to do in your actual research. The supervisor needs to approve the proposal before you even embark upon the research. To avoid having your proposal rejected, you should take note of the following guidelines.

The Stages/Information Needs

- **Title**

  Clearly indicate title, which reflects your content and approach.

- **Aims of the Dissertation**

  A general introductory, that reflects investigative aims and leads to the statement of specific aims, which are to be achieved, should be written. Your introductory part needs to be problem-based approach.

- **Relation to Previous Work**

  You need to place the work in relevant context. Remember no investigation starts from scratch. So it is not sufficient to say there has been little work done in this area before.

  Relationship can be:

  Subject-based – *what has been done in this subject before*

  Problem-based – *how has this kind of problem been approached in previous studies*

  Methodology-based – *which methods have been deemed appropriate to this kind of studies in previous work*

  You need to move beyond these previous works in some clearly identified ways for originality in your work.

- **Methodology**

  This section details the way in which the aims of the investigation are to be achieved. It involves a description of the method to be used, and also a discussion of the logic of those methods.

  You need to include consideration of alternatives and demonstrate the suitability of the approach being proposed. It goes beyond merely itemizing a data-collection approach.

- **Plan of Work**

  Explain the practical way in which the project is to be carried out. It includes details of means for ensuring that the aims are achieved. Include a time schedule. You can use a Gantt chart if you are familiar with it. Note: Refer to Appendix VII, page 191 if you are not familiar with Gantt chart.

- **Resources Required**

  This section refers to other sources of factual material and the kinds of access that will be required together with the arrangements for making that possible.

- **Bibliography**

  A full bibliography of all items referred to in the proposal should be provided. Only those items actually referred to in the proposal should be included. Use consistent and recognized convention.

- **Forms of Presentation**

  It will be assumed that your dissertation will be presented in the traditional format outlined unless you specify in this section any deviation from the usual.

Are You Ready for Research Proposal?

Look at the following pointers.

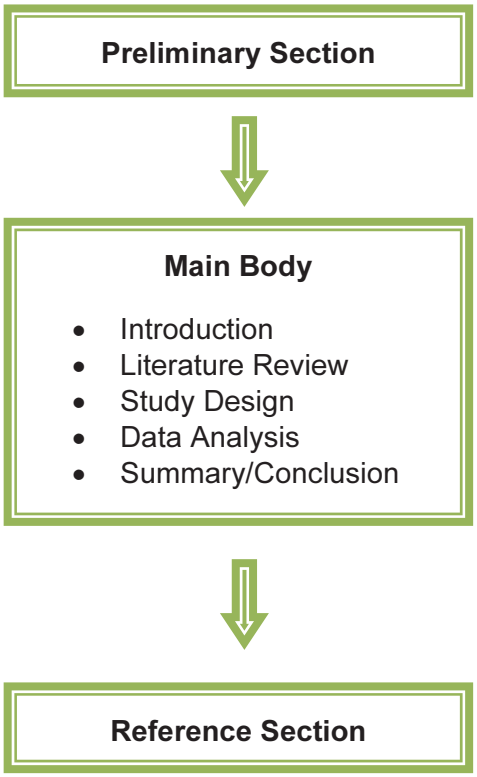
- *Have you gone broadly and deeply in the area of the research topic you have selected?*
- *Have you spent enough time thinking critically about your research topic?*
- *Did you discuss your research topic with others who are knowledgeable?*
- *Have you found out how people in other disciplines think about your research topic?*
- *Do you feel ready to begin writing your research proposal?*

If your answer is **'YES'** to all the questions, then you are confident and comfortable with the research topic. Begin the proposal writing. If **'NO'**, review the problem areas you have identified once more.

Now the Question is How to Write a Good Research Report/Format

The most widely used research report format is given below.

If no specific guidelines are provided, there are certain elements that must be considered when preparing the report. These elements can be grouped into three sections as shown below,



Note: See Appendix I, page 165 for research process

1.3 Organization of Dissertation

A. Preliminary Section

Title Page	*Tell what the paper investigated, be specific.*
Abstract	*Give a summary of hypotheses, methods, and major findings.* *Refer to Appendix II, page 166.*
Acknowledgment	*If any specific help was received, acknowledge it.*
Table of Contents	*Give the details of contents along with respective page numbers.*
List of Tables	*Provide it, if there is any.*
List of Figures	*Provide it, if there is any.*

B. Main Body

Introduction	Short reports on: *a. Statement of the problem* – *a general introduction.* *b. Significance of the problem* – *comments on why the issue is being investigated.* *c. Purpose* – *why a better understanding is important.* *d. Hypothesis/Research Question* – *the question being investigated and the expected results.* *e. Assumptions* – *everything that is taken to be true, for the investigation to succeed.* *f. Limitations* – *areas that limit the accuracy of study or generalization..* *g. Definitions* – *any terms and concepts used.*

Literature Review	State a necessary background for readers to understand the study.
Study Design	Include necessary information to replicate the study. *a. Describe design/procedure* *– step by step process of what was done.* *b. Data source* *– how data were collected.* *c. Sampling procedure* *– limitation and how a representative sample was obtained.* *d. Methods/instruments of data gathering* *– the procedure for obtaining data collected, forms, and manner in which they were recorded.* *e. Statistical treatment* *– mathematical procedure used to analyze data and determine the significance of results.*
Data Analysis	Explain the patterns observed in data, use tables and figures. ***A. Begin with exploratory analysis:*** *i. Show specific values* *– use a table to summarize the data.* *– group data into categories.* *ii. Show highest and lowest values* *– emphasize values using alternate fonts.* *– use charts, bar, histogram.* *iii. Show trends* *– explore trends using line charts, time series.* *iv. Show proportion* *– use pie charts.* *v. Show distribution of values* *– use box plots.* *vi. Select a combination most appropriate from the above for your Analysis.* ***B. Next go for Comparing variables:*** *i. Show specific values, interdependence* *– cross-tabulation, and contingency table.* *ii. Compare highest and lowest values* *– use multiple bar charts.*

Data Analysis	*iii. Compare proportion* – *percentage component bar chart or two or more pie charts.* *iv. Compare trends* – *multiple line chart.* *v. Compare totals* – *stacked bar chart.* *vi. Compare proportions and totals* – *comparative proportional pie charts.* *vii. Compare distribution of values* – *use multiple box plots.* *viii. Show relationship between cases for variables* – *use scatter plots.* *ix. Select a combination most appropriate from the above for your Analysis.* ***C. Use of Descriptive statistics*** ***- exploratory data from above can be numerically described using the following statistics:*** *i. Central tendency* *– description of the mean, mode and median values of your observation.* *ii. Dispersion* *– use of range, inter-quartile range and standard deviation.* ***D. Examine relationships, differences and trends with statistics:*** *i. Test for significant relationships and differences* *– use of p-value, chi square test, t-test, ANOVA, correlation, regression analysis and other tests.* *ii. Select tests applicable to your research, those that reflect your research question/s and objectives.*
Summary and Conclusion	Condense previous sections. *a. Restatement of problem* *– a short reiteration of the problem.* *b. Description of procedure* *– brief reiteration of important elements of the design of the study.* *c. Major findings* *– presentation of final analysis, hypothesis/research question restated, decision on hypothesis/research restated.*

	d. Conclusions – *comments on implication of findings presented.* e. *Recommendations* – *knowledge and experience gained, means and areas for further improvement, other possible hypothesis for future research.*

C. Reference Section

Bibliography	List references for the works cited in the text.
Appendix	Show tables, figures, or other materials not central to analysis but need to be included.

DOs and DONTs in Research Report Writing

- Avoid abbreviations.

  For example avoid using words like "Reps & Certs" which are not commonly understood for words like "Representations and Certification".

- Avoid one-sentence paragraphs.

- Avoid flowery language.

  Research is factual report and you need to capture the information accurately so avoid flowery language.

- Maintain consistency in using same tense in a sentence.

  For example, use past tense in reporting research of others and present perfect for research of primary importance to you and the research in the area is still ongoing. Present tense can be used to indicate your general position relative to reported research, for generally accepted scientific fact and for statement made by you.

- Write in simple sentences as far as possible.

  If it is necessary to use compound sentences, strike a balance between the two.

- Avoid repetitions.
- Try not to start or end a sentence with a conjunction unless it is necessary (but, and, because, however).
- Don't end a sentence with a proposition.

  Proposition is similar to a hypothesis, but its main purpose is to suggest a link between two concepts in a situation
- Don't write, not even a paragraph, in capital letters.
- Identify a person by family name, when you mention him/her for the first time in the report. Refer to referencing guide on page 22.
- Avoid quoting secondary works.

  Sometimes you may find an author cites a useful quotation from another text, which you want to reference in your work. Avoid such reference and where possible go for the original article which is referenced in the book or article to avoid misrepresentation.
- Try to the maximum extent possible, not to put facts/data in point form.

  Try to work ideas into sentences in order to maintain a flow.
- Report should be written in third person; do not use 'I' and 'My'.
- Try to avoid giving advice for readers or to express conclusions or speculations that extend far beyond researched samples.

  Do not give advice like "All the motorists of Yanbu should be given training on road safety'' or speculation like "High labor turnover in the industry *may be* due to job dissatisfaction.''
- Avoid unsubstantiated or overly general statements/statistics.

  Majority of the population in Saudi Arabia have an educational background of degree or more.
- Capitalize when referring to a specific concept. (e.g.: **G**overnor {Name} but not governors. **P**arliament but not parliamentary.)
- Always make sure your grammar and spelling are correct.

Structure and Layout of the Report

Before you print your final submission copy, make sure you follow the pointers below:

- Paper
  - *Always use good quality white paper.*
  - *The standard size of paper used for reports is 8 ½" x 11".*
- Margins
  - *Leave a margin of at least 1 inch (2.5 cm.) at the top, bottom, and sides of each and every page you print.*
- Spacing
  - *All narratives should be double-spaced.*
  - *Leave one space between words and after a comma, colon and semi-colon.*
  - *Do not leave a space in front of a punctuation mark.*
- Indentation
  - *You may or may not indent the paragraphs as you wish; but be consistent whichever one you choose to use.*
  - *If you wish to indent, indent ½" from the left margin.*
- Page numbering
  - *Put page numbers in Arabic numerals either in the footer or header.*
  - *Don't do any fancy work like -1-, *1*, (1), #1 etc. to make it attractive. Just write the number. There is no need to put a full stop after the number.*
- Fonts
  - *Always use simple and easy to read fonts.*
  - *Font size should be more than 10 points. Twelve points is regarded as standard and is our recommendation.*
  - *Use a single type of font for the whole report.*
- Printing
  - *You can print the report or type the report using a typewriter.*
  - *Always start a new chapter/section in a new page.*
  - *Print only on one side of the paper.*
  - *Give proper headings and subheadings.*
  - *Ensure that there is a logical flow of ideas for the whole narrative.*

1.4 Literature Review

Literature review is always a standard chapter in dissertation. It can be defined as the effective evaluation of what has already been written or said on a certain field or topic. It refers to gathering any information relevant to a specific subject stated in books, journal articles, newspaper articles, government reports, theses, etc. The idea behind this work is to carry on from where others have already reached. This is to avoid reinventing the wheel. It involves "reading what other people have written about your area of interest" (Bell, 1993). It is not necessary that the writer agrees or disagrees with this information.

Literature review actually comes in the beginning of your work although it can be a continuous practice until the end.

The Purposes of Literature Review include the followings

- To provide a solid background and justification for the research undertaken (Bruce, 1994).
- To keep the readers up to date with what is current in the field.
- To state some evidence that you are aware of the last knowledge on the subject you are going to write about.
- To trace the intellectual progression of the field.
- To give interpretation of materials and discuss the debates to support the current research's findings.
- To provide a framework for establishing the importance of the study, as well as a benchmarking for comparing the results of a study with other findings.
- To identify gaps in the existing literature in concerned areas of study.

The Contents of the Literature Review

The focus of literature review is to summarize the arguments and ideas of others without adding new contributions. However, the writer should keep his voice in front and center such as starting and ending the paragraphs with his own ideas and his own words.

Literature review contains a critical assessment of the relevant literature to a particular field or topic. This part of the dissertation signifies that the writer has grasped the subject very well. The sources of literature review include anything relevant to the topic: book, journal article, newspaper article, government reports, theses and dissertations. Literature review provides a background to the study being proposed like what has already been

established on the topic, what their strengths and weaknesses are, what the dominant theories are, who the key scholars are, and any other relevant information.

Literature review usually has an introduction, body, and conclusion. Introduction provides a quick idea of the topic, the body contains the discussion, and the conclusion identifies what the writer has drawn from this literature review. Paragraphs should be well formed and structured logically. In other words, there should be a logical flow of ideas with critical analysis. The evaluation of the literature should lead to the research question.

Processes of Writing Literature Review

To write a good literature review, you need to have knowledge of the use of indexes and abstracts in order to search for the references effectively. Moreover, as there are thousands of articles and books on each area of study which you cannot read all; you have to narrow your topic which limits the numbers of the sources you need to read. The processes of writing literature review may include the following:

First, you have to determine how many references roughly you would include and the types of such sources (books, journal articles, and websites). In general, the reasonable number of references are 5 to 20 titles for undergraduate dissertation, 40 and 50 and above for master and doctoral theses respectively (Academic Skills Program, 2010). You should use search engines or library resources for this purpose. However, it is advisable to start research from an academic library because any journal there can be considered credible, while search engine through internet cannot judge the quality and credibility of information unless you are using data bases like **ERIC** or other digital library educational literatures or resources.

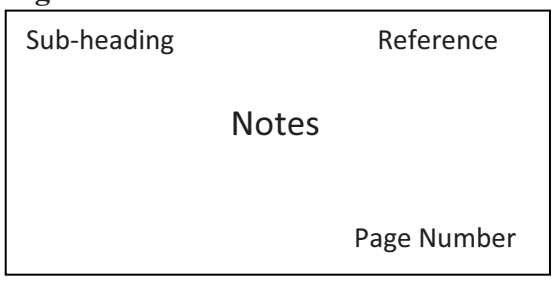
Second, you have to sort and prioritize such studies according to their relevance and significance to the topic of the dissertation.

Third, you start reading and taking notes using the note cards. Each card should contain a single idea and on its top has a reference and the title of the section where you think it will fit. It is also necessary to make sure you add the full details of each reference to the bibliography.

The note card should contain the following information:

1. author/s
2. sub-heading
3. page numbers
4. your note: a quotation, paraphrase, or summary
5. any comments

Figure 1.1a Note Card



Fourth, when you finish reading you should sort these cards again based on the specific sections of the literature review.

Finally, the content of your literature should be written; the sentences should be grouped in a way that expresses and develops one aspect of the topic. Use a new paragraph for another aspect or another topic. This is because literature review should be organized around ideas, not the sources themselves as an annotated bibliography.

Writing literature review should also encompass the following (Hart, 1998):

- Expressing one idea in a sentence.
- Using a consistent grammar and correct punctuation.
- Using transition words that link paragraphs and which show contrast and development to your argument, e.g. hence, therefore, but, as a result, in contrast, so we link between concepts and processes.
- Use past tense

Paraphrase, Summary and Quotation

Literature review includes borrowing information from a source *without plagiarizing*. There are three ways you can avoid plagiarism: paraphrasing, summarizing and quoting.

Paraphrasing is the process of presenting the author's information or opinions in your own words. It is used for two main purposes. Firstly, it is used as a study technique since it tests the person's understanding of writer's ideas, so it helps the researcher to learn better. Secondly, it is used as a writing technique where it helps a writer to summarize another writer's ideas.

Steps to Effective Paraphrasing:

- Read the original passage several times until you understand it well.
- Set the original aside, and write your paraphrase on the note card.
- Write a few words at the bottom of the card note about how you will use this material and at the top of the card, write the title of the paraphrase.
- Compare the paraphrase with the original to make sure you present the information accurately and in a new form.
- Write the reference including the page number on the note card.

Summarizing includes putting only the main ideas of other writer into your own words. It should be shorter than a paraphrase and the source because it "moves much farther than paraphrase away from point-by-point translation" (Plotnick, 2002).

Steps to Effective Summarizing:

- Read the article several times until you understand the meaning very well.
- Put circle around the key terms and underline essential ideas and the main points of the article.
- Divide the article into parts according to thoughts or main ideas.
- Write a brief outline of each part.
- Use your own expression; write the main ideas of the article as the first draft of the summary.
- Review your summary and combine the important information you gathered into paragraphs.
- Make sure you expressed the same meaning of the original material accurately and delete any repetitions.
- Compare the summary with the source (Plotnick, 2002).

Quotation on the other hand is copying the other statements exactly as they are. So it must be identical to the original source and it should be put between quotation marks. Quotations can add to the body of a paper, assisting with flow, clarity, and the support of the main idea. It is allowed to use about 10 percent of the paper as a quotation.

For short quotation, less than three lines or fewer than 40 words make it as part of the main text and it must be enclosed in quotation marks.

For long quotation, more than three lines or more than 40 words it must be set apart from the main body of your essay and indented from the left and right margin.

The original passage:
Several Studies show that job autonomy is significantly correlated to pay satisfaction and organizational commitment. Heart and Spell (2009) argue that, when employees are given more control over their own jobs, the employees will feel as if the company treats them more fairly (Froese, F and S. Xiao, 2012).

Quotation:

"Several Studies show that job autonomy is significantly correlated to pay satisfaction and organizational commitment" (Froese, F and S. Xiao, 2012, p. 2150).

Figure 1.1b: An example of short quotation

It is important to mention that in all the above three ways we should attribute the work to the original author. Otherwise, it will be considered a plagiarism. The only difference in the case of quotation is that it is important to indicate from which page in the source document the quote is taken.

When should we paraphrase, and when should we summarize?

Writers usually need to use paraphrase and summary because they are the only important tools which allow him to borrow other people's ideas without using too much quotation. So using these writing techniques helps them to control their dissertation. But the question is that when should we paraphrase and when should we summarize? The answer is that it depends on "how much of the detail from your source is relevant to your argument" (Plotnick, 2002); if your reader needs to know the main ideas of the article then summarize. In fact, a summary presents the most important elements of the source. Therefore, you need to absorb the meaning of the passage very well before you start writing. Summary is usually shorter than a paraphrase. However, it is advisable that we should not rely on one of these techniques but we use the one which serves our writing purpose better.

The original passage:

Students frequently overuse direct quotation in taking notes, and as a result they overuse quotations in the final [research] paper. Probably only about 10% of your final manuscript should appear as directly quoted matter. Therefore, you should strive to limit the amount of exact transcribing of source materials while taking notes. Lester, James D. Writing Research Papers. 2nd ed. (1976): 46-47.

A legitimate paraphrase:

In research papers students often quote excessively, failing to keep quoted material down to a desirable level. Since the problem usually originates during note taking, it is essential to minimize the material recorded verbatim (Lester, 1976, 46-47).

An acceptable summary:

Students should take just a few notes in direct quotation from sources to help minimize the amount of quoted material in a research paper (Lester, 1976, 46-47).

A plagiarized version:

Students often use too many direct quotations when they take notes, resulting in too many of them in the final research paper. In fact, probably only about 10% of the final copy should consist of directly quoted material. So it is important to limit the amount of source material copied while taking notes.

Figure 1.1c Key Pointers to Paraphrase and Summary

An illustration on how to paraphrase, summarize and quote is given below:

Original Passage (85 words)

Language is the main means of communication between peoples. But so many different languages have developed that language has often been a barrier rather than an aid to understanding among peoples. For many years, people have dreamed of setting up an international universal language which all people could speak and understand. The arguments in favor of a universal language are simple and obvious. If all peoples spoke the same tongue, cultural and economic ties might be much closer, and good will might increase between countries (Kispert)

Paraphrase (63 words, about the same length as the original)

Humans communicate through language. Because there are so many languages, however, people around the world have a difficult time understanding one another. Some people have wished for a universal international language that speakers all over the world could understand. Their reasons are straightforward and clear. A universal language would build cultural and economic bonds. It would also create better feelings among countries (Kispert).

Summary (28 words, much shorter)

People communicate through language; however, having different languages creates communication barriers. A universal Language could bring countries together culturally and economically as well as increase good feelings among them (Kispert).

Short Quotation: (Less than than3 lines or fewer than 40 words, in Inverted Commas)

"For many years, people have dreamed of setting up an international universal language which all people could speak and understand."

Long Quotation: (More than or equal to 3 lines or more than 40 words, Indented)

Language is the primary means by which people throughout the world communicate. But the existence of many languages throughout the world is a barrier to communication. A solution is to have a common language as the author points out its benefits:

> For many years, people have dreamed of setting up an international universal language which all people could speak and understand. The arguments in favor of a universal language are simple and obvious. If all peoples spoke the same tongue, cultural and economic ties might be much closer, and good will might increase between countries.

Communication can be direct between parties concerned without the need for third party. As a result some of the current global problems can be surmounted.

Figure 1.2 Illustrations on Paraphrase, Summary and Quotation

Source: Oshima Alice and Ann Hogu (2006), Writing Academic English, Pearson Longman, USA, 4th ed.

Take stock of all you have read

A simple tabulation of what you have read, written, and summarized in Note Cards etc. can help and guide you through the flow of events and later in writing the literature review. A simple format you smay follow is given below.

No.	Source	Author/s	Year	Title	Method/s	Result/s	Comments/Gaps
1.							
2.							
3.							
4.							
5.							

Figure 1.3 Summary of Literature Review

As you read through the relevant articles, keep a record of the sources, author/s, date when the article was published, the title, methods used in the analyses, and the outcome of those analyses. Further, you can critically look at the articles by including your opinions and any gaps in those analyses in the comments/gaps analyses column.

Conclusion

Literature review includes gathered materials or the critical points of current knowledge on a topic. This will help us to support our arguments and our findings. In other words, it shows that what we are saying is valid. It could be information from the company pamphlets or journals or any other sources.

To present this information in their own words, the writers usually use three levels of borrowing from other's work: paraphrase, summary and quotation. In all cases, the references should be stated.

It is advisable to avoid annotated bibliography which means listing sources and going into detail about each one of them, one at a time. The required result of the literature review is to establish a solid background for the current research.

1.5 Referencing

How to do Referencing

There are msany ways are possible, but it is institute regulations or journal citation procedure select an approach and *stick to it throughout*. Here, we have given a simple and effective approach to commonly used referencing.

Two types of references need to be mentioned:

1. In-text reference (**in the essay itself**)
2. End-text reference (**bibliography**)

Examples:

1. **Name:** Dr. Abdulmonem Alzalabani

   In text: … (Alzalabani, 2006) or Alzalabani (2006)…

   - ✓ You need to use **family name** here.
   - ✓ In the absence of family name, write the name itself here.

   End text: Alzalabani, A., (2006). "Training and Development in Saudi Arabia", *International Journal of Training and Development*, Vol. 6, No.2, pp.125-140.

   - ✓ If you don't want ***italics***, you can **underline** the book title or ***enclose*** the book title with ***double quotation*** marks.
   - ✓ It is an International norm to use italics for title of books and journals.

2. **Name:** Juan Dela Cruz

   In Text: … (Dela Cruz, 2007) or Dela Cruz (2007)…

   - ✓ You need to use **family name** here.

End text: Dela Cruz, J., (2007). Introduction to Management, 4th Ed, Sun Publication.

> ✓ if you don't want to **underline the book title**, you can ***italicize it***.

3. **Name:** Dr. N.N. Rajiva Menon

   In text: … (Menon, 2007) or Menon (2007)…

> ✓ You need to use **family name** here.

   End text: Menon, N. N. R., (2007). *Introduction to Management,* 4th ed, Sun Publication.

> ✓ If you don't want to **underline the book title**, you can ***italicize it***.

4. **Name:** Ng Ah Kow

   In text: (Ng, 1991)

   End text: Ng, A.K, (1991). "Financial Patterns in Industries", *Journal of Finance*, 28, 3:389-396)

> ✓ The citation refers to volume 28 number 3 pages 389-396.
> ✓ The second line must be indented.

5. **Name:** Rob Brown

   In text: (Brown, 2002, page 11, Table 1)

> ✓ This reference is to specific page or diagram.

   ibid, page 11

> ✓ This refers to the above author and page 11 now.

(Brown, 2002[a], 2002[b], 2002[c], 2002[d])

- ✓ This style is used if the author has published many works in the same year.

op cit

- ✓ This means the source is somewhere, mentioned earlier in your essay.

End text: Brown, 2002, ---

If two authors (include both authors)

Name: Praveen Balakrishnsan Nair and Ng Ah Kow

In text: (Praveen & Ng, 1998, p.96)

End text: Praveen, B. N., Ng, A.K, (1998), "Financial Patterns in Industries", *Journal of Finance*, 28, 3:389-396

For more than two but less than six authors

First-time citation includes all authors, second-time citation includes the first author and use *et al* plus the date and third-time citation uses the first-named author and *et al* but does not include the date anymore. All this is in the same paragraph.

In text: (Uma, Nair & Segaran, 2003)

- ✓ Use the format the first time you cite the authors.
- ✓ If there is **no family name**, use the name as it is.

Uma, et al. (2003)

- ✓ Second-time citation.

Uma, et al.

- ✓ Third and subsequent mention.

End text: *List all the names in the manner given as in the example for two authors.*

Author Referred to by another Author

In text: Nair (cited in Segaran, 2004, p.54) says that…**or**

Nair (2000) is quoted by Segaran (2004, p.54) as saying that…

End text: Cite Segaran

Segaran, X.(2004), "_____"

For Newspaper Articles

In text: L*ist name, year and page as in earlier examples*

End text: Bala, K. 1996. Art of Management. New Straits Times Press (11 November 1996):15

For Internet References

Example:

In text: (Bala, 1999)

End text: Bala, K. (1999). *Image of Man; Report on Performance* (on-line). Available WWW: http://www.abc.abc (11 November 1996)

✓ Full Web link and the date accessed must be given.

Accounting References

Examples

In text: [Accounting Standard Board (ASB), 1997, para. 15]

✓ First-time citation.

✓ Second-time citation.

(ASB, 1997, para. 15)

End text: Accounting Standard Board (ASB). (1997). Accounting *Standard Board ASB 1997: Earnings*. Accounting Standards Board.

Special Note

Date is uncertain: Can be written as (Johns, c. 1910, p.54) or (John, n.d.)

✓ The **c** stands for *around that time* and **n.d**. stands for *no date*

No author: *The title or the shortened version of the title can replace the author's name.*

Note: Refer to Appendix III, page 164 for other citation styles

1.6 Research Data Collection

There are two broad categories of data collection methods: qualitative and quantitative.

Qualitative Research

Qualitative methods are used for interpretation of situations. Data are narrative in nature and collected through interviews, on-site observations, and focused groups. The data collected are then analyzed and explored for themes and patterns.

This approach is good for method of inquiry employed in market research and other related fields where an in-depth understanding of behavior and the reasons for such behavior are needed. The issue is why and how a decision is made, hence, smaller and focused samples are more appropriate.

Data can be analyzed through coding (a word or phrase that links to research objectives) to demarcate the sections. This can help organize the data and provides a mean to introduce quantitative methods for interpretation.

Some data can be analyzed without coding. The data sets are summarized and the summaries are then further refined until a compact summary is obtained.

Quantitative Research

Quantitative inquiries are based on numerical and the use of statistical processes to answer specific questions. Data are collected through questionnaires, surveys and data bases. Statistics is used widely in a variety of ways to support an inquiry or program assessment and evaluation. Note: refer to Part II, page 49 for more detailed analysis of data types.

Phenomena and observations are developed into models, theories, hypotheses, and quantitative analysis can provide a testing platform for your empirical observation. With correct sampling technique and sample size, it is hoped that observations can be generalized to some larger population.

Tools Used to Collect Research Data

Method: Questionnaires, Surveys, and Checklists

This method is good when you need information quickly and/or easily. It induces lots of information from people in a non-threatening manner.

This method is anonymous, inexpensive to administer, easy to compare and analyze. It can be administered to as many people as possible and there are many sample

questionnaires available. Search for scholarly articles for sample questionnaires for survey research over the Internet.

However, this approach might not get careful feedback. The wording can be biased to client's responses and the tone personal, and this method may require surveys sampling experts. In addition, a researcher may not get the full story through this procedure.

Note: A more detailed analysis of questionnaire will be introduced in Section 1.7, page 30.

Method: Interviews/Focus Groups

Interview is appropriate when you want to fully understand someone's impressions or experiences, or learn more about their answers to questionnaires. This method gets full range and depth of information, develops relationship with client, and can be flexible.

However, interview can take much time, can be hard to analyze and compare, can be costly, and the interviewer can be partial to client's responses.

Focus group is appropriate when you explore a topic in depth through group discussion, e.g., about reactions to an experience or suggestion, understanding common complaints. This is useful in evaluation and marketing. The method allows you to quickly and reliably get common impressions. It is an efficient way to get much range and depth of information in a short period of time besides conveying key information about programs.

However, focus group can be hard to analyze responses and this method needs a good facilitator for safety and closure. It is also difficult to schedule 6-8 people together.

Method: Documentation Review

This method is appropriate when you want an impression of how a program operates without interrupting the program. It is appropriate for documents such as applications, finances, memos, minutes, etc.

This method is comprehensive and ideal in getting historical information without interrupting a program or client's routine in a program. The information already exists and there are few biases about the information.

However, this approach often takes much time, and information may be incomplete. There is a need to be quite clear about what you are looking for because information is not flexible and data is restricted to what already exists.

Method: Observation

This method is appropriate when you want to gather accurate information about how a program actually operates, particularly about processes.

This method views operations of a program as they are actually occurring and can adapt to events as they occur.

However, it can be difficult to interpret observable behavior and it can be complex to categorize observations. Moreover, the behavior of participants can be influenced by external factors, and this method is somewhat costly.

Method: Case Studies

This method is appropriate when you want to fully understand or depict client's experiences in a program, and conduct comprehensive examination through cross comparison of cases.

This method is a powerful means to portray program to outsiders.

1.7 The Research Questionnaire

Importance of Good Questionnaire

Questionnaire is one among the important data collection tools used for survey and research purposes for many research students. But it takes time to collect, organize and describe information collected. . The choice of questions, words, and even the order of putting questions can be very influential in affecting how respondents (people who answer the questions) react to a survey.

A questionnaire should translate research objectives into specific questions. It should create interest in the mind of the respondents and should keep them motivated throughout the setting. It should be simple, easy to administer and brief. Questionnaire development is an iterative process. It should go through a series of drafts before it reaches its final format. When you write each question, you should think about the probable answers also. This will help you to minimize the 'question bias', that is, the capability of question format to influence respondents' answers.

Steps in the Development of a Questionnaire

1. First define your survey/research objectives.
2. Develop suitable questions with the survey objectives in your mind.
3. Check whether the questions are conveying what information you really intend to get.
4. Check the language and grammar.
5. Arrange the questions in a logical order.
6. Evaluate the questions.
7. Pretest the questionnaire.
8. Make revisions as needed.
9. Finalize.

DOs and DONTs in Questionnaire Development

- Be brief and clear in number of items and in using words for each statement. Make the questionnaire as simple as possible by using words that everyone know its meaning. Avoid ambiguous words like 'often', 'occasionally', 'frequently' and other similar words whenever possible. For example, 'frequent reading' outside teaching hours may be one or two hours a week for one person and twice as much for another.
- Avoid "double-barreled questions". Double-barreled questions are those which carry two different issues at once. An example of a double-barreled question would be the following: "Do you think that a business undergraduate should have more classes about management and liberal studies?" Combining both questions into one may confuse the respondents. Some respondents would answer "yes" to both and some "no" to both, some would like to answer both "yes" and "no".
- Questions should not lead the respondent to a particular answer. Such questions present only one aspect of an issue. A question should be constructed in as neutral a way as possible and avoid specific name of a brand or a company or by presenting all the sides of an issue.
- Arrange the questions in a logical sequence. Use simple and interesting opening questions to gain the respondent's co-operation. Use the funnel approach by beginning with a very general question on a topic and gradually leading up to a narrowly focused question on the same topic. Arrange questions in logical order as sudden changes in subject confuse the respondent and cause indecision. Place difficult or sensitive questions near the end. Sensitive questions should be relegated towards the end of the questionnaire, once the respondent has become involved in the study.
- If possible, include 'cross-check' questions. Questions can be posed from both positive and negative side of the topic so that at the end of the session, it can be cross checked for consistency or if there is any contradiction. In case if contradiction is noticed then the responses cannot be taken as correct. Such methodology will make the research work free from errors and foolproof.
- Try to minimize the number of open-ended questions. Open-ended questions require the respondents to provide their own answers to the questions and too many such questions may pose problems enumerating them.
- Avoid personal questions if possible. Most people will not answer them. Avoid questions like: "How old are you?" But rather such question can be framed as: "Which age categories do you belong to?"
- Always add a cover letter if it is a mailed questionnaire. Refer to example page 32.

- Attach pre paid envelopes if the questionnaire is to be returned by mail.
- Categorize the questions into sections based on your Objectives. This process will help in data analysis and recording and ease the respondents' task and minimize the jumping ideas.
- Do a pilot test. A small sample of questionnaire is tested for validity, reliability, and clarity before distribution to minimize problems later. Tabulating the results of the pre-test is also very useful to ensure that all the required information will be obtained.

1.7.1 The Cover Letter

A cover letter is a letter from the researcher to the respondent, usually attached with the questionnaire, introducing the research and its purpose. It may also guide the respondent on how to fill in the questionnaire.

A cover letter should be brief and clear and should be drafted in such a way to motivate the reader to fill in the questionnaire.

A sample cover letter is given below.

A Sample Cover Letter

School of Business,

YIC College,

Yanbu Al Sinaiyah

Dear Respondent,

This is a survey undertaken to analyze the buying behavior for Sweetlife confectionary products as part of the requirements for the Bachelor of Industrial Management degree. Participation in this study is purely voluntary. I would like to solicit your kind cooperation and take a couple of minutes of your precious time to fill up this questionnaire. All your responses will be treated with strictest confidence, and will be solely used for the purpose of this research. Your involvement is highly appreciated as valuable contribution to the success of this study. Please complete the questionnaire based on your honest opinion.

Should you have any queries please feel free to contact me.

I thank you in advance for your cooperation.

Sincerely,

Saud Abdul Aziz

Phone no. 011 12345

E mail: Saud-Abdul@yiccollege.com

1.7.2 Sample Questionnaire

You will notice in the given sample below a variety of question types being asked in a single questionnaire, viz. ***closed ended***, ***open ended***, ***scale rated*** etc. It is based in experience to decide on the type of questions to be used to suit your need. Keep in mind that the main objective is to elicit the information from the respondent in the best possible way. Try your level best to make a maximum number of questions closed ended as it will make the questionnaire unambiguous and facilitate tabulation.

SAMPLE QUESTIONNAIRE

GENERAL INFORMATION:

1. Gender: [ ] Male [ ] Female
2. Age: [ ] Less than 22 years
   [ ] 22 to 25 years
   [ ] More than 25 years
3. Residency: [ ] Saudi
   [ ] Non Saudi

(**Note:** *It is a common practice to* ***avoid asking personal details*** *if the questionnaire is intended to collect personal or confidential data.*

You may add more questions in this section as you deem fit, like marital status, number of children, income class etc.)

SURVEY QUESTIONS:

1. Of the varieties given below which one do you prefer more?

   Please tick (√) the relevant box once only.

Plain Laban	
Low fat Laban	
Strawberry Laban	
Mixed fruit Laban	
Others (Specify) ____	

2. Who buys Laban for you?

Myself	
Others (Specify) _____	

3. Usually how many Laban are bought at a time?

One or Two	
Three to Five	
Six to Ten	
Above Ten	

4. How often do you purchase the Laban? _____ times a week

5. How much money are you willing to spend for one Laban?

Less than 50 hallala	
50 hallala	
75 hallala	
1 SR	
Above 1 SR.	

4. Rank the attributes considered by you before buying Laban

Value for money	
Taste	
Availability	
Brand name	
Packing	

5. 7. Are you pre decided on purchasing some Laban when you go for shopping?

YES / NO

If no, what motivates you to buy?

Advertisement	
Attraction of Laban/Display	
Point of purchase aids	
Others (Specify) _____	

8. Do you ask for any specific brand of Laban? YES / NO

9. If your brand is not available,

Settle for another brand	
Go to another outlet	
Postpone the purchases	

10. Factors influencing the decision to settle for another brand

Display of Laban	
Shopkeepers Choice	
Point of purchase advertisements	
Others (Specify) _____	

11. If you have no brand preference, your buying decision will be based on:

Laban that catches attention	
Point of sale display	
Preference of shopkeeper	
Any offers attached to Laban	
Others (specify) _____	

12. Which brands come to your mind when you think of Laban? List ten preferences without necessarily ranking them.

1		6	
2		7	
3		8	
4		9	
5		10	

13. Which advertisements from magazines or TV can you recall when you think of Laban?

1	
2	
3	
4	
5	

14. What all brands of *Healthy Life* Laban are you aware of?

1		5	
2		6	
3		7	
4		8	

15. Tick the columns for each brand if they apply to your experience.

***HEALTHY LIFE* BRANDS**	**Aware**	**Tried**
Health Delight		
Natural Delight		
Family Delight		
Fruity Delight		

16. Tick your agreement or disagreement level with the specified attributes given below in relation to *Health Life* Laban:

ATTRIBUTES	**Strongly Agree**	**Somewht Agree**	**Neutral**	**Somewhat Disagree**	**Strongly Disagree**
Good taste					
Value for money					
Refreshing					
Homemade taste					
Easily available					
Suitable with all types of meal					
Suitable during meal					
Attractive brand name					
Attractive container					
Meant for children					
Meant for family					

Thank you very much for your time and consideration in completing this questionnaire.

Note: Question 16 can also be written in this format:

Please **circle** the number which represents the extent to which you agree or disagree with the specified attributes given.	1. strongly agree 2. somewhat agree 3. neutral 4. somewhat disagree 5. strongly disagree

a.	good taste	1	2	3	4	5
b.	value for money	1	2	3	4	5
c.	refreshing	1	2	3	4	5
d.	thirst quenching	1	2	3	4	5
e.	easily available	1	2	3	4	5
f.	suitable for leisure time	1	2	3	4	5
g.	suitable after meal	1	2	3	4	5
h.	attractive brand name	1	2	3	4	5
i.	attractive wrapper	1	2	3	4	5
j.	meant for children	1	2	3	4	5
k.	meant for adult	1	2	3	4	5

1.7.3 Uniformity of Questions

Questions must be either positively or negatively written, you should not have a combination of both otherwise your *summary data will be affected* and your inference misleading. The usual approach is to have questions positively placed with the exception of having one or two questions to check the honesty of respondent.

Illustration: On the Survey Question on Students Satisfaction, one of the researches' Objectives can be Lecturer's Performance.

The following questions can be devised for Lecturer's Performance.

1. Lecturers have always take concern about my academic advancement.
2. I found my lecturers helpful if I encountered difficulties in my subject area.
3. Lecturers marked assigned work fairly and returned it in a reasonable time.
4. Lecturers **Do Not** always give encouragement or motivation towards my study.
5. My lecturers have **No** good knowledge and skills in teaching their subjects well.

Questions 1 through 3 are positively placed, but Questions 4 and 5 is negatively placed. You should remove the word **'Do Not'** from Question 4 and **'No'** from Question 5 to make it positively placed.

1.7.4 Common Approaches to Questions

You can use the following guide to device Questions.

For Questions that are Behavioural in nature, you can start with;

Have/Do you ever …?

Who/which do you know …?

When did you …?

In what ways do you …?

For Questions that are Attitudinal in nature, you can start with;

What do you think …?

Why do you …?

Do you agree …?

How do you rate …?

Which is the best …?

As you design your questions, Always keep track of the following:

- Research Question/s and Objectives in mind
- How to Kindle the Knowledge and Interest of respondents
- Order of Questions, Types of Questions and Possible answers
- Ease of Processing Data through the use of Software.
- Do not pose Questions that are biased and ambiguous but use short and simple ones.
- Use response Range for sensitive questions like Age and Salary.
- For fixed response question, allow an option 'Others' as part of the answer provided.

1.7.5 Questionnaire Distribution Size

Questionnaire is necessary in survey based research and the number of questionnaires you should distribute depends on the accuracy you require.

Basically, it depends on:

a. the purpose of your study.

b. the size of population in your study.

c. the sample → the smaller the error, the larger the sample should be.

d. the sample size → should not be smaller than 30, otherwise it will not bring out the pattern in your study.

(*There are techniques to handle smaller sample sizes; however, you should refer them to a statistician as it is beyond the book scope*)

Usually a sufficient size beyond 30 where patterns can be discerned should be a fair number; nevertheless, for those who want the size to reflect sampling techniques where *confidence level, level of variability, and the level of precision* matter, they can use the following techniques to come out with a suitable sample size.

For this you need to know some terms, they are:-

Level of precision (sampling error)

It is the range in which the true value of the population is estimated. Thus, if a teacher finds that 70% of the students in a sample from a school have adopted IT in their studies with a precision rate of ±5%, then he or she can conclude that between 65% and 75% of the students in the population have adopted IT in their studies.

Confidence Level

95% confidence level means 95 out of 100 samples will have the true population value within the range of interval specified above.

Degree of Variability

This measures the distribution of attributes in a population. The more heterogeneous a population, the larger the sample size, and for a more homogeneous population, the smaller the sample size must be for a given level of precision. A proportion of 50% indicates a greater level of variability than say 20%. Twenty percent indicates that a large majority do not have the attribute of interest. An 80% indicates that a large majority do have the attribute of interest. As a proportion of 0.5 (50%) indicates the

maximum variability in a population, it is often used in determining a more conservative sample size.

Once you have understood the above and decided on a set of values for the above measures, you can either use a:-

A. Ready Published Table

- This table is for a given combination of precision, confidence levels, and variability only.

B. Calculation through Use of Formulae

- For any other combination of precision, confidence, and variability, that is appropriate.

***Sample Size Calculation* (for 95% confidence level)**

You can use the formula below by Yamane (1967:886) who provides a simplified formula to calculate sample sizes. A 95% confidence level and P = 0.5 are assumed for the equation. *This is sufficient and suitable for most of our needs*.

$$n = \frac{N}{1 + Ne^2} \qquad \text{----- (1)}$$

Where N – population size
n – sample size
e – error

Example1:

If the population Size is 500 and the margin of error is 5%.

Your sample size is, $n = 500/\ \{(1 + 500(0.05)^2\}$

$= 222$ (rounded)

That is, a sample size of 222 is selected, at a 95% certainty and a 5% margin of error for a population of 500.

Example2:

If the population Size is 500 and the Margin of Error is 1%.

Your sample size is, $n = 500/\ \{(1 + 500(0.01)^2\}$

$= 476$ (rounded)

That is, a sample size of 476 is selected, at a 95% certainty and a 1% margin of error for a population of 500.

For other margin of errors, you can use 0.03 (3%), 0.02 (2%) and 0.01 (1%). However, an estimation of the population's characteristics to within plus or minus 3-5 per cent of its true value is a norm for most business and management research.

Sample Size Calculation for Proportion (for 99% confidence level and other levels)

You can use the following formula by Cochran (1963:75) for other confidence intervals *if the need arises*.

$$n_0 = \frac{Z^2 pq}{e^2} \qquad \text{----- (2)}$$

Where:

Z – the z-score based on Normal table, for 95% Confidence Interval, Z = 1.96 and for 99%, it is 2.58. For other values, refer to Normal Table.

P – the degree of variability (where q = 1- p)

e – precision level

Example

If Margin of Error required is 5% and the variability is taken as 0.5, with a confidence interval of 99% that is Z = 2.58.

Your sample size is, $n_0 = \frac{2.58^2 \times 0.5 \times 0.5}{0.05^2} = 666$ (for 99% confidence interval

$n_0 = \frac{1.96^2 \times 0.5 \times 0.5}{0.05^2} = 384$ (for 95% confidence interval)

And if the Population is Finite (N), an adjustment is made,

Adjusting Formula, $$n = \frac{n_0}{1+(\frac{n_0 - 1}{N})}$$

----- (3)

For the *above example*, say the finite population is known and is 2000,

The adjusted sample size is for 99% confidence interval;

$$n = \frac{666}{1+(\frac{666-1}{2000})}$$

$$= 500$$

The adjusted sample size is for 95% confidence interval;

$$n = \frac{384}{1+(\frac{384-1}{2000})}$$

$$= 322$$

Formula (2) can be adjusted if you know the *variance* (σ^2) of the attributes in the population.

$$n_0 = \frac{Z^2\sigma^2}{e^2} \qquad \text{----- (4)}$$

Example

If Margin of Error required is 5% and σ is known and say it is 0.7 with a confidence interval of 99% that is Z = 2.58.

Your sample size is, $n_0 = \frac{2.58^2 \text{ x } 0.7 \text{ x } 0.7}{0.05^2} = 1305$ (rounded)

1.7.6 Questionnaire Distribution and Sampling Techniques

The choice of sampling technique depends on research question/s and objectives and the need to make inferences. Sampling can be either probability based or non-probability based. For the probability based sampling the technique selection can be based on how the attributes of interest to you are distributed across the population. If these attributes are uniformly distributed having equal chance of selection across the population then the techniques under uniform population can be applied, otherwise use non-uniform population.

The following will help you in identifying the distribution and sampling approach most appropriate for your study.

Random Sampling: If the Population of Your Study is Uniform

i. Use Simple Random Sampling

Give a number as reference (like 1, 2, 3 etc.) for each of the members in the population, and select the required sample size based on random selection. Random numbers form random number table, calculator (for Casio scientific calculator use Ran#) or computer software (e.g. Excel has Random Number Generation under Data Analysis) can be used to select the required sample.

Example: If the Population is 25 and a Sample Size of 5 is required.

First give a reference number 1 to 25 to each member in the population. Five random numbers are generated (Casio scientific calculator through the use of Ran#) 1, 7, 12, 5, 23.

So you select members from population with reference number 1, 7, 12, 5 and 23 as your sample of size five.

This method is purest form of probability sampling and is ideal for small sample selections. The process itself is simple and easy to understand and the calculation of means and variance are easy.

However, this is not the most efficient method and not good for very large population.

It requires knowledge of complete sampling frame and you cannot always be certain that there is an equal chance of selection.

ii. Use of Systematic Sampling

This sampling process spreads the sample more evenly over the population and is easier to conduct than a simple random sample; however, the system may interact with some hidden pattern in the population that you need to be careful about.

Here again, give a number as reference (like 1, 2, 3 etc.) for each of the members in the population and select each member on a number multiple. If your number multiple is five, select 1, 6, 11, 16, etc. To know the number multiple, you need to divide the population size by sample size.

Example: If the Population is 500 and Sample Size is 25.

Your multiple is 20 (i.e. 500/25)

First, select a random number from 1 to 20,

Say your random number is 1.

Then, you select 1, 21, 41, 61 etc. until you get a sample of size 25.

Alternatively, you can use sampling fraction (sampling fraction = actual sample size/total population).

Example: If the Population is 500 and Sample Size is 25.

Your sampling fraction is 1/20 (i.e. 25/500)

Again, select a random number from 1 to 20,

Say your random number is 3.

Then, you select 3, 23, 43, 63 etc. until you get a sample of size 25

(Note: If you encounter any problems with sampling fraction, round the population to the nearest tens or hundreds and increase the sample size until a manageable sampling fraction is obtained.)

In this method sampling frame does not need to be defined in advance and it is easier to implement in the field. If there are unrecognized trends in the sample frame, systematic sample ensures coverage of the spectrum of units. It is as good as random sampling and the technique is simple.

Random Sampling: If the Population of Your Study is Non-Uniform

For non uniform population, divide the population into number of uniform groups (strata) like Group 1, Group 2 etc.

The strata may be based on a single criterion or a combination of two or more criteria. Single criterion can be survey participants' gender that can be divided into two strata like male and female or it can be their residence urban and rural. A two criteria can be age and gender of participants.

And from each group, you can select the required sample size using simple random sampling or systematic sampling.

To give emphasis to the size of each group, (like group 1 can have fewer members than group 2) you can use proportion in your selection to reflect more equitable distribution. For this, the sample size selected in each group is based on the proportion of that group size relative to population size.

Example

If your population is 1000 and Group 1 has 200 members, your proportion is 0.2 (i.e. 200/1000). Therefore, for Group 1, you select 0.2 proportion of the overall sample size for Group 1. If overall sample size required is 500, Group 1 contributes 100 members (i.e. 0.2x500). The figure illustrates when you have three uniform (G_1, G_2 and G_3) attributes groups.

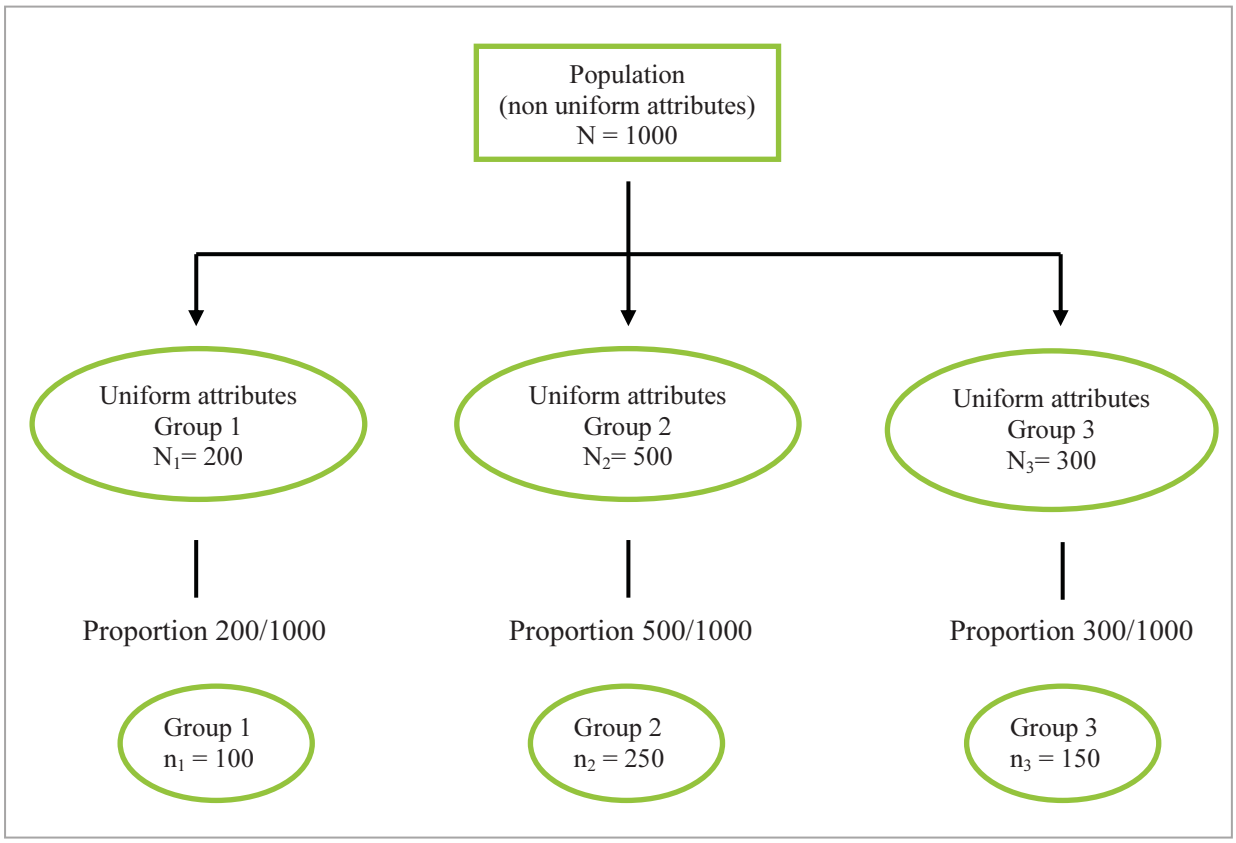


Figure 1.4 Strata Sampling

Other Probability Sampling Methods

Cluster Sampling

Selection of groups of study units (clusters) instead of selection of study units individually.

Population is divided into clusters and a subset of the cluster is randomly selected.

Example:

A study of purchasing behavior of rural community of a region, list of villages is made and random sample of villages is chosen and all study units in selected villages are interviewed. This approach needs a large sample size.

Multi-stage Sampling

A combination of two or more sampling methods are used as stage by stage for large survey

For example start with simple random sampling for selecting regions followed by cluster sampling of districts for each selected region, then stratified sampling as part of cluster sampling (rural and urban) and lastly a simple random sampling from within each cluster.

Example:

A study of utilization of computers in a district where 150 houses are to be visited and the district is composed of 6 sub districts and each sub district has between 6-9 villages. The steps to be taken are select 3 sub districts out of 6 by simple random sampling, select 5 villages from each sub districts by simple random sampling (say out of 15 villages) and lastly select 10 houses from each village by systematic sampling or stratified sampling.

Non- Probability Sampling Methods

Non-Probability sampling is used for surveys of hard to identify groups and for pilot surveys. It includes;

Purpose sampling- subjects are selected for special reasons, e.g. effect of new teaching technique among volunteers.

Convenience sampling – selection of respondents are conveniently done, e.g. at a bus stop, patients in waiting room of clinics.

Snow ball or chain sampling- used in studying population that is hard to reach and /or hidden

Observers enter to the specified area with a team containing only few members. Street to street or house to house sample selection is done, become large group of people at the end of sample selection.

2 *ORGANIZING DATA Using MS Excel*

By the end of this part you should be able:

- *To appreciate the different types of data;*
- *To summarize data using graphs and diagrams;*
- *To use data analysis tools in Excel;*
- *To use and interpret correlation;*
- *To use and interpret regression;*
- *To handle non-linear relationship;*
- *To use and interpret Chi-square Test results;*
- *To use and interpret ANOVA and F-test;*
- *To use post hoc test.*

Before we go into organization of data, you should have an understanding of the different categories of data types. It is important that you identifying the data types correctly so that you can use the right statistical tools to analyze the data.

Types of Data Categories

1. Nominal data (unordered categories)

The name 'nominal' comes from the Latin "nomen", meaning 'name'. This data have no measurement scale.

2. Ordinal data (ordered categories)

These data are not measured on a measurement scale. They form an ordinal (ordered, ranked) variable but the differences are meaningless. It contains more information than nominal level.

3. Interval data

There are meaningful differences between data except there is no zero or starting point. The distances between each interval on the scale are equivalent along the scale from low interval to high interval.

4. Ratio data

Same as interval data, but Ratio has a true zero as a starting point.

The following illustrations will further reinforce your understanding of data.

Illustration 1

Sample Classifications of Data Types for Part of Sample Questionnaire (General Questions 2-3, and specific questions 16, 17).

GENERAL INFORMATION:

2. Age: [ ] Less than 22 years
[ ] 22 to 25 years
[ ] More than 25 years

Interval

3. Residency: [ ] Saudi
[ ] Non Saudi

Nominal

16. Tick the columns for each brand if they apply to your experience.

HEALTH LIFE BRANDS	Aware	Tried
Health Delight		
Natural Delight		

Nominal

17. Tick your agreement or disagreement level with the specified attributes given below in relation to HEALTH LIFE Laban:

ATTRIBUTES	Strongly Agree	Somewhat Agree	Neutral	Somewhat Disagree	Strongly Disagree
Good taste					
Value for money					

Ordinal

Illustration 2: Relationship between Data Types

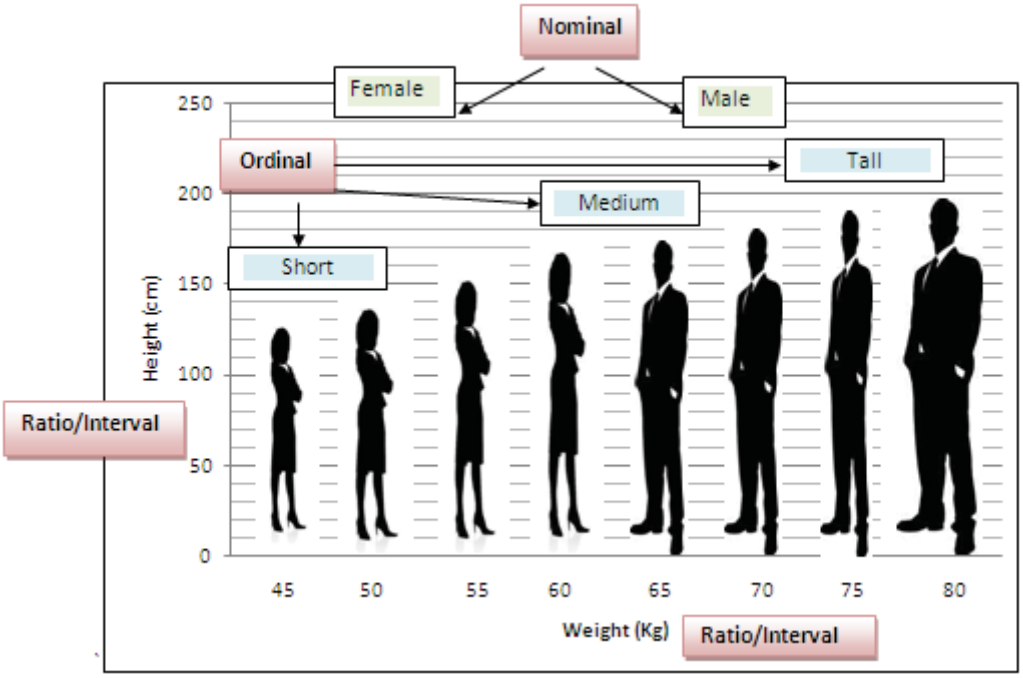


Figure 2.1a Data Types

Illustration 3: Data Type and Statistical Tool

As an illustration, the table below gives you an idea on the importance of identifying data type correctively so that an appropriate test can be applied.

Variable 1 (Data)	Variable 2 (Data)	Summary	Statistical Test
Nominal Ordinal	Nominal Ordinal	Counts Proportion	Chi Square Proportions
Nominal Ordinal	Ratio Interval	Mean	t-test z-test
Ratio Interval	Ratio Interval	Scatter Plot	Correlation

Figure 2.1b Link between Data Type and Statistical Test

Illustration 4: Data Reduction

Data can be reduced from Interval to Ordinal and Nominal, but not the other way round.

Example: Age

Interval (continuous): 20, 21, 22,…….. 80

Grouped 20-29, 30-39, 40-49, etc.

Ordinal: Twenties, Thirties, Forties, etc.

Nominal: Young, Old.

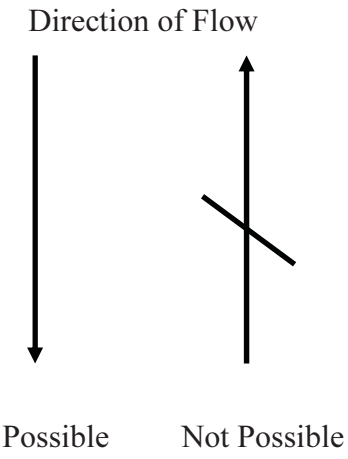


Illustration 5: Data Collection

In terms of collection, data are classified by source.

i. Primary Data: Data collected directly by the researcher

ii. Secondary Data: Data collected by other people

iii. Tertiary Data: Data from text books, encyclopedias, etc.

Illustration 6: Data Types and Analysis

i. Qualitative Data are generally Nominal and Ordinal

ii. Quantitative Data are generally Discrete (gaps between possible values) and Continuous (theoretically, no gaps between possible values)

Illustration 7: Four Common Situations

Data Type	Sample Size	Summary Stats	Stats Test
Ratio, Interval	Large	Means SD	Z-test
Ratio, Interval	Small	Means SD	t-test
Nominal, Ordinal	Large	Proportions	Test of Proportions
Nominal, Ordinal	Large, small	Counts	Chi-Square

Figure 2.1c Link between Data Type and Statistical Test

2.1 Graphs and Diagrams

Now we are ready to analyze the data. The first thing a researching student needs to do with data is to *summarize* and where possible to express those data in diagrammatic form. Once this is done, only then is he ready for advanced analysis of data using various statistical tools.

We shall first focus on drawing some common diagrams using EXCEL and then go on to summarizing data and lastly to applying various statistical tests commonly available to business students.

Many summaries, diagrams, and analysis can be readily constructed using EXCEL which is very handy software for basic analysis and some further analysis of data. Further, this software is readily available as it comes with microsoft package.

2.1.1 Getting Started

You need the **Data Analysis** tool of Microsoft Excel.

A. Analysis ToolPak for Microsoft Excel 2010 users.

You need to load the Analysis ToolPak *if it is not yet installed*. Follow the instruction below:

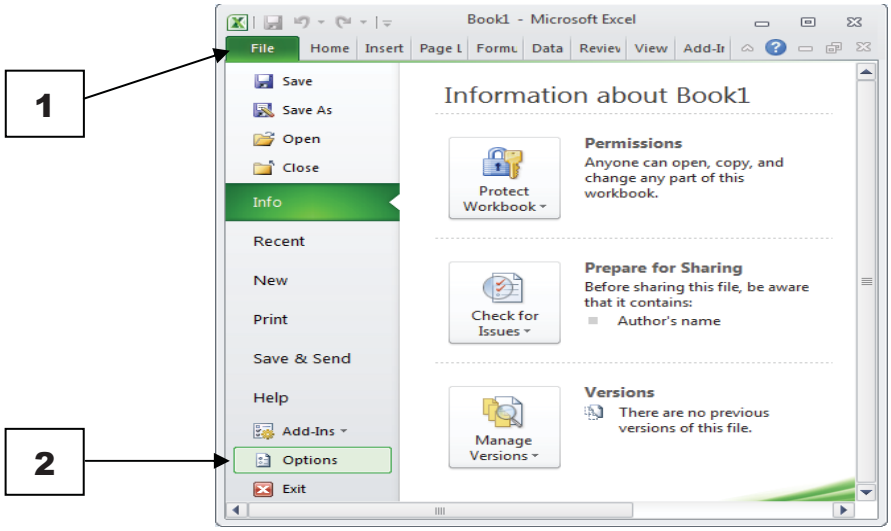


Figure 2.2 Analysis ToolPak

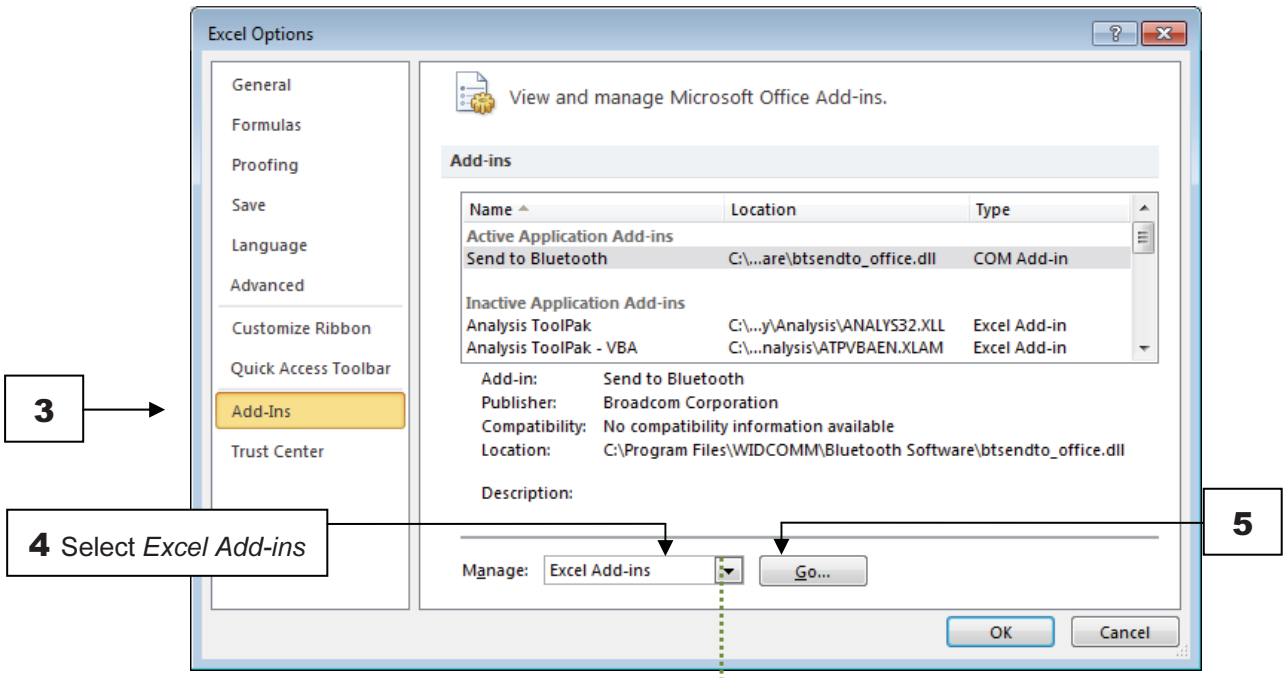


Figure 2.3 Excel Options

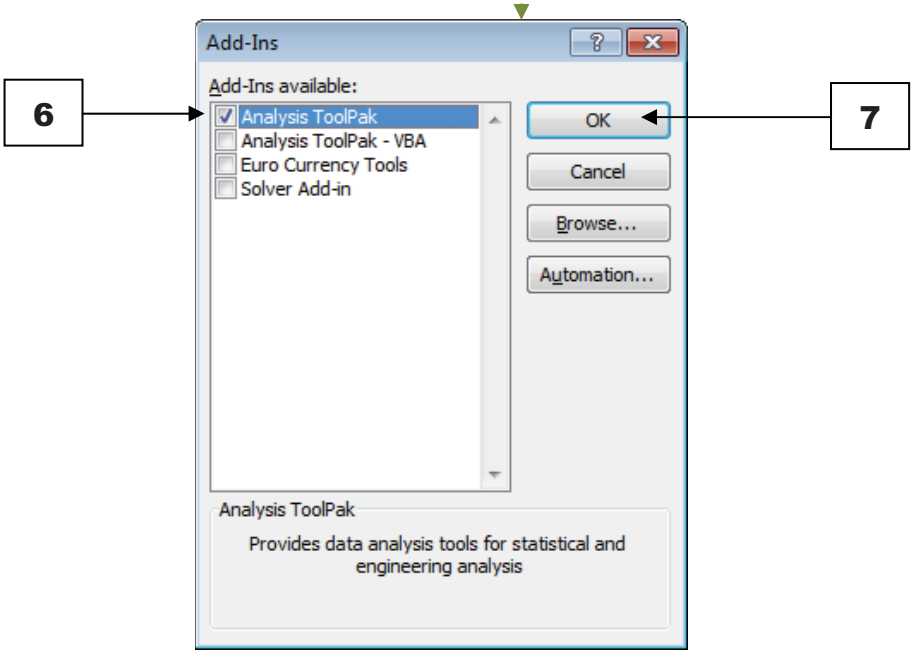


Figure 2.4 Add-Ins

*After you load the Analysis ToolPak, the **Data Analysis** command is available in the **Analysis** group on the **Data** tab (next figure).*

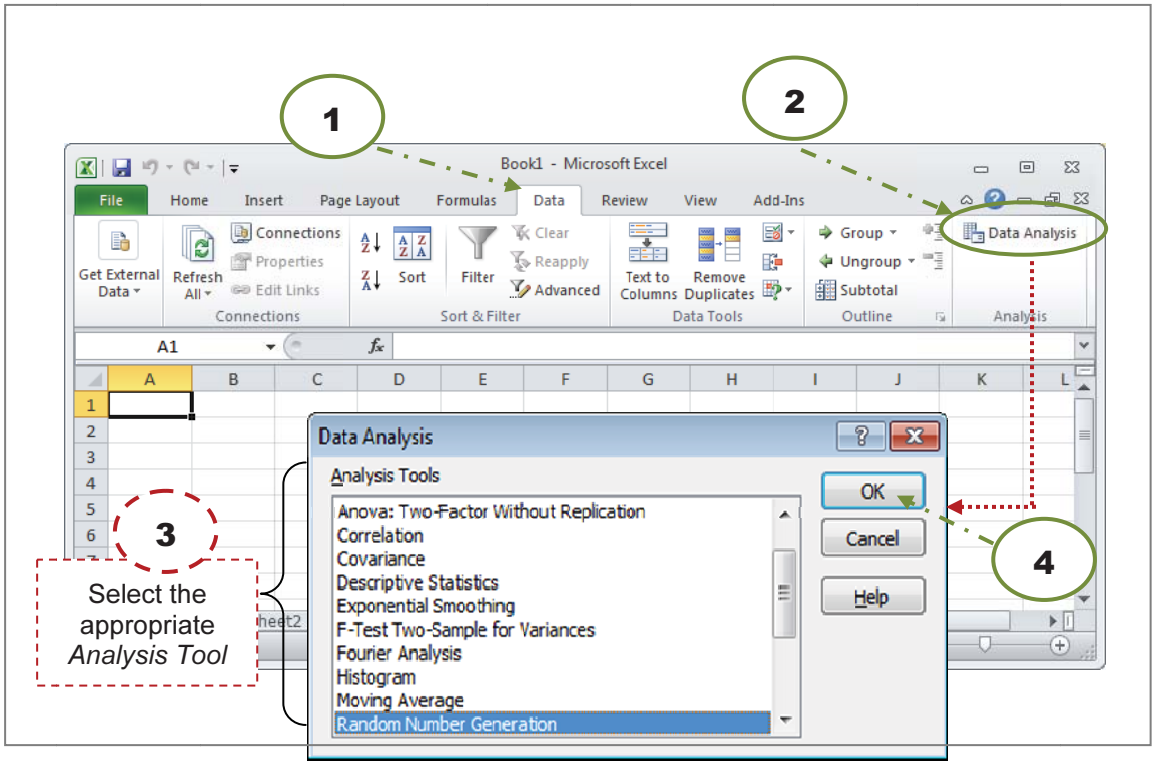


Figure 2.5 The MS Excel 2010 Data Analysis Tools

B. Analysis ToolPak for Microsoft Excel 2007 users.

Load the Analysis ToolPak (*MS Excel Help 2007*)

The Analysis ToolPak is a Microsoft Office Excel add-in program that is available when you install Microsoft Office or Excel. To use it in Excel, however, you need to load it first.

1- Click the **Microsoft Office Button** , and then click **Excel Options**.

2- Click **Add-Ins**, and then in the **Manage** box, select **Excel Add-ins**.

3- Click **Go**.

4- In the **Add-Ins available** box, select the **Analysis ToolPak** check box, and then click **OK**.

If **Analysis ToolPak** is not listed in the **Add-Ins available** box, click **Browse** to locate it.

If you get prompted that the Analysis ToolPak is not currently installed on your computer, click **Yes** to install it.

5- After you load the Analysis ToolPak, the **Data Analysis** command is available in the **Analysis** group on the **Data** tab.

Then follow similar instruction as for MS Excel Version 2010.

C. Analysis ToolPak for MS Excel 2003 users

- Open an Excel file, at the top click on the **Tools**; from the menu list select **Data Analysis**.

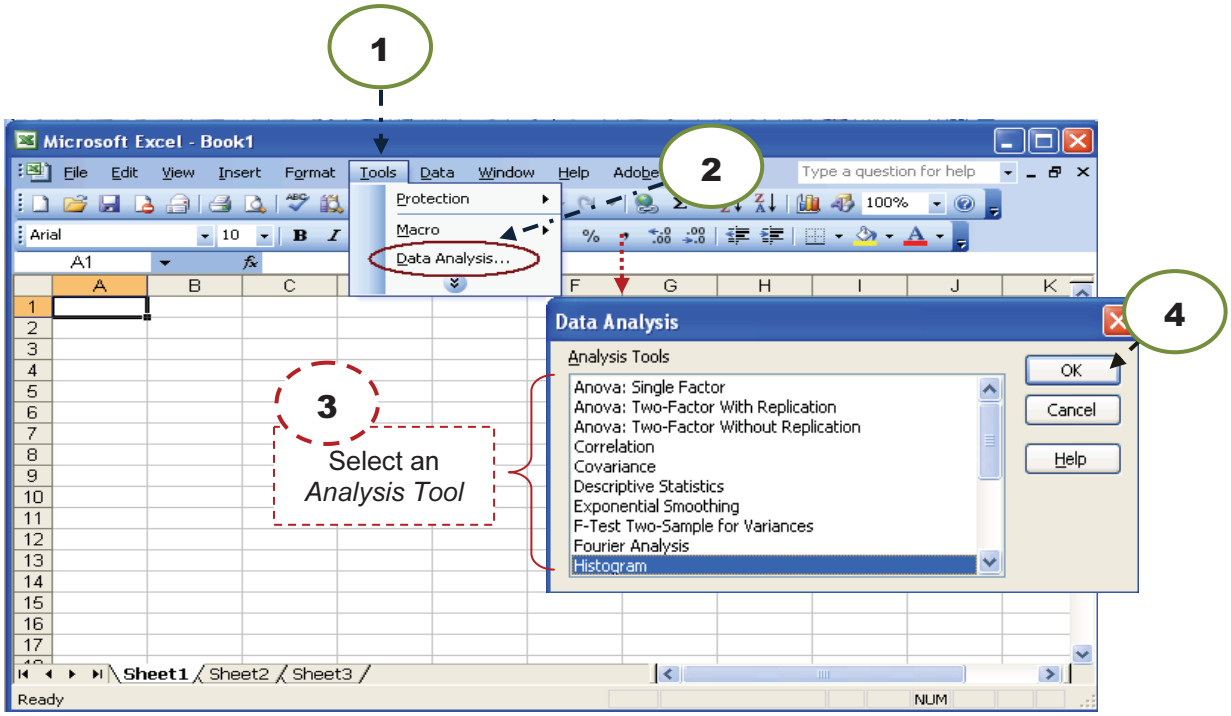


Figure 2.6 Steps in Accessing the MS Excel 2003 Data Analysis Tools

If it is *not installed* in your Excel, follow the procedure below to install it.

- Click on the **Tool** menu item at the top, from the drop list, click on Add-Ins. Select **Analysis ToolPak** and **Analysis ToolPak-VBA** and click **OK**.
- Now check again to see whether the Data Analysis is installed by going back to top of Excel file and clicking on Tools.

2.1.2 Histogram

Section 2.1.2 through 2.1.4 will take you through the steps to draw the various common charts, diagrams and descriptive statistics using data analysis tools available in Excel.

We will start with Histogram based on the data given in table 2.1 for students' total score in an exam.

Steps:

1. In a new worksheet, enter the data below:

Table 2.1 Raw Data on ID and Score

	A	B	C	D
1	Student ID	Total Scores		Bin Range
2	217127	65		60
3	217486	17		65
4	217592	61		70
5	217608	67		75
6	217613	60		80
7	217619	65		85
8	217630	75		90
9	217646	71		95
10	217650	39		100
11	217675	61		
12	217996	60		
13	226026	62		
14	226066	75		
15	226173	85		
16	226228	80		
17	226246	60		
18	226366	76		
19	226728	75		
20	226912	80		

2. Select the analysis tool **Histogram** from the options shown in Figure 2.5 or Figure 2.6 or Figure 2.7 (*depending on the version of your MS Excel*).
3. Click **OK** button.
4. To enter the **Input Range** (that is your data), click 1 **Input Range** input box shown in below (Figure 2.8). Then, click **on the first cell** (the first position of your data, in our case it is A2) **and drag down to the bottom** (B20). The column will be highlighted by dotted line and the range will appear in the dialogue box.

Or, type the cell range in the **Input Range** BOX.

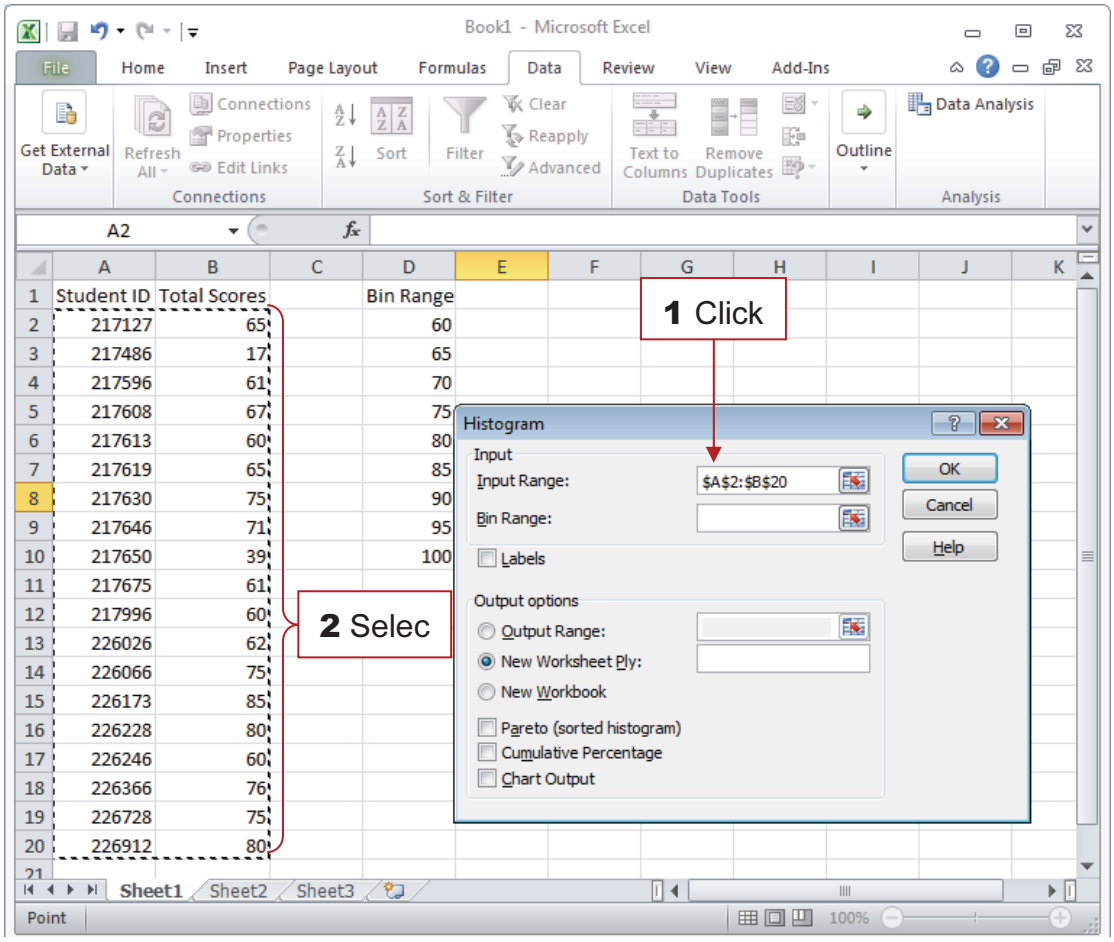
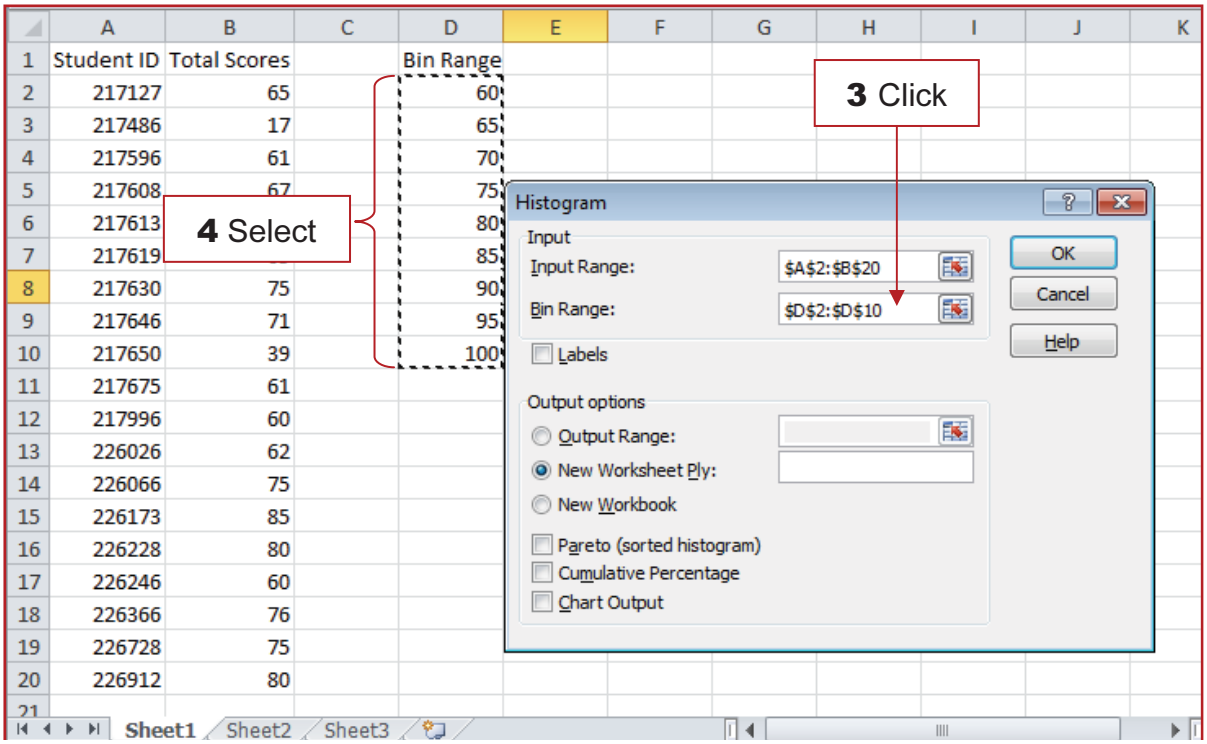


Figure 2.7 Input Range Data

You need to decide the **Bin Range**. Check your data, and decide the number of histogram bars you want and the positioning of the Histogram. This is the use of Bin Ranges. The Bin Range we have select for our illustration are 60, 65, 70...100.

5. Click in the **Bin Range** box and then highlight the Bin column and drag down to the bottom (D2:D10).



	A	B	C	D
1	Student ID	Total Scores		Bin Range
2	217127	65		60
3	217486	17		65
4	217596	61		70
5	217608	67		75
6	217613			80
7	217619			85
8	217630	75		90
9	217646	71		95
10	217650	39		100
11	217675	61		
12	217996	60		
13	226026	62		
14	226066	75		
15	226173	85		
16	226228	80		
17	226246	60		
18	226366	76		
19	226728	75		
20	226912	80		

Figure 2.8 Bin Range Data

6. For the Output Options, click on the **Output Range** and the C**hart Output**. You should have a dialogue box like the following:

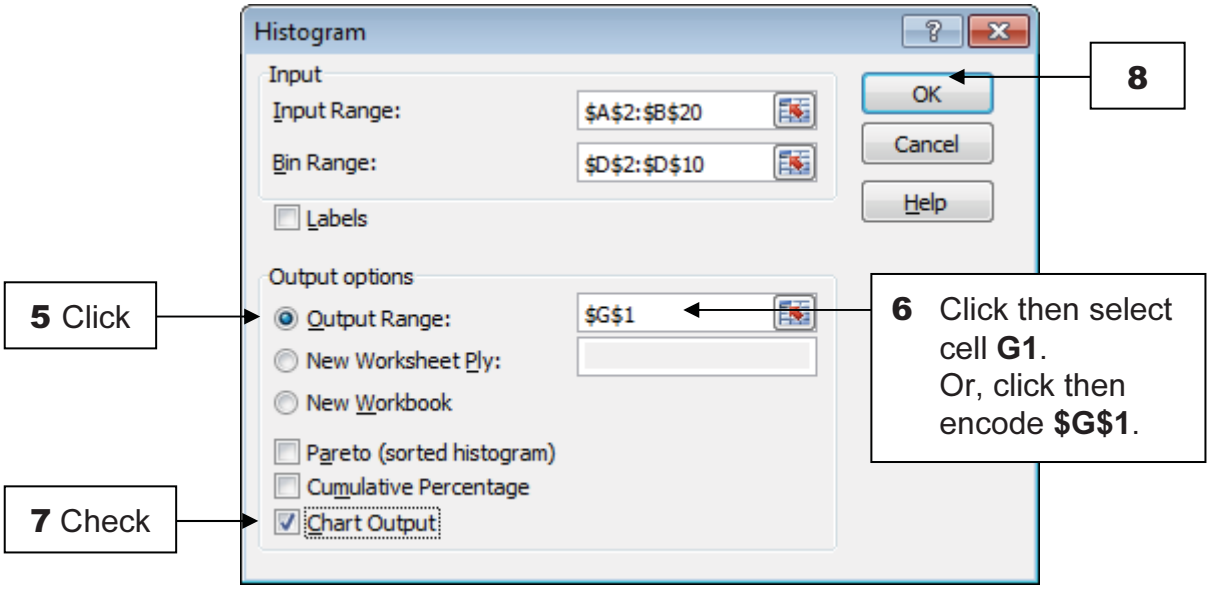
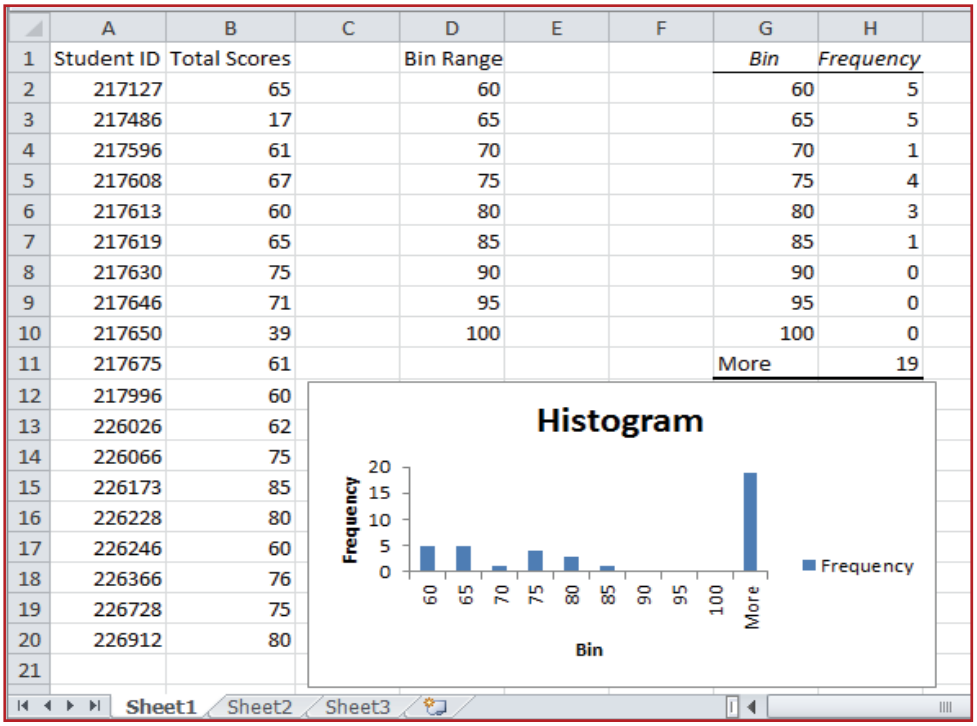


Figure 2.9 Output Range and Chart Output



	A	B	C	D	E	F	G	H
1	Student ID	Total Scores		Bin Range			*Bin*	*Frequency*
2	217127	65		60			60	5
3	217486	17		65			65	5
4	217596	61		70			70	1
5	217608	67		75			75	4
6	217613	60		80			80	3
7	217619	65		85			85	1
8	217630	75		90			90	0
9	217646	71		95			95	0
10	217650	39		100			100	0
11	217675	61					More	19
12	217996	60						
13	226026	62						
14	226066	75						
15	226173	85						
16	226228	80						
17	226246	60						
18	226366	76						
19	226728	75						
20	226912	80						
21								

Figure 2.10 Histogram

❖ You can adjust the histogram in terms of color, shading, changing the size of width and adding a title.

❖To print the histogram, click on the graph to highlight it, and then use **File ---Print**

Usefulness of Histogram:

1. to show frequency of occurrences of categories or values for one variable so that the highest and the lowest values are clear.
2. to show trends for variable.
3. to show the proportion of occurrences of categories or values for one variable.
4. to show the distribution of values for one variable.

2.1.3 DATA Summaries

Steps:

1. The following data will be analyzed:

Table 2.2 Raw Data on ID and Score

	A	B
1	Student ID	Total Scores
2	217127	65
3	217486	17
4	217592	61
5	217608	67
6	217613	60
7	217619	65
8	217630	75
9	217646	71
10	217650	39
11	217675	61
12	217996	60
13	226026	62
14	226066	75
15	226173	85
16	226228	80
17	226246	60
18	226366	76
19	226728	75
20	226912	80

2. Select DESCRIPTIVE STATISTICS from the Analysis Tools (steps shown in Figure 2.5 or Figure 2.6 or Figure 2.7 *depending on the version of your MS Excel*).

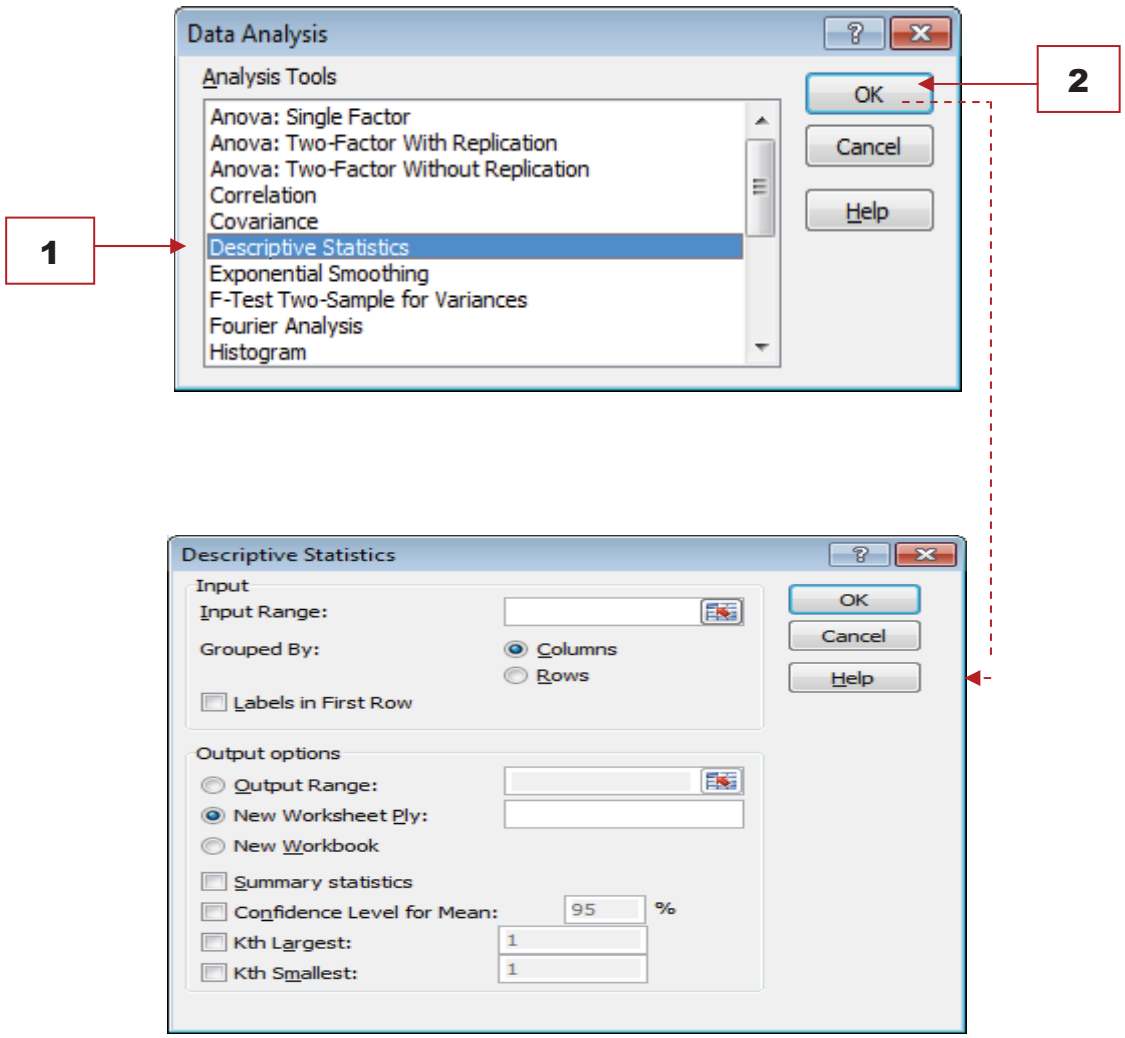


Figure 2.11 Descriptive Statistics Input and Output Options

Input:

1. Click on the **Input Range** box.
2. Highlight cells B2 to B20. (your data starting point to data ending point)

Output options:

1. Select **Output Range**, then click on the box.
2. Click cell D1 (or type in the box D1).
3. Select **Summary Statistics**.

4. Click **OK** button.

You should see the result as shown below: (adjust the column width of D to show the complete labels)

	A	B	C	D	E
1	Student ID	Total Scores		*Column1*	
2	217127	65			
3	217486	17		Mean	64.94737
4	217596	61		Standard Error	3.589681
5	217608	67		Median	65
6	217613	60		Mode	60
7	217619	65		Standard Deviation	15.64706
8	217630	75		Sample Variance	244.8304
9	217646	71		Kurtosis	4.262896
10	217650	39		Skewness	-1.76178
11	217675	61		Range	68
12	217996	60		Minimum	17
13	226026	62		Maximum	85
14	226066	75		Sum	1234
15	226173	85		Count	19
16	226228	80			
17	226246	60			
18	226366	76			
19	226728	75			
20	226912	80			
21					

Sheet1 Sheet2 Sheet3

Figure 2.12 Data Summaries

Note: by clicking on the 'Confidence Level for Mean' box in Figure 2.12, you will also have the confidence interval for your data.

Usefulness of Summary Data:

1. for exploratory purposes, where the raw data are summarized to manageable form.
2. to understand the central tendency of data → summary data describe the mean, median and mode.
3. to understand dispersion → summary data tell you the range, inter-quartile range, and standard deviation.

2.1.4 Bar Graphs, Scatter Plots and Others

MS Excel 2010 Graphs:

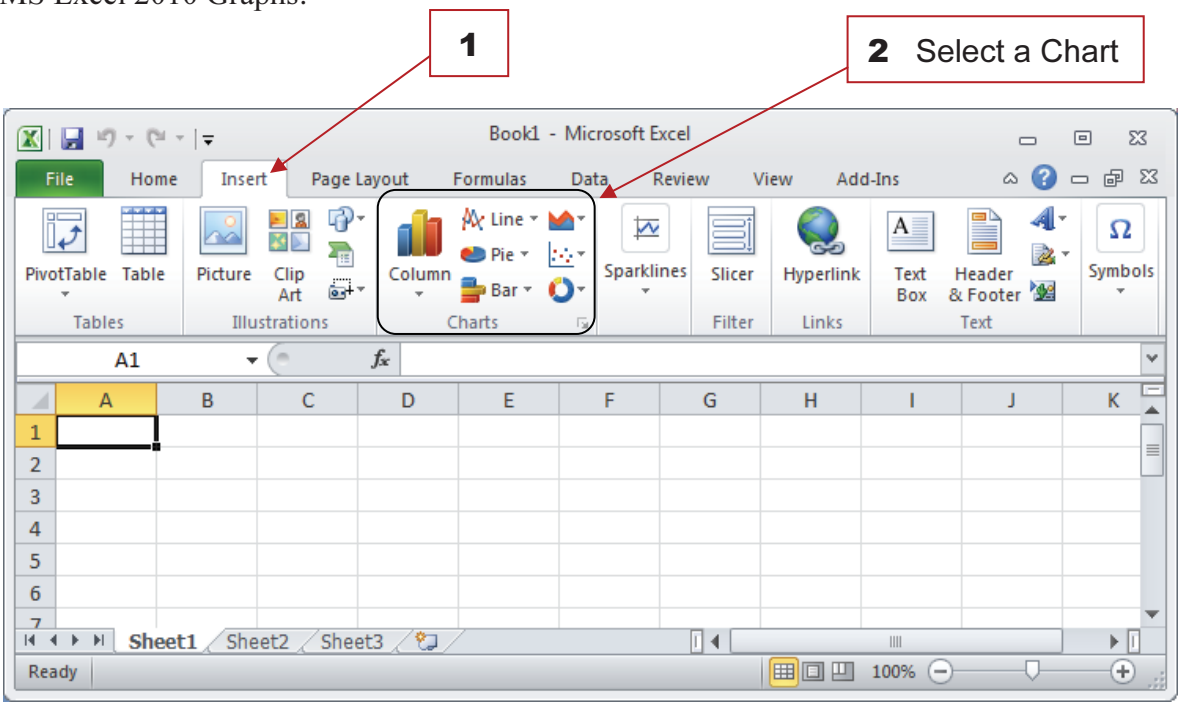


Figure 2.13 MS Excel 2010 Chart Options

MS Excel 2007 Graphs:

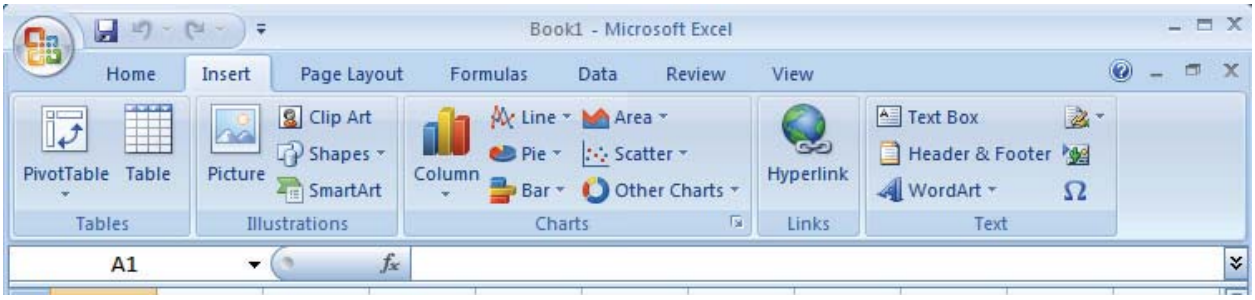


Figure 2.14 MS Excel 2007 Chart Options

Note: MS Excel 2003 Graphs and Diagrams → refer to Appendix V page 172

(a) Bar Graph

Based from the data given (Table 2.2), do the following steps below (Figure 2.16):

1 Click **any cell** from A1 to B20. In this example, cell **A1** is selected.

2 Click **Insert** tab.

3 Click **Bar** chart.

4 Select one type of a Bar chart. In this example, click the first 3-D Bar option.

The chart appears as shown in Figure 2.16.

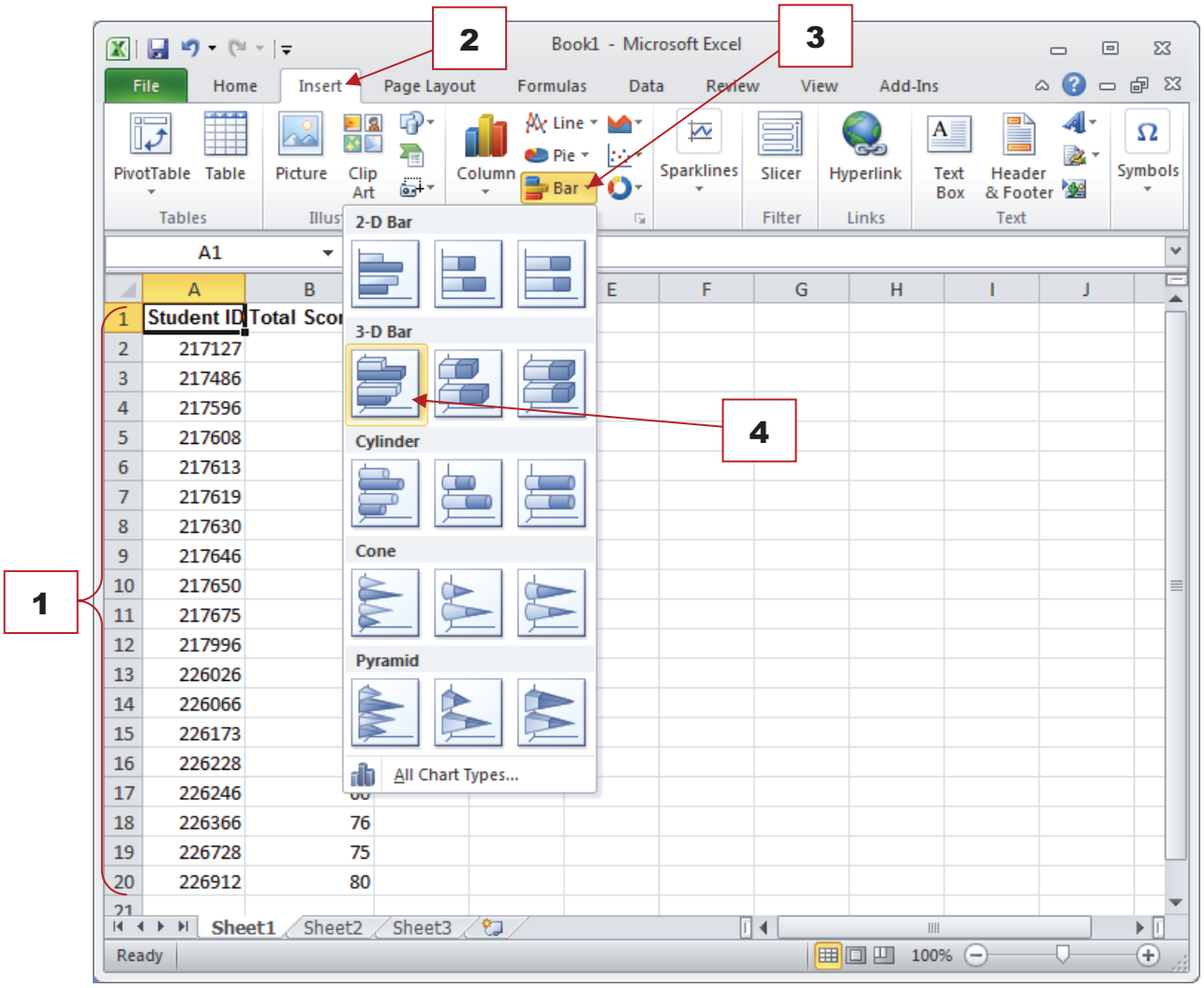


Figure 2.15 Bar Graph Initial Steps

5 Adjust or change the selected **Data range** only from B2 to B20 by dragging the border as shown below. Notice that the Chart will be changed, shown in Figure 2.17.

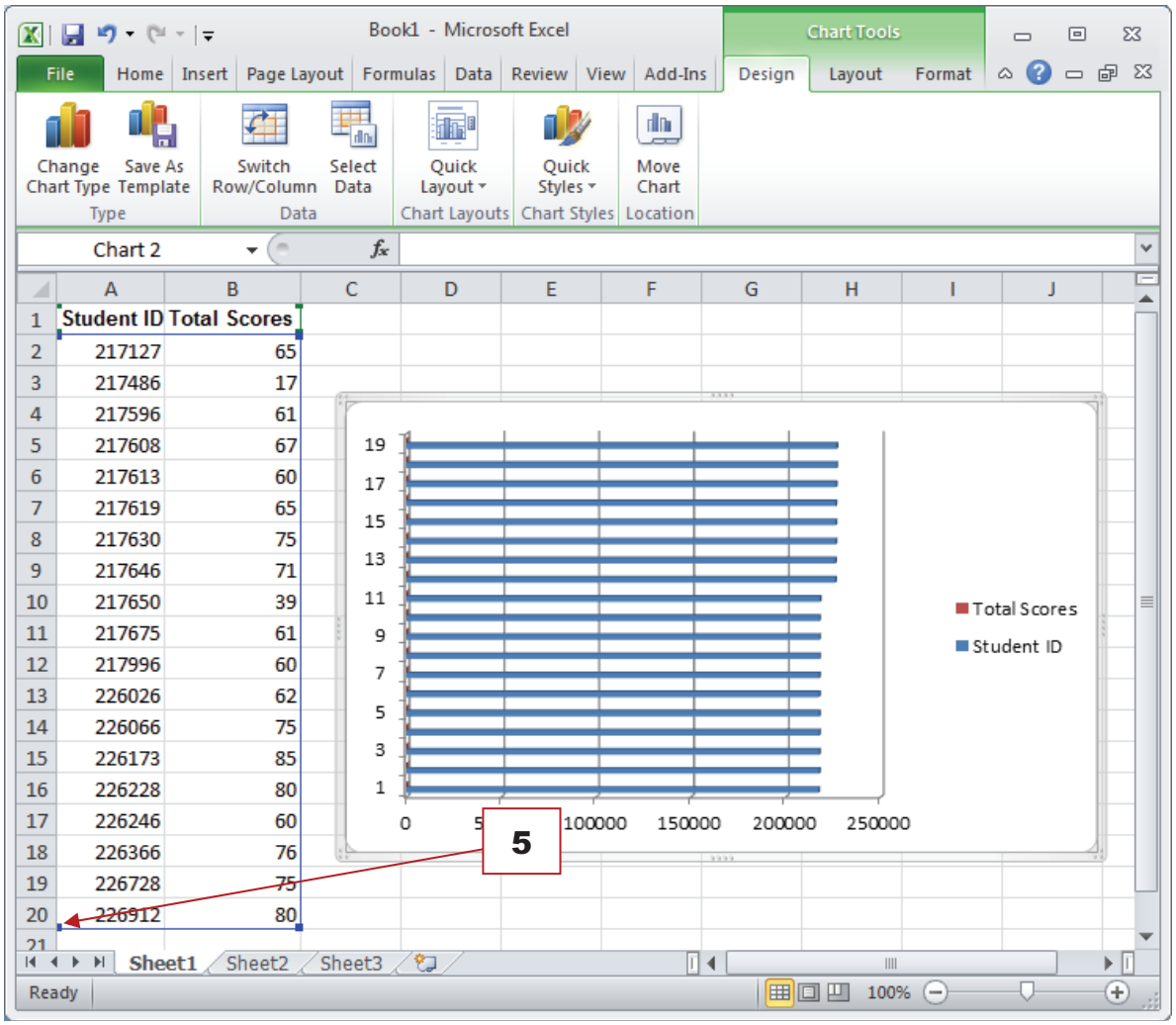


Figure 2.16 Bar Chart and the Active Data Range

To modify the LAYOUT of the chart such as Labels, Axes, etc.

6 Make sure that the chart is selected. Just click on the border or 'white' area of the graph.

7 Click the Layout tab. Check out the dropdown options in the Labels, Axes, and others.

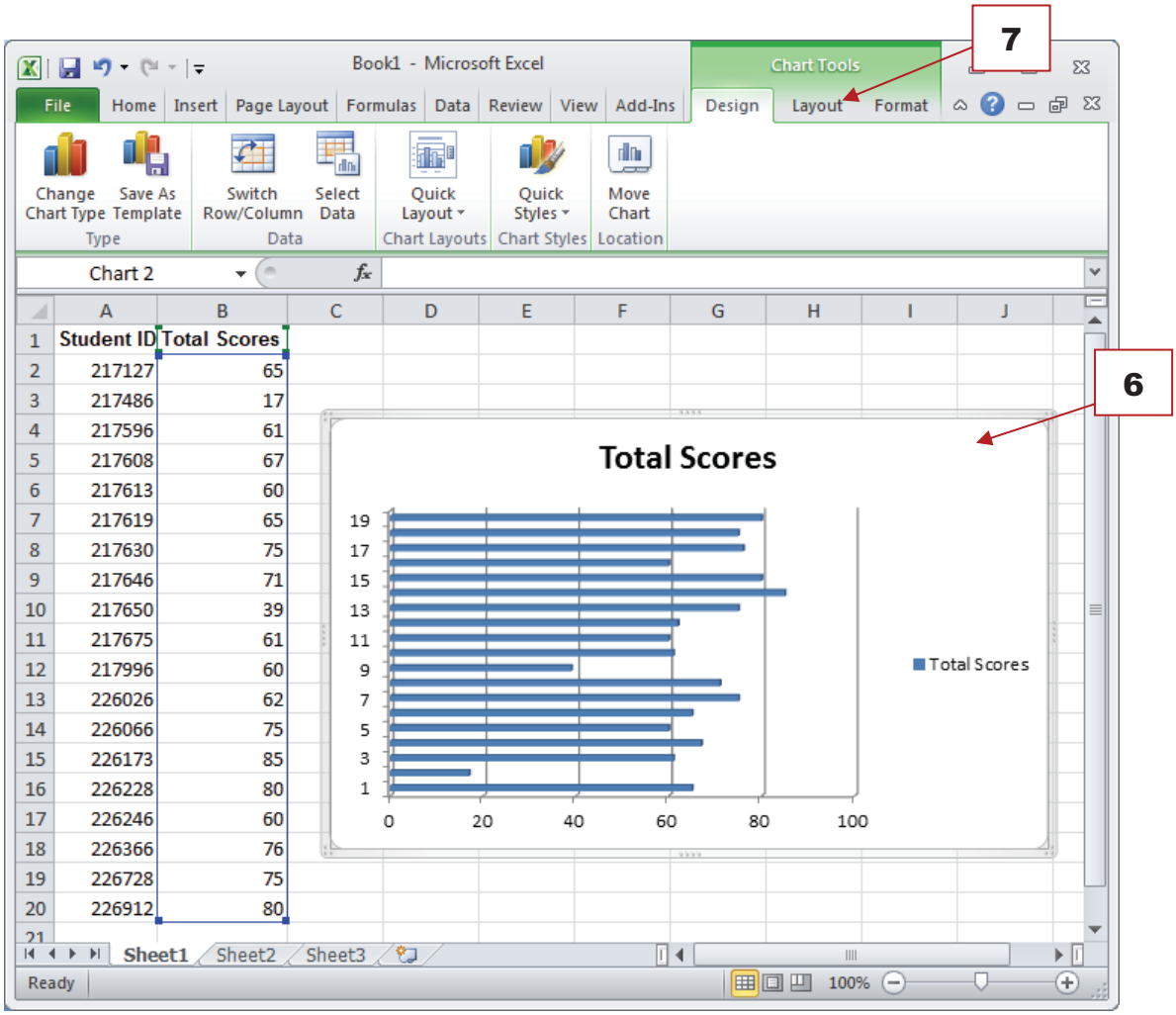


Figure 2.17 Modified Data Range and Chart Layout

Modify the following parts of the chart:

1) for Chart title, **Student Performance**.

8 Select the object title by clicking on the 'Total Scores'. And then, click again. You can modify the title, change to Student Performance.

Or

Right-click the title object 'Total Scores', then select Edit Text.

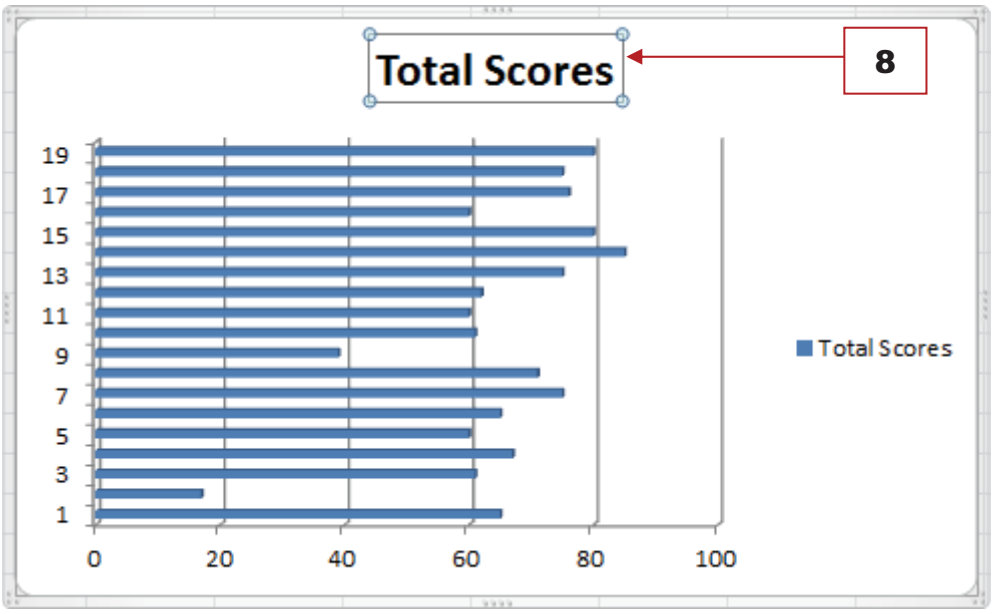


Figure 2.18 Edit functions

2) for Horizontal axis Title, type, **Total Scores**.

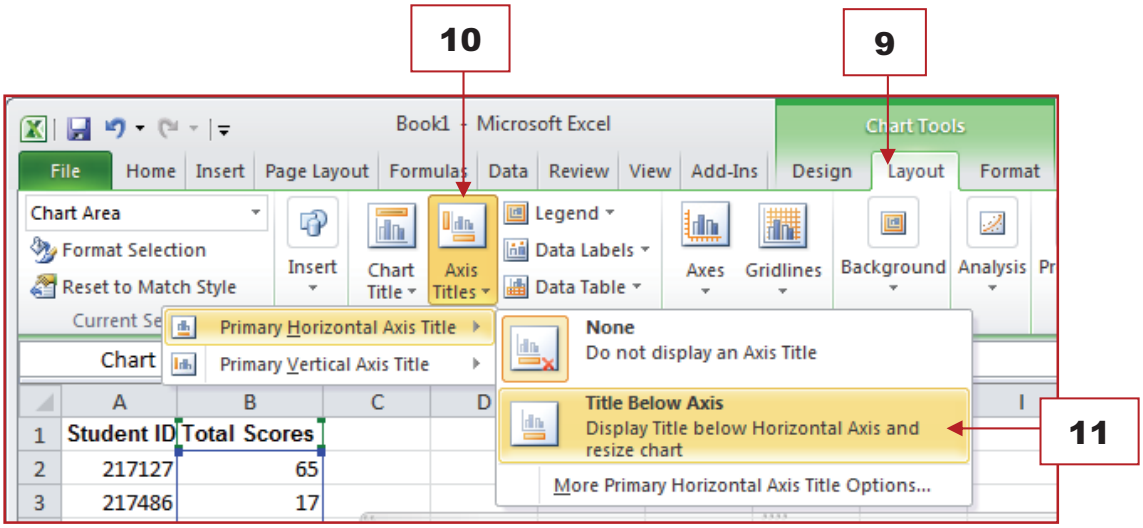


Figure 2.19 Edit functions

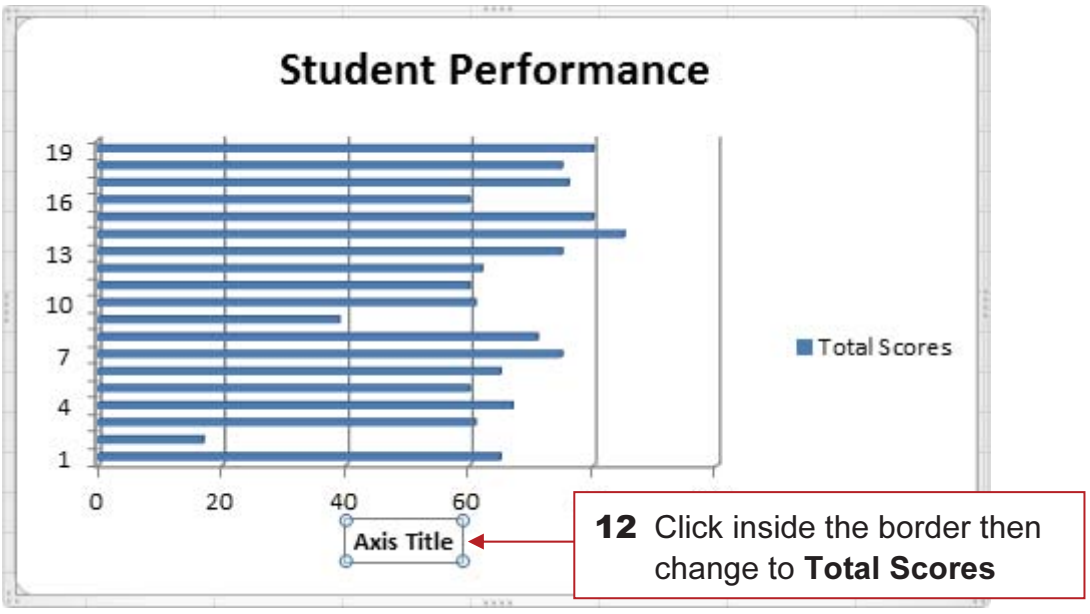


Figure 2.20 Edit functions

3) for Vertical axis, type, **Students**.

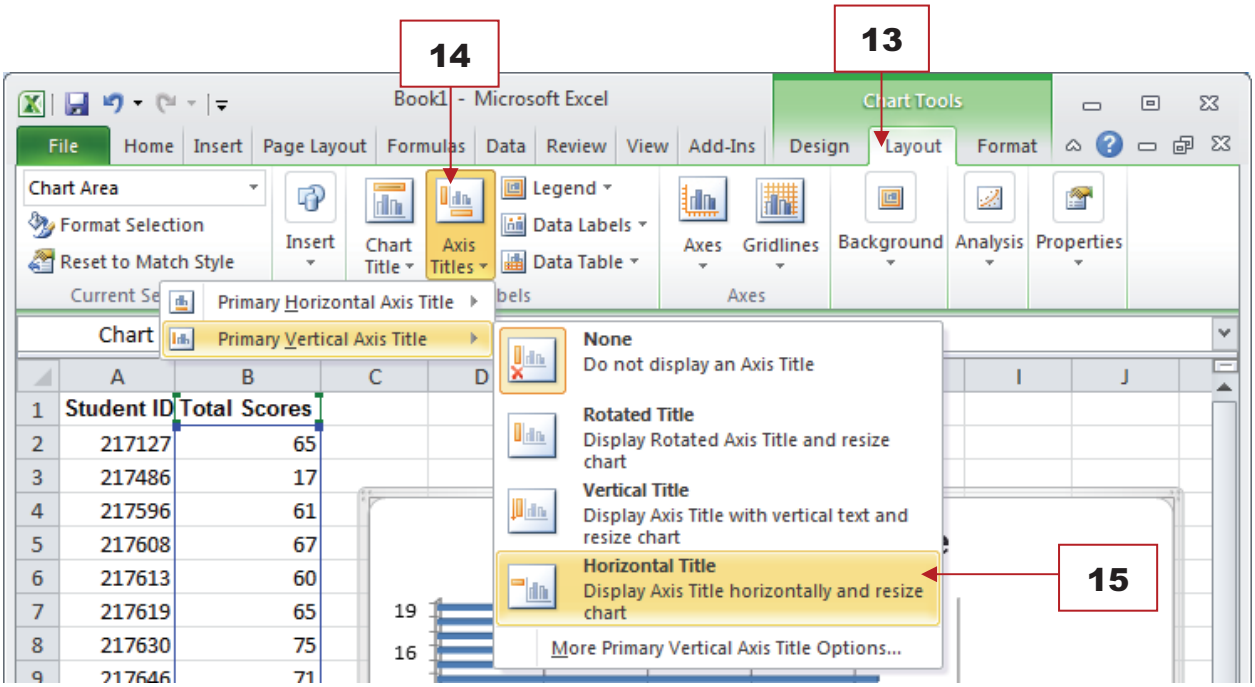


Figure 2.21 Edit functions

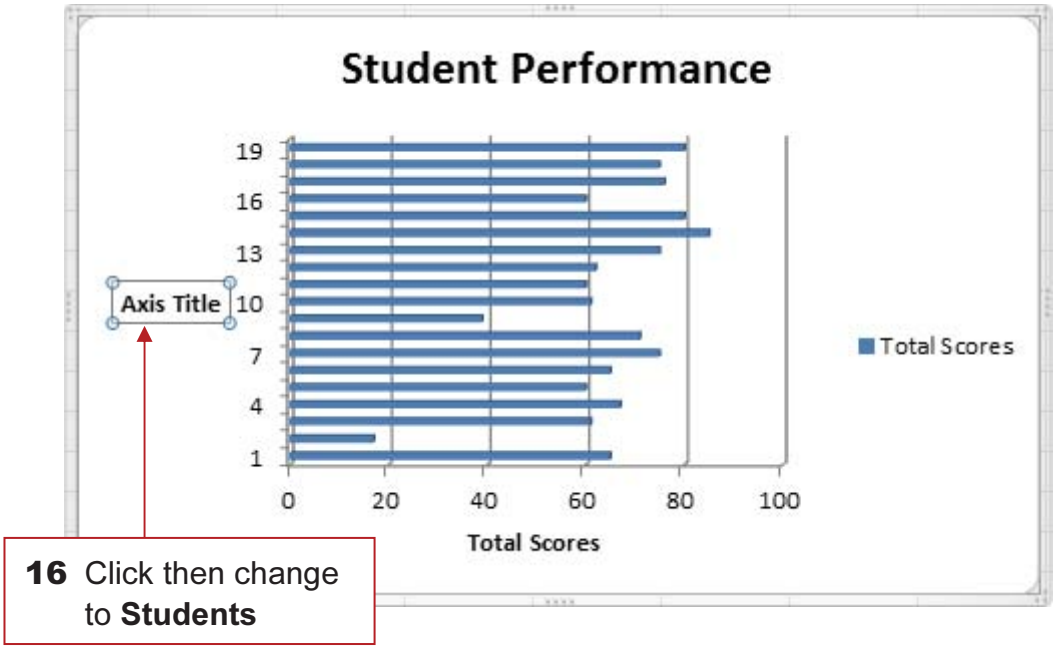


Figure 2.22 Edit functions

The result will look like the following figure:

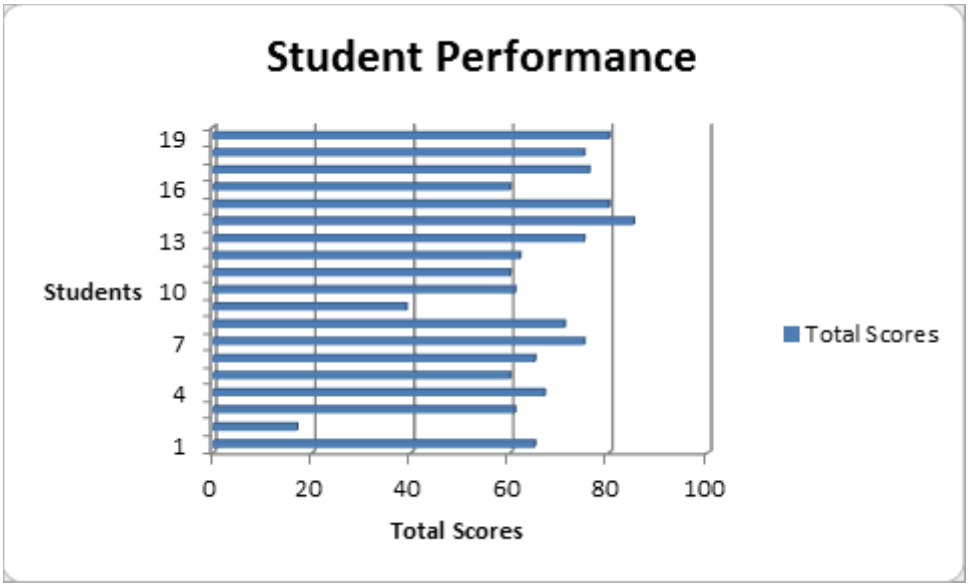


Figure 2.23 Bar Chart Sample

(b) Column Chart

Using the same data earlier, repeat the steps presented in the previous example, only the choice of the type of chart is different. This time choose 3-D Column chart.

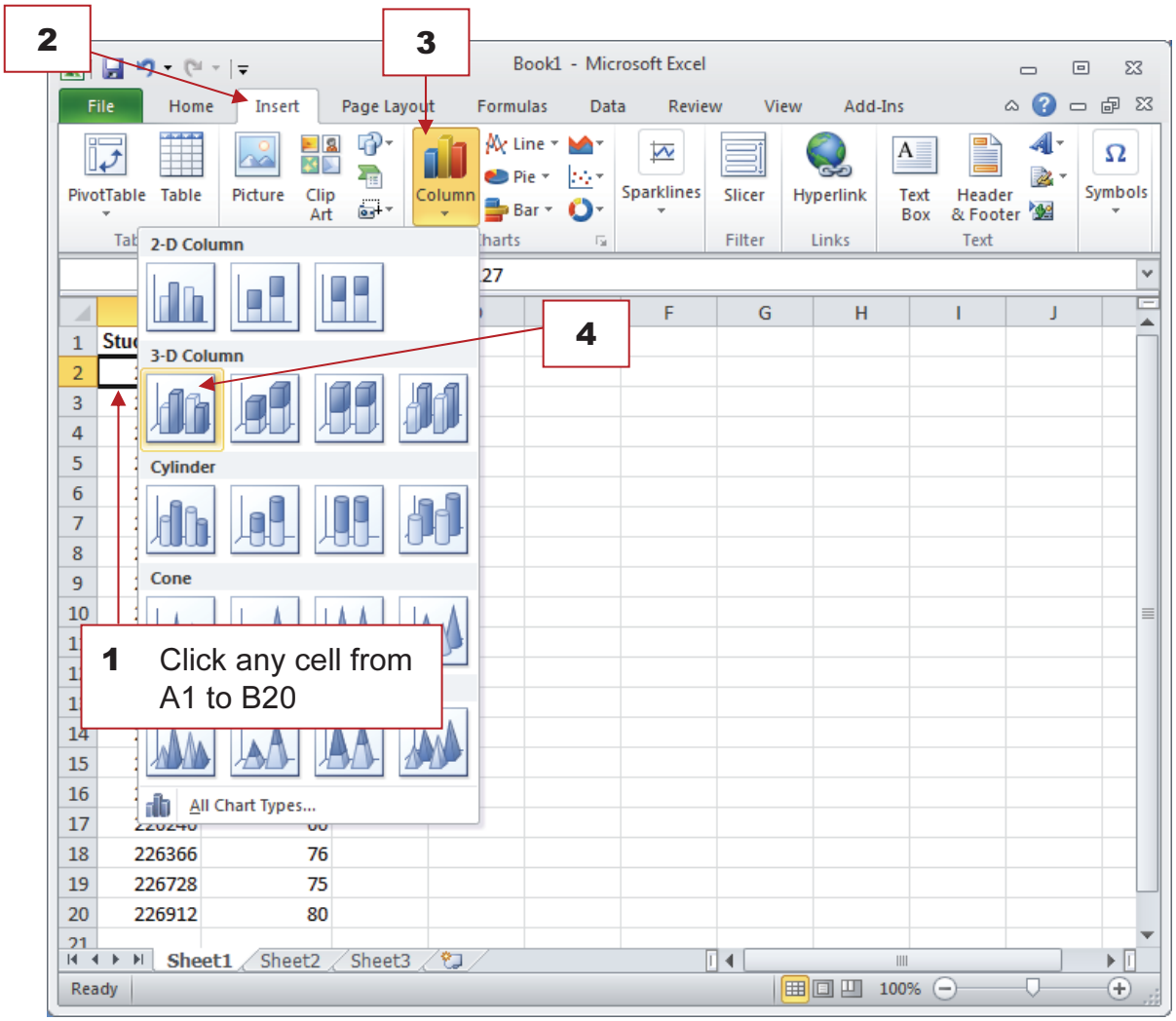


Figure 2.24 Column Chart type options

→ Repeat steps [5] to [16] of the previous example (Bar graph) to modify the layout.

NOTE: the horizontal and vertical titles must be interchanged, that is, ***Total Scores*** *(vertical axis)* and ***Students*** *(horizontal axis)*.

The figure below shows the result of these steps.

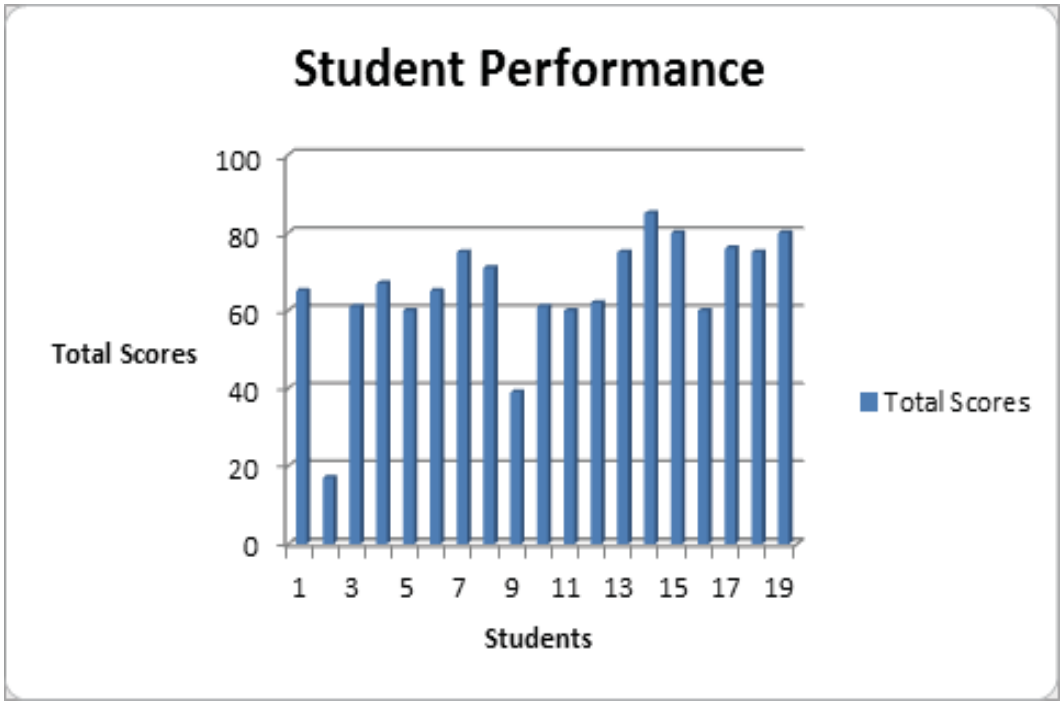


Figure 2.25 Column Chart Sample

(c) Scatter Plots

A scatter plot is an extremely useful tool when you are looking at association between two variables. It allows you to simultaneously view the values of two variables on a case-by-case basis. Here we shall illustrate the association between students and scores.

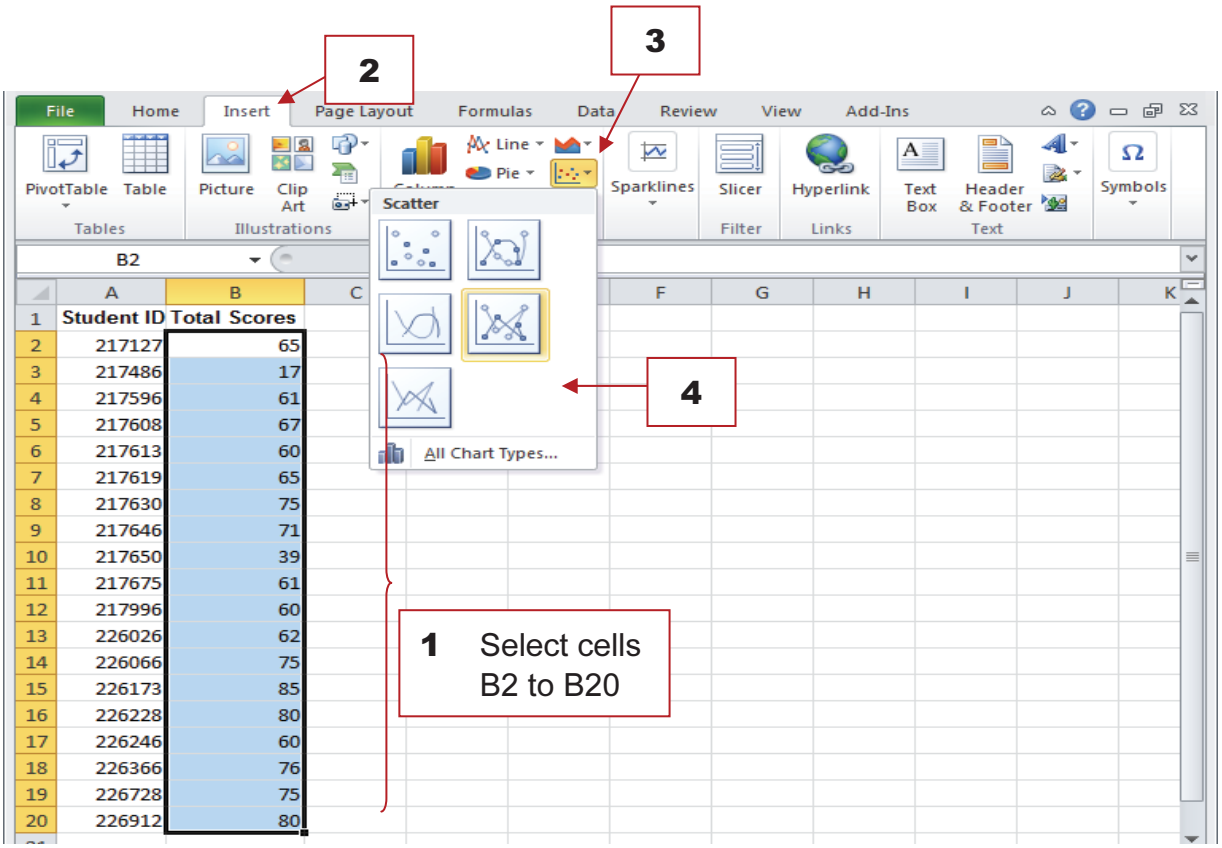


Figure 2.26 Scatter Chart Options

Follow the same steps above. The result is:

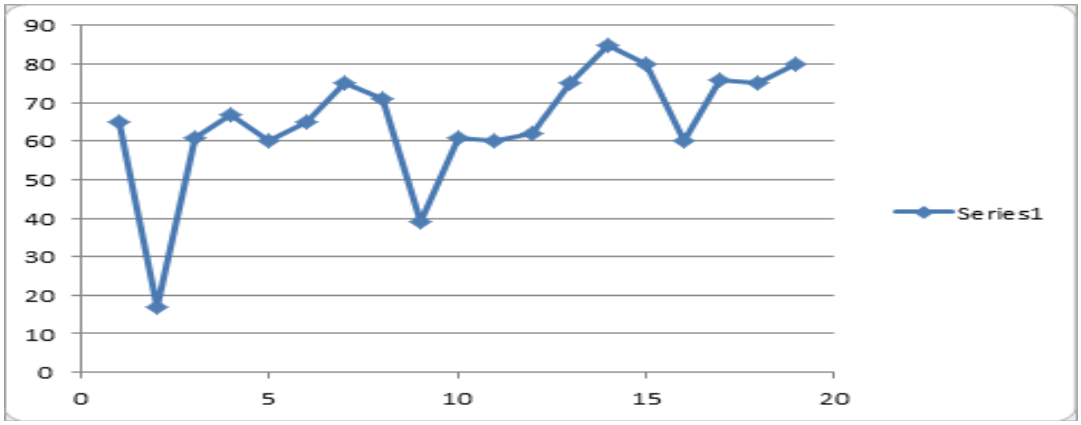


Figure 2.27 Scatter Chart Sample

(d) BOX Plot (also known as Whisker plot)

The plot is useful for analyzing *minimum, maximum, first quartile, second quartile, third quartile, spread and symmetry* in a single diagram. It is also possible to compare the above Data for two or more sample in a single diagram. *You cannot plot directly using Excel.*

An indirect guide to plotting Box plot:

- For minimum and maximum values:

Input data → **DATA → Data Analysis → Descriptive Statistics** (same steps as earlier)

(from the summary statistics you will know your maximum and minimum value - Figure 2.12)

- For the Quartile: Q1

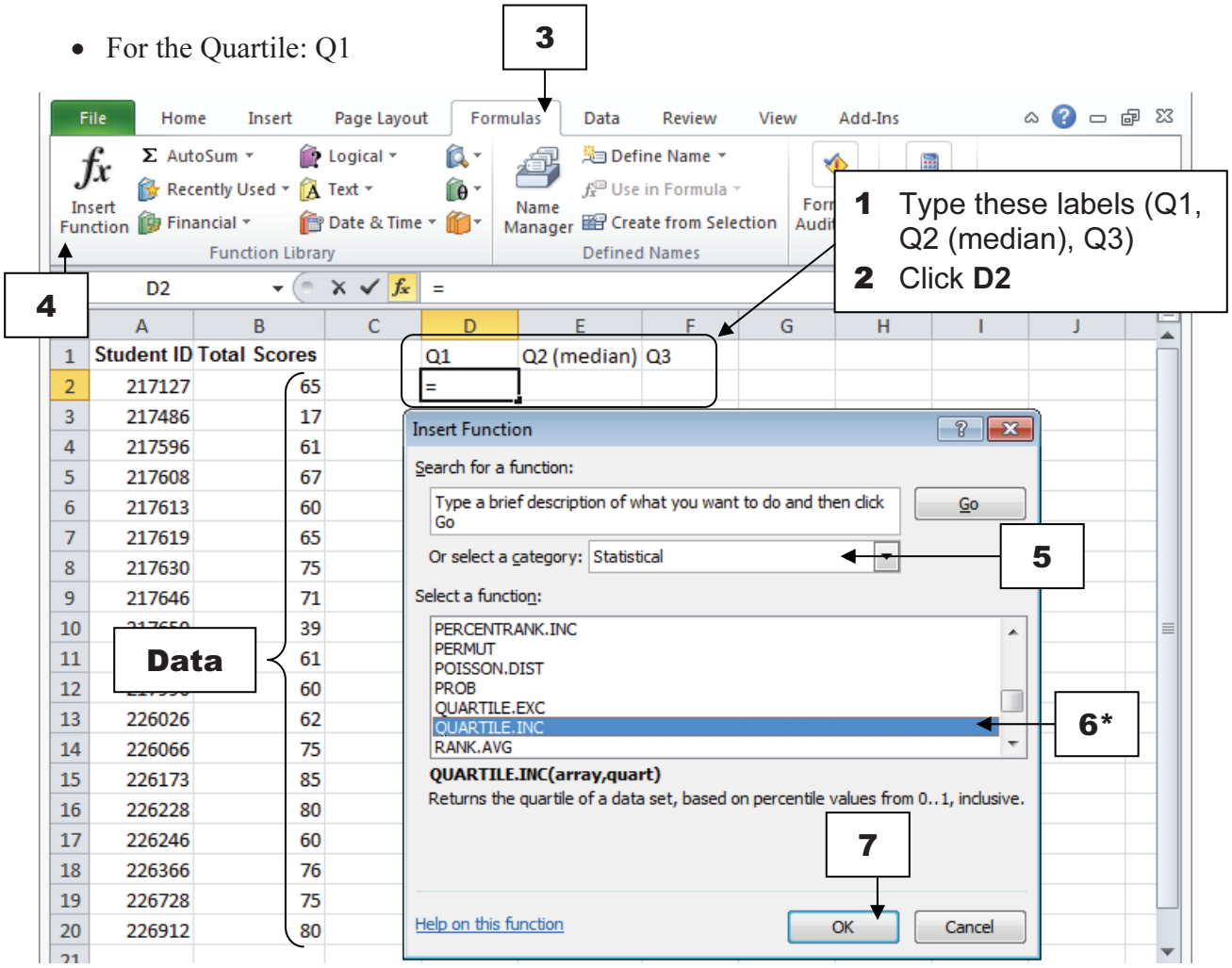


Figure 2.28 Statistical Quartile Function – *long method*

6* The function is new in Excel 2010 and so is not available in earlier versions of Excel. **Quartile.Inc** is simply a renamed version of the old Quartile function, that is available in earlier versions of Excel.

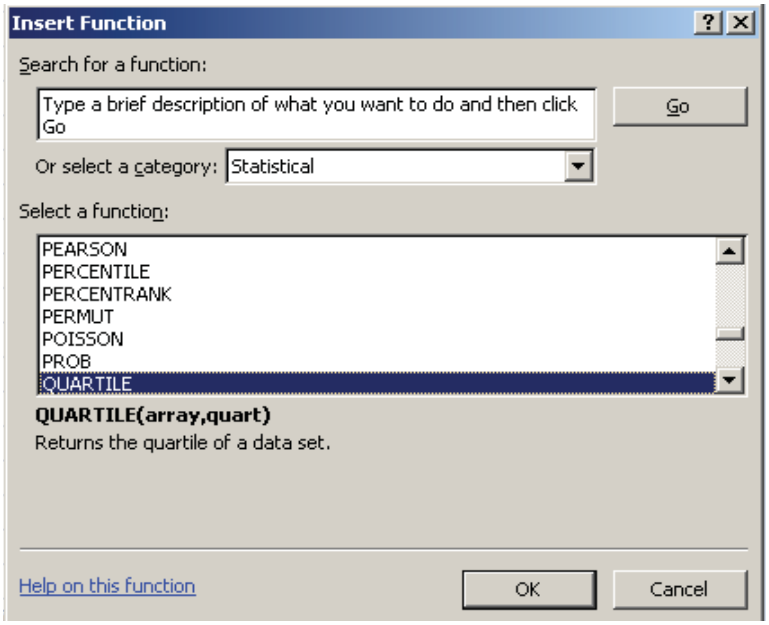


Figure 2.29 Old Quartile function available in 2007 and earlier versions of Excel

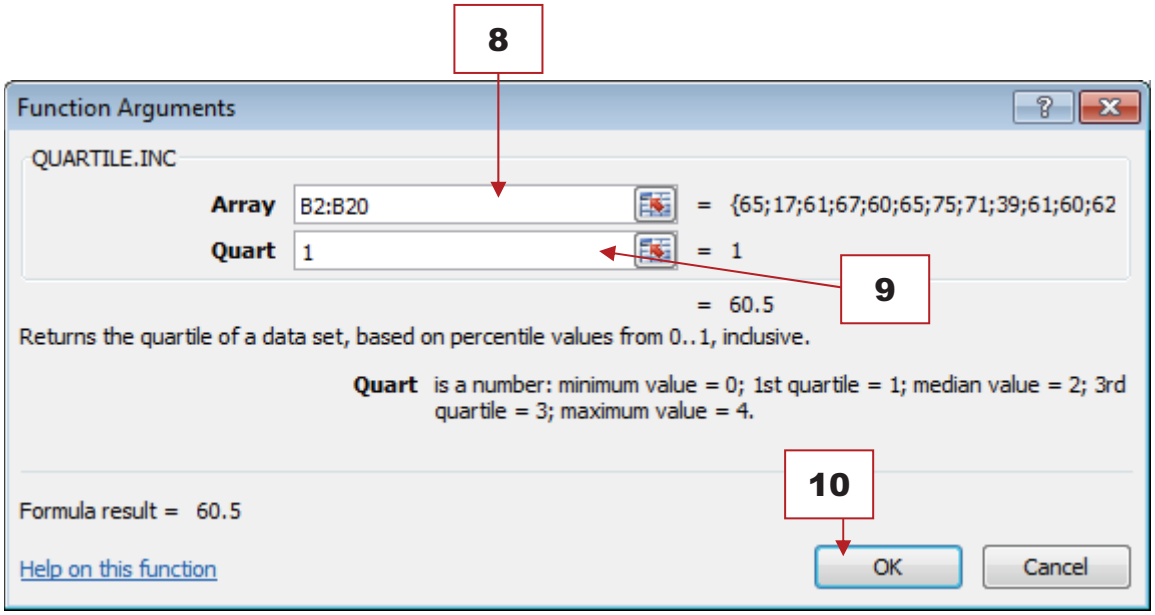


Figure 2.30 Function Arguments

- For the Quartile: Q2

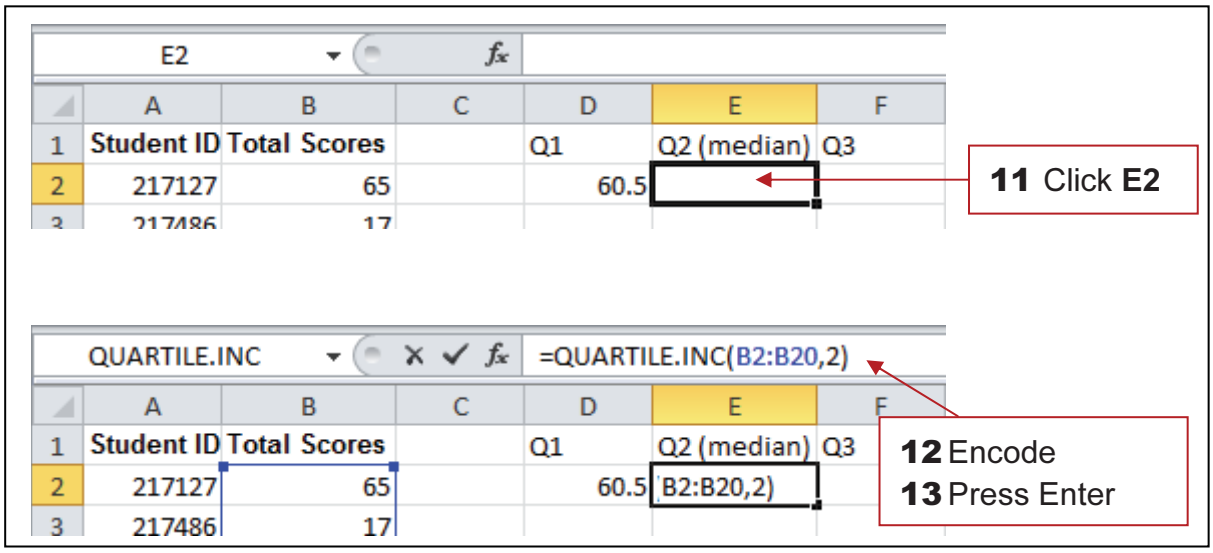


Figure 2.31 Statistical QUARTILE.INC Function (MS Excel version 2010) – *other method*

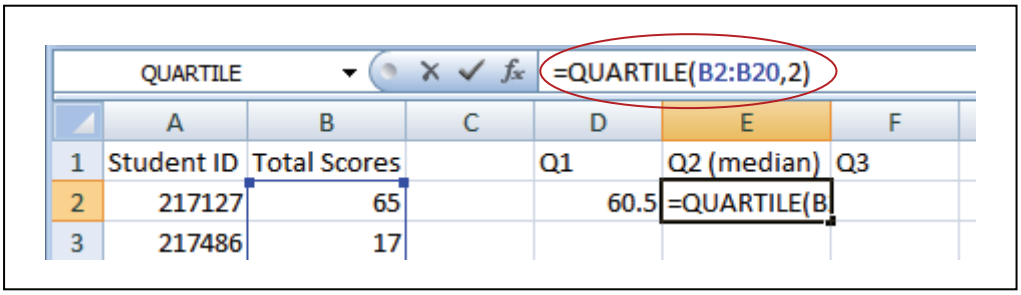


Figure 2.32 Statistical QUARTILE Function (earlier versions of Excel) – *other method*

- For the Quartile: Q3

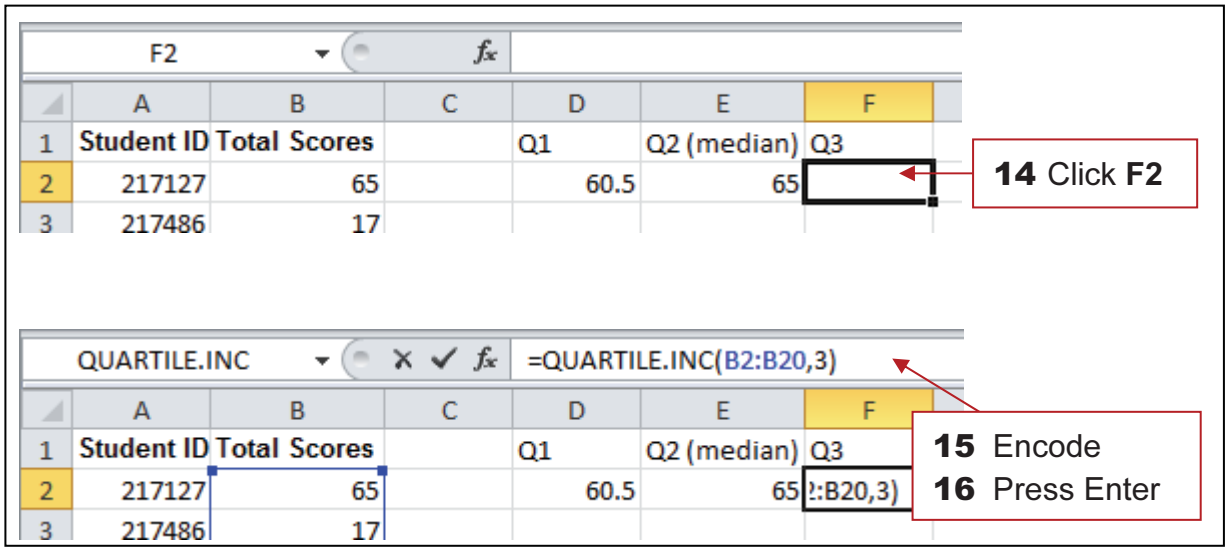


Figure 2.33 Encoding quartiles

Result:

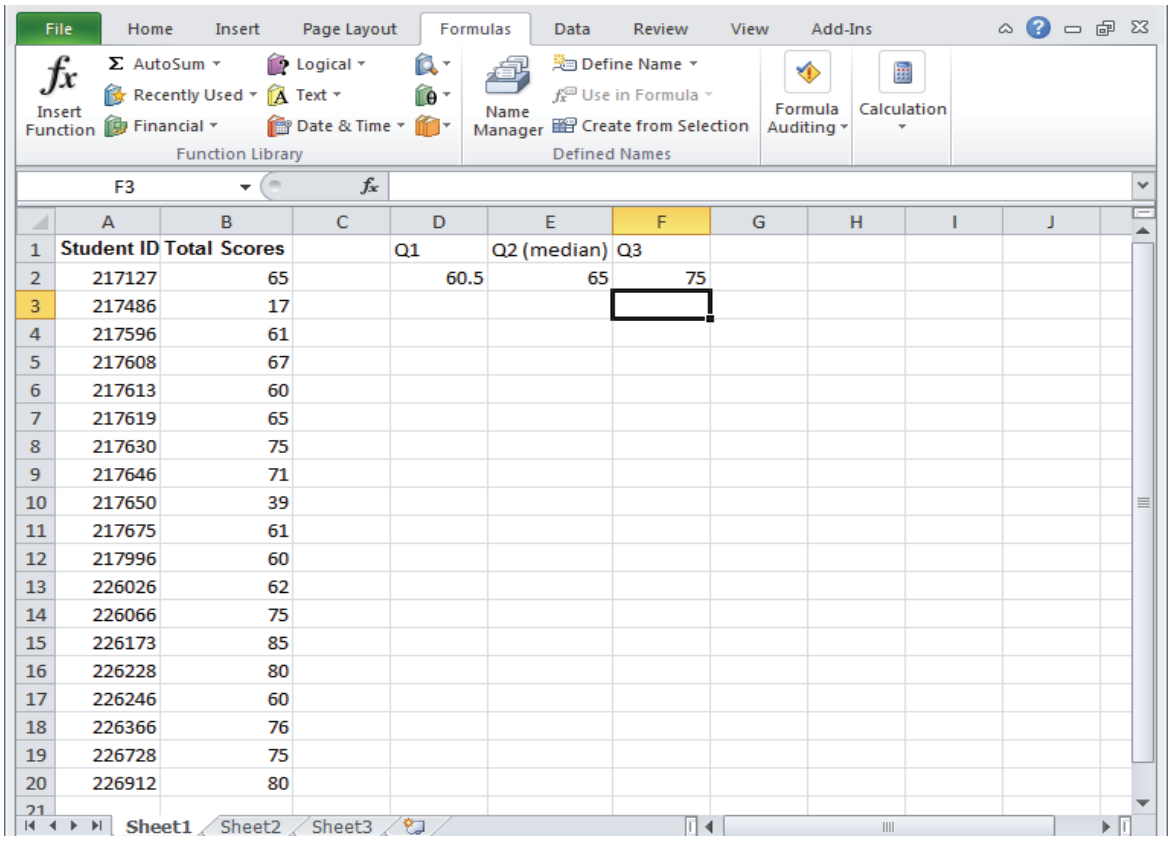


Figure 2.34 Quartile output

With minimum, maximum, first, second and third quartiles you can draw the box plot.

Examples of Box Plots

The Box Plot for a Single Set of Data

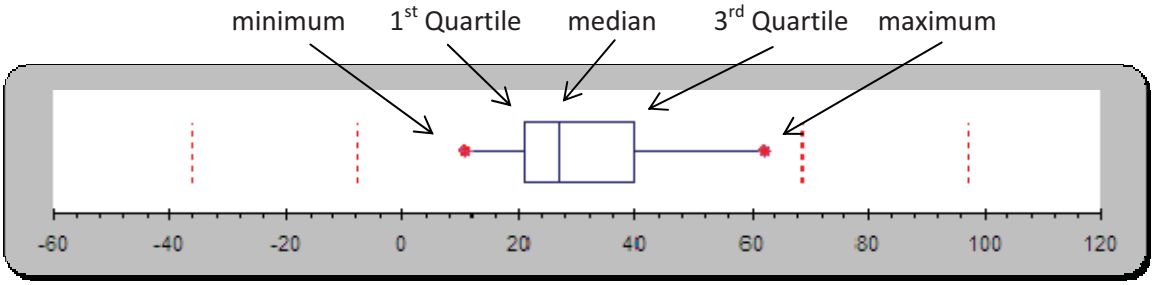


Figure 2.35a Box Plot for a Single Set of Data

Comparing Two Data Sets Using Box Plot

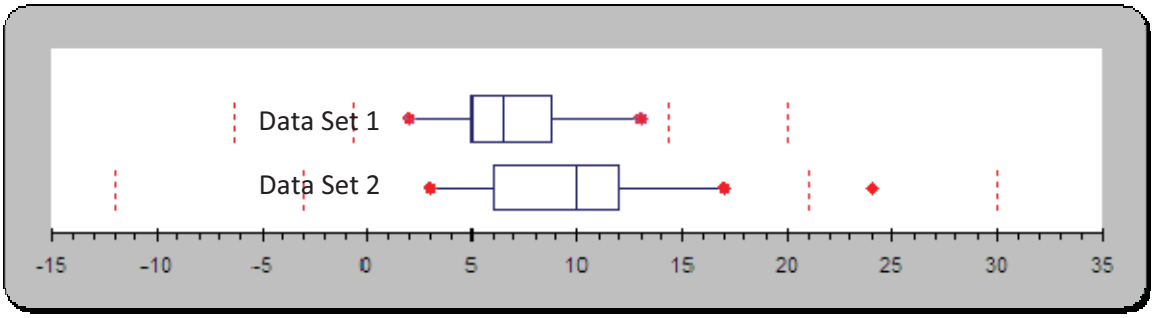


Figure 2.35b Box Plot for Two Data Sets

(Refer to statistics book for more details.)

Chart Usefulness:

Bar Charts

1. to show frequency of occurrences of categories or values for one variable so that the highest and the lowest values are clear.

2. to show trends for variable.

3. to show the proportion of occurrences of categories or values for one variable.

4. to show the distribution of values for one variable.

5. to compare the frequency of occurrences of categories or values for two or more variables so that highest and lowest are clear- use multiple bar chart.

6. to compare the trends for two or more variables- use multiple bar chart.

7. to compare proportions of occurrences of categories or values for two or more variables- use percentage component bar chart.

8. to compare the frequency of occurrences of categories or values for two or more variables so that totals are clear- use stacked bar chart.

Line Charts

1. to show trends for variable.

2. to compare trends – use multiple line graph.

Scatter Plots

1. to show the relationship between two variables.

Box Plots

1. to show the distribution of values for one variable.

2. to compare the distribution of values for two or more variables – use multiple box plot.

2.2 Correlation and Regression

2.2.1 Correlation

When two variables are associated, they can be associated in several ways. A scatter plot can provide a graphic and concise statement as to the general relationship or association between two variables. However, to quantify the association, correlation analysis is used. It provides the strength and direction of association between two variables.

Statistically, it is the cross products of the Z-scores (of the two variables X, Y) divided by the degree of freedom (n-1) where n is the sample size.

The Pearson's correlation coefficient is:-

$$\mathbf{r = \sum (Z_X \, . \, Z_Y) \, / \, n - 1}$$

But for computational purpose, the formula can be presented as:

$$r = \{n \sum \mathbf{XY} - (\sum \mathbf{X}) \, (\sum \mathbf{Y})\} \, / \, \sqrt{\{[\, n \sum \mathbf{X^2} - (\sum \mathbf{X})^2] \, [\, n \sum \mathbf{Y^2} - (\sum \mathbf{Y})^2]\}}$$

Given the increasing use of computers and statistical software, however, the important issue is whether or not you have a good understanding of what lies behind the procedure and how to interpret the results. Therefore, we *need not* go into the derivation of formula but will focus on the use of software (Excel) to derive the coefficient and its interpretation.

Table 2.3 shows a hypothetical distribution of GPA and IQ for 10 cases. The following steps will show you how to analyze the relationship using the correlation tool in Excel.

Directions:

1. Encode the following data in a worksheet. *(We assume GPA the dependent variable as Y and the independent variable IQ as X)*

Table 2.3 Raw Data on GPA and IQ

	A	B
1	GPA	IQ
2	3.70	117
3	2.60	90
4	3.30	105
5	2.20	116
6	2.80	107
7	3.20	105
8	3.80	121
9	4.00	113
10	3.00	90
11	2.40	87

2. For correlation coefficient 'r', go to: **Data → Data analysis → Correlation**

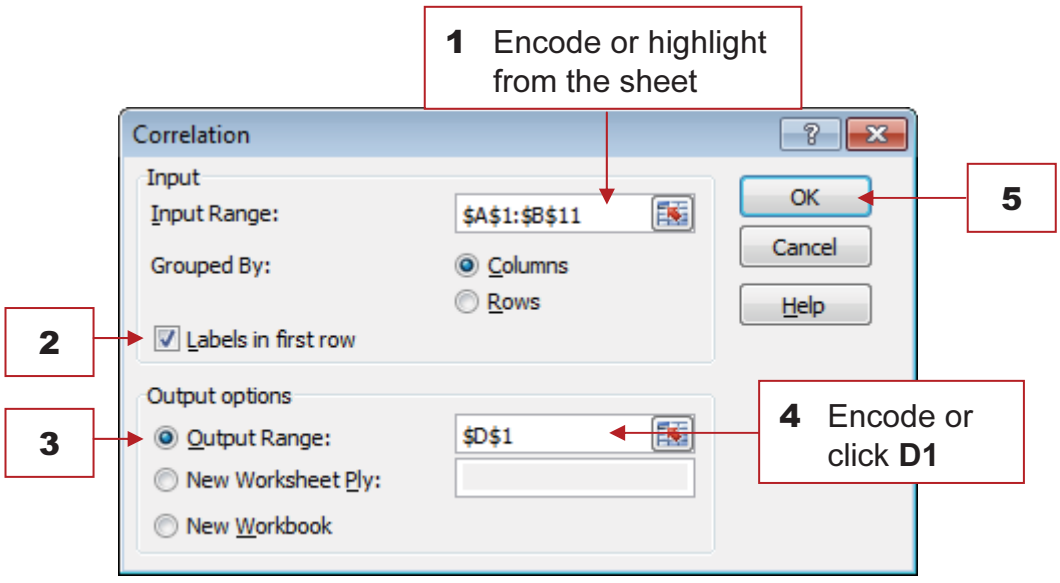


Figure 2.36 Correlation Input and Output Options

	A	B	C	D	E	F
1	GPA	IQ			*GPA*	*IQ*
2	3.70	117		GPA	1	
3	2.60	90		IQ	0.542331	1
4	3.30	105				
5	2.20	116				
6	2.80	107				
7	3.20	105				
8	3.80	121				
9	4.00	113				
10	3.00	90				
11	2.40	87				

Figure 2.37 Correlation Output

The correlation coefficient between GPA and IQ is 0.542331 (the value where the GPA and IQ cross in the table of output).

Use the following values of "r' to interpret the strength of relationship between the variables.

Table 2.4 Interpretation of r values

r = 0	r =1	.90 < r < .99	.70 < r < .89	.50 < r < .69	.30 < r < .49	.01 < r < .29
No correlation	Perfect positive	Very strong positive	Strong positive	Moderate positive	Weak positive	Very weak positive

r = -1	-.99 < r < -.90	-.89 < r < -.70	-.69 < r < -.50	-.49 < r < -.30	-.29 < r < .01
Perfect negative	Very strong negative	Strong negative	Moderate negative	Weak negative	Very weak negative

Note: *r^2 is known as coefficient of determination, and it is the amount of variation in one variable that is attributable to variation in another variable.*

In our example r =0.5423, and r^2 = 0.2941

Multiplying r^2 by 100% tells us that the variation in the level of IQ explains 29.41% of the variation in the GPA.

Usefulness:

Correlation

1. to assess the strength of relationship between two variables – correlation coefficient.
2. to assess the strength of relationship between one dependent variable and one or more independent variables – regression analysis.

2.2.2 Simple Linear Regression

A regression line is a line of best fit that passes through the points in a scatter plot in such a way that it provides the best representation of the overall association between the variables. This line is also known as least square line and allows us to predict the value of Y (dependent variable) on the basis of the value of X (independent variable).

As a predicting tool, we need to rely on regression equation. It is defined as follows:

$$Y' = a + bX$$

The elements of the formula are:

Y' - value that we are attempting to predict (in this case, a particular GPA); the symbol is read as Y-prime and stands for a predicted value. It can also be written as $\hat{y}$.

X- value we are given (in this case, IQ score)

a- the point where the regression line, cuts the Y-axis, also known as Y-intercept

b- the slope of the regression line (the amount of change in Y that is associated with a unit change in X)

A further useful term that you need to know is the standard error of estimate.

Standard Error of the Estimate- Any prediction you make will be subject to some amount of error. In regression, overall expression of potential error in an estimate of Y' is referred to as the standard error of the estimate. It is an overall measure of the extent to which the predicted Y' values deviate from the actual Y values.

Statistical Analysis Using Excel Software

The use of software (Excel) helps us in determining the values of *a* and *b*, estimate of standard error, t-statistics, P-values and confidence interval and other useful information for analyzing regression line without the need for elaborate calculations. The following section will introduce the steps to regression analysis.

Directions:

1. Use the same data as in **Figure 2.27**.
2. For regression with fitted line, go to: **Data → Data analysis → Regression**

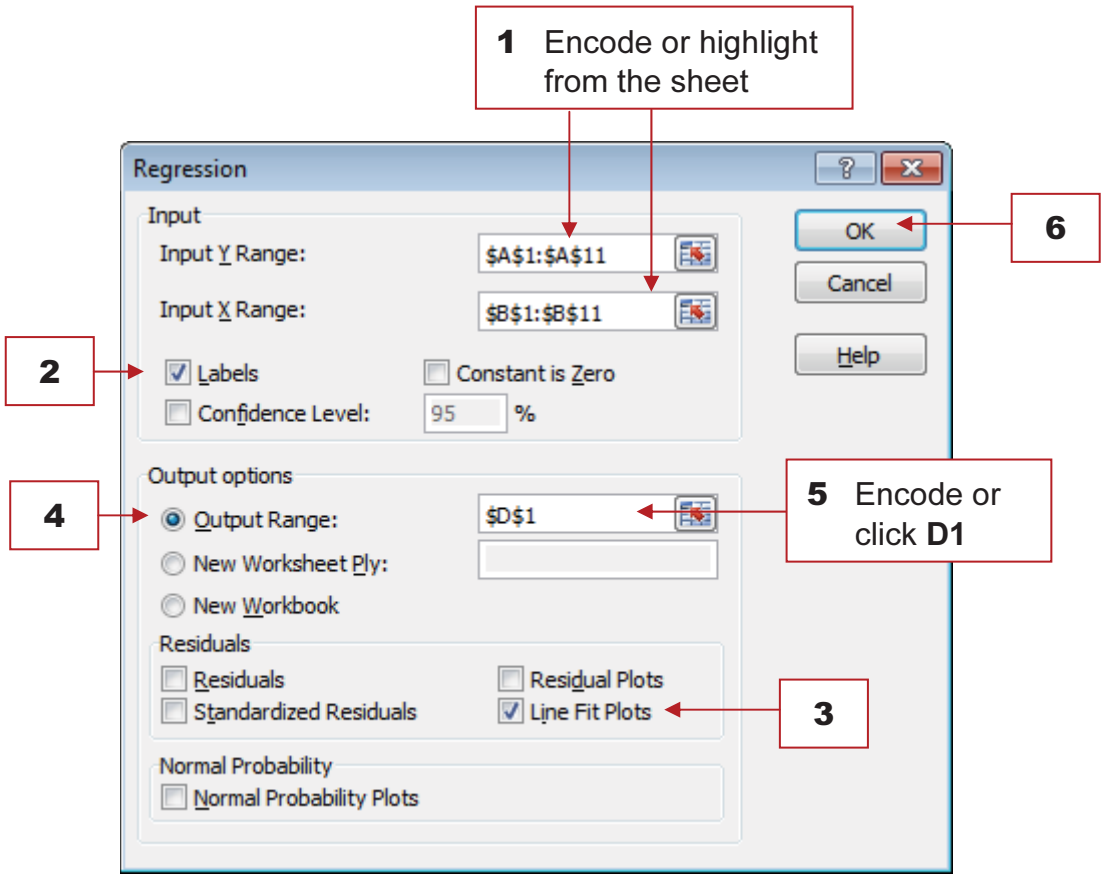


Figure 2.38 Regression Input, Output, Residual, and Normal Probability Options

	A	B	C	D	E	F	G	H	I	J	K	L
1	GPA	IQ		SUMMARY OUTPUT								
2	3.70	117										
3	2.60	90		*Regression Statistics*								
4	3.30	105		Multiple R	0.542331							
5	2.20	116		R Square	0.29412291							
6	2.80	107		Adjusted R Square	0.20588828							
7	3.20	105		Standard Error	0.5444891							
8	3.80	121		Observations	10							
9	4.00	113										
10	3.00	90		ANOVA								
11	2.40	87			*df*	*SS*	*MS*	*F*	*Significance F*			
12				Regression	1	0.98825299	0.98825	3.33342	0.105318185			
13				Residual	8	2.37174701	0.29647					
14				Total	9	3.36						
15												
16					*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
17				Intercept	0.2698804	1.559634241	0.17304	0.86692	-3.32664261	3.8664034	-3.3266426	3.86640341
18				IQ	0.02692787	0.014748817	1.82577	0.10532	-0.00708296	0.0609387	-0.007083	0.06093871
19												
20												
21												
22				RESIDUAL OUTPUT								
23												
24				*Observation*	*Predicted GPA*	*Residuals*						
25				1	3.42044171	0.279558295						
26				2	2.6933891	-0.093389097						
27				3	3.09730721	0.202692787						
28				4	3.39351383	-1.193513831						
29				5	3.15116296	-0.351162961						
30				6	3.09730721	0.102692787						
31				7	3.5281532	0.271846797						
32				8	3.31273021	0.687269792						
33				9	2.6933891	0.306610903						
34				10	2.61260547	-0.212605474						

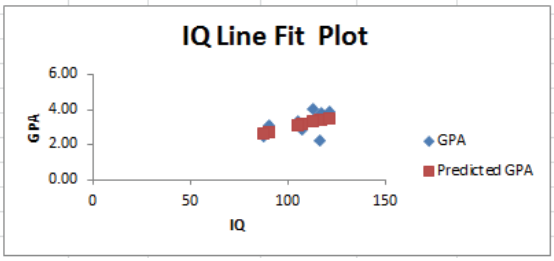


Figure 2.39 Regression result

Note: You can edit the titles (chart, X, Y) of the chart the same way as the examples above.

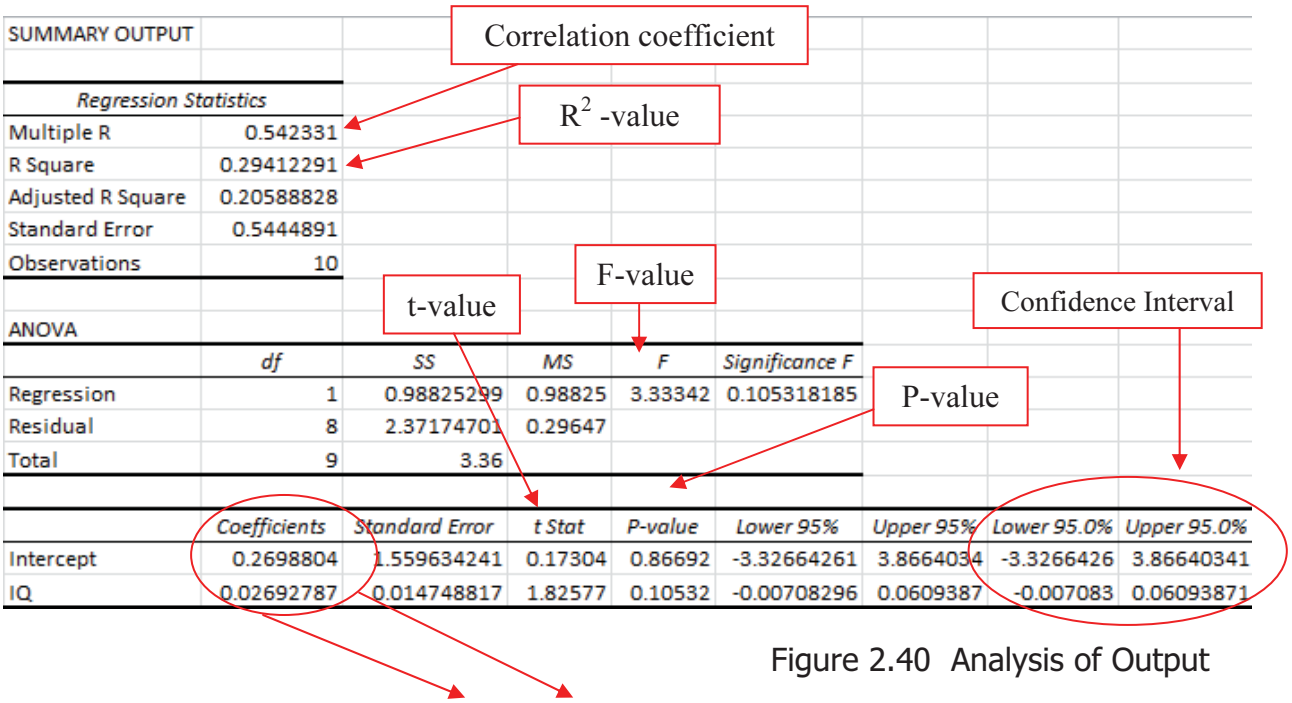
Usefulness:

Regression Equation

1. to predict the value of a dependent variable from the values of one or more independent variables.

2. to determine trends over time of a series of data.

Analysis of Regression Data:



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.542331
R Square	0.29412291
Adjusted R Square	0.20588828
Standard Error	0.5444891
Observations	10

ANOVA

	df	SS	MS	F	Significance F
Regression	1	0.98825299	0.98825	3.33342	0.105318185
Residual	8	2.37174701	0.29647		
Total	9	3.36			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.2698804	1.559634241	0.17304	0.86692	-3.32664261	3.8664034	-3.3266426	3.86640341
IQ	0.02692787	0.014748817	1.82577	0.10532	-0.00708296	0.0609387	-0.007083	0.06093871

Figure 2.40 Analysis of Output

The Regression Equation is $Y' = 0.270 + 0.027X$ (rounded)

When replacing X with independent variable name (IQ) and Y' by dependent variable name (GPA), the equation becomes:

$$GPA = 0.270 + 0.027IQ$$

A simple rule of the thumb on estimates 0.270 and 0.027

Rule of the thumb say if the absolute value of a t-statistics is greater than or equal to 2, then the corresponding parameter estimate is statistically different from zero.

The t-statistics reading from the spreadsheet output for IQ is 1.826 and this value is not greater than or equal to 2. There is a possibility of the parameter estimate for IQ being zero, the equation becomes:

$$GPA = 0.27 + 0IQ$$

Therefore, GPA score in this case is not dependent on IQ.

In regression we do not look at significant level of intercept because it is the only term that stays in the equation if a parameter estimate of regression is not significant.

Or Use of P-value

The P-value from the regression output:

For the IQ, the P value is 0.105

P value indicates the probability or likelihood of obtaining a result at least as extreme as that observed in a study by chance alone, assuming that there is truly no association.

The P value of 0.105 indicates that there is 10.5 percent or about 1 in 9.5 probability of observing the parameter estimate for IQ in the study by chance alone, given that there is really no association.

In research a P-value of less than 0.05 is considered statistically significant and since the test P value is more than 0.05, we are statistically non significant. Therefore, the parameter estimate for IQ slope has a chance of being zero.

Notice that the lower the P-value for an estimate, the more confident you are in the parameter estimate. Note: For more information on P value Appendix VI page 185.

Analysis of Overall Fit of Regression Line

a. Use of R-square (R^2)

R^2 value from the regression output is 0.294,

This means that the regression Equation explains 29% (0.294 convert to percentage) of the total variation in GPA. The closer the R^2 is to 1 (100%), the better is the overall fit, that the independent variable IQ can accurately predict dependent variable GPA.

Adjusted R^2 is an unbiased estimate of R^2 and is used to generalize from the sample to the larger population.

b. Or Use of F-statistics

From the ANOVA table of the regression output:

F-Statistics is 3.333

Significance F is 0.105

The "F" column provides a statistic for testing the hypothesis that the slope of the regression line is not zero against the null hypothesis that it is zero. The test statistic is the ratio MSM/MSE, the mean square model term divided by the mean square error term. When the MSM term is large relative to the MSE term, then the ratio is large and there is evidence against the null hypothesis. For simple linear regression, the statistic MSM/MSE has an *F*

distribution with degrees of freedom (DFM, DFE) = (1, n - 2). Note: Refer to Appendix VI, page 182.

In the ANOVA table the F statistic is equal to 0.98825/0.29647 = 3.3333 (which is exactly the value of regression output). The distribution is $F(1, 8)$, and the probability of observing a value greater than or equal to 3.33 (rounded) is less than 0.001. There is strong evidence that coefficient of IQ is equal to zero.

Confidence Interval

Refer Output Data for $Y' = 0.27 + 0.027X$ (i.e. GPA = 0.27 + 0.027IQ)

Confidence interval for estimate 0.27 is (-3.327, 3.866)

Confidence interval for estimate 0.027 is (-0.007, 0.061)

Both at 95% confidence interval, and given in spreadsheet Figure 2.29.

Regression of More than One Independent Variable (Multiple Regressions)

Many experiments and research do not depend on a single independent variable but many independent variables. The following example will illustrate the use of Excel in more than one independent variable for the purchasing behavior of 10 respondents. The three independent variables used in the analysis are price, advertising and distance and the dependent variable is quantity.

(Note: The analysis for multiple independent variables follows the same approach as for single independent explanatory variable)

Directions:

For regression data, go to: **Data → Data analysis → Regression**

(we use quantity, the dependent variable and its relationship to independent variables; price, advertisement and distance).

Encode the data below in a new worksheet:

Table 2.5 Raw data on multiple inputs

	A	B	C	D	E
1	Observation	quantity	price	advertising	distance
2	1	30	25	11	12
3	2	70	40	24	6
4	3	45	45	15	5
5	4	32	55	31	7
6	5	42	57	34	4
7	6	75	37	22	2
8	7	66	37	12	5
9	8	50	45	24	7
10	9	70	40	22	4
11	10	60	38	10	5

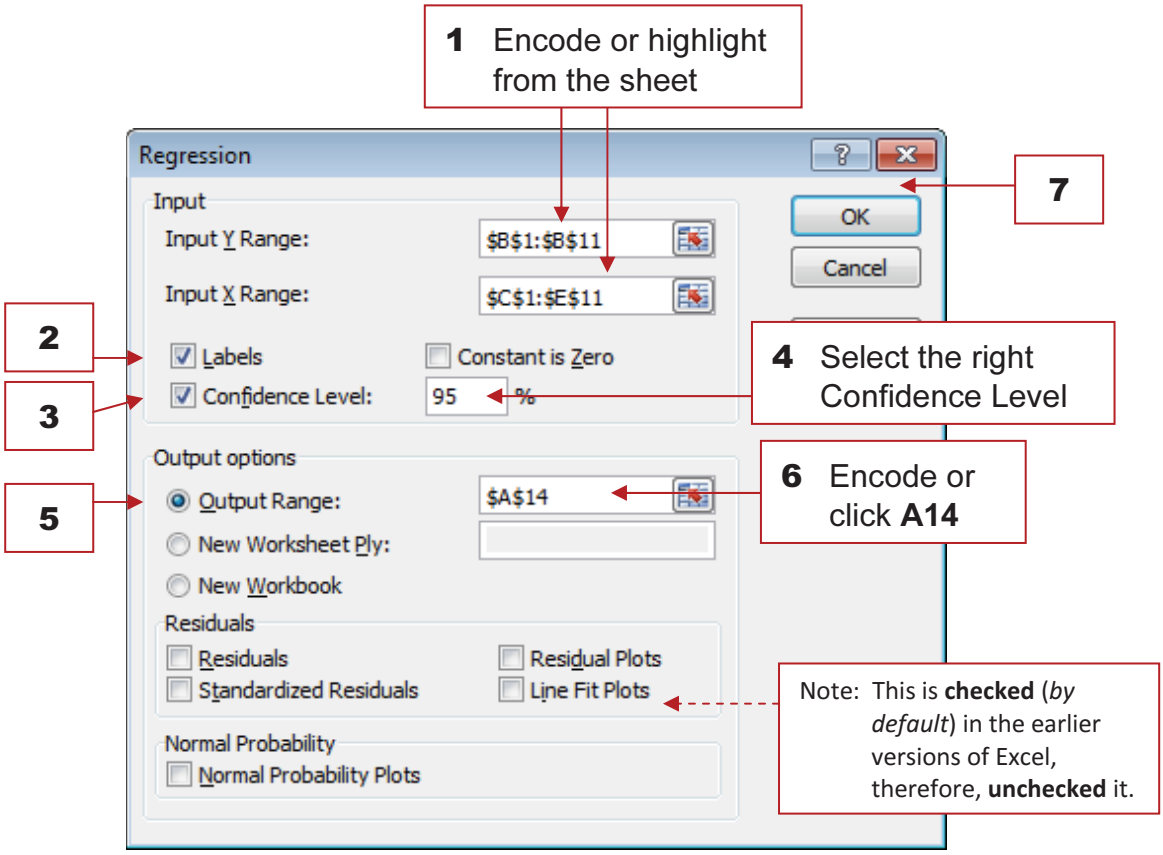


Figure 2.41 Multiple Regression Input and Output Options

The Output gives the following information.

	A	B	C	D	E	F	G	H	I
14	SUMMARY OUTPUT								
15									
16	*Regression Statistics*								
17	Multiple R	0.902844012							
18	R Square	0.81512731							
19	Adjusted R Square	0.722690965							
20	Standard Error	8.660062815							
21	Observations	10							
22									
23	ANOVA								
24		*df*	*SS*	*MS*	*F*	*Significance F*			
25	Regression	3	1984.019872	661.34	8.818256	0.012825159			
26	Residual	6	449.9801277	74.99669					
27	Total	9	2434						
28									
29		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
30	Intercept	138.0184567	19.24229403	7.172661	0.000371	90.93425938	185.102654	90.93425938	185.102654
31	price	-1.483456818	0.551470572	-2.690002	0.036053	-2.832856695	-0.13405694	-2.832856695	-0.13405694
32	advertising	0.537390075	0.58863362	0.912945	0.396472	-0.902944505	1.977724656	-0.902944505	1.977724656
33	distance	-5.768089926	1.176392819	-4.903201	0.002702	-8.646619456	-2.889560396	-8.646619456	-2.889560396

Figure 2.42 Multiple regression Output

Analysis of Regression Data:

- *Equation*: Quantity = 138.02 - 1.48 Price + 0.54 Advertisement - 5.77 Distance

Not significant, P-value is greater than 0.05

Use of t-statistics

If the absolute value of a t-statistics is greater than or equal to 2, then the corresponding parameter estimate is statistically different from zero. (*As before, simple rule of the thumb applies here).*

For price, t-value is |-2.69| is ≥2,

For advertisement, t-value is |0.91| is <2

For distance, t-value is |-4.90| is ≥2,

When the modulus of the t-value is more than or equal to 2, we retain the factor in the equation.

If it is a prediction, non significant factors should be dropped from the prediction equation. In some cases where the goal is to know the weight of each factor in the model then we keep non significant factors.

Therefore, being a predictive model, we retain the factors price and distance in the equation and drop the factor advertisement from the equation.

The analysis of constant or intercept in this model is the same as for single independent variable. (Note: Refer to page 87)

Or Use of P-value

The P-value from the regression output:

For Price, P-value = 0.036053

For Advertisement, P-value = 0.396472

For Distance, P-value = 0.002702

The P value of 0.036053 for price indicates that there is 3.6 percent or about 1 in 27.74 probability of observing the parameter estimate for price being zero. For distance it is 0.2 percent and for advertisement it is 39.64 percent. So you are more confident on your estimate when the percentage is smaller.

Note: For more information on P value and significance refer to Appendix VI page 185.

Analysis of Overall Fit of Regression Line

a. Use of R-square

R-square value from the regression output is 0.815

The ratio SSM/SST is R^2, also known as the *squared multiple correlation coefficient*. This value is the proportion of the variation in the response variable that is explained by the response variables. The square root of R^2 is called the *multiple correlation coefficient.*

The squared multiple correlation is;

R^2 = SSM/SST = 1984.0198/2434 = 0.815, (which is exactly the value of the regression output) indicating that 81.5% of the variability in the "Purchasing" variable is explained by the independent variables of price, distance and advertisement.

b. Use of F-statistics

The output table gives a significance F = 0.0128.

This means there is 1.28% (change 0.0128 to %) chance that the estimated regression model fits the data purely by accident and not through your experimental data.

In general, F-statistics with significance values of 5% or less are generally a considered significant and a good fit. In our case here, the value of 0.0128 is less than 5% and we say that the graph overall is a good fit. This is also supported by a small F-value of 8.818.

2.2.3 Further Exploration: Non-Linear Relationship

If you suspect your research data is non-linear, you can follow the guidelines given in the illustration below to analyze the relationship.

An example of a set of non-linear data is shown below, first plot the scatter plot and check for the type of non- linearity.

Table 2.6 Hypothetical Experimental Data

X	1.2	2.4	2.8	3.5	3.9	4.3
Y	9.13	3.23	2.56	1.83	1.56	1.35

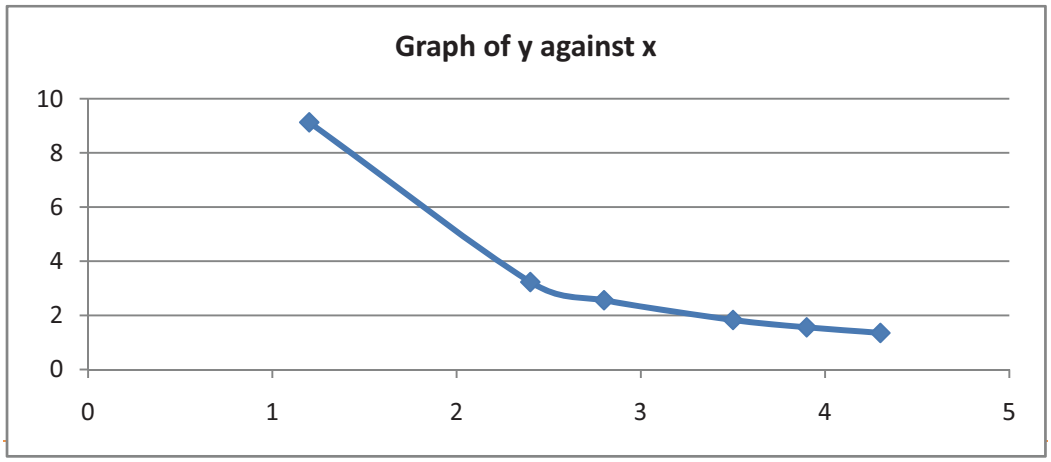


Figure 2.43 Scatter plot for the data

The above graph shows a non-linear relationship between x and y. We will try to fit a non-liner equation to the data, say, in this case we take the relationship as: $\boldsymbol{yx^a = b}$

We need to calculate the unknown constants **a** and **b** before the equation can be used for predicting purpose. First, we need to convert the equation using the concept of Log (base 10) as shown below to make it into linear form:-

$$yx^a = b$$

$$log_{10} yx^a = log_{10} b$$

$$log_{10} y + log_{10} x^a = log_{10} b$$

$$log_{10} y = -alog_{10} x + log_{10} b$$

(this is an equivalent linear form: Y = a + bX)

Next find the log values for x and y as shown below:-

Table 2.7 Log value of Data

X	1.2	2.4	2.8	3.5	3.9	4.3
Y	9.13	3.23	2.56	1.83	1.56	1.35
$\boldsymbol{log_{10} x}$	0.079	0.380	0.447	0.544	0.591	0.633
$\boldsymbol{log_{10} y}$	0.96	0.51	0.41	0.26	0.19	0.13

Lastly, show Graph of Relationship between log x and log y

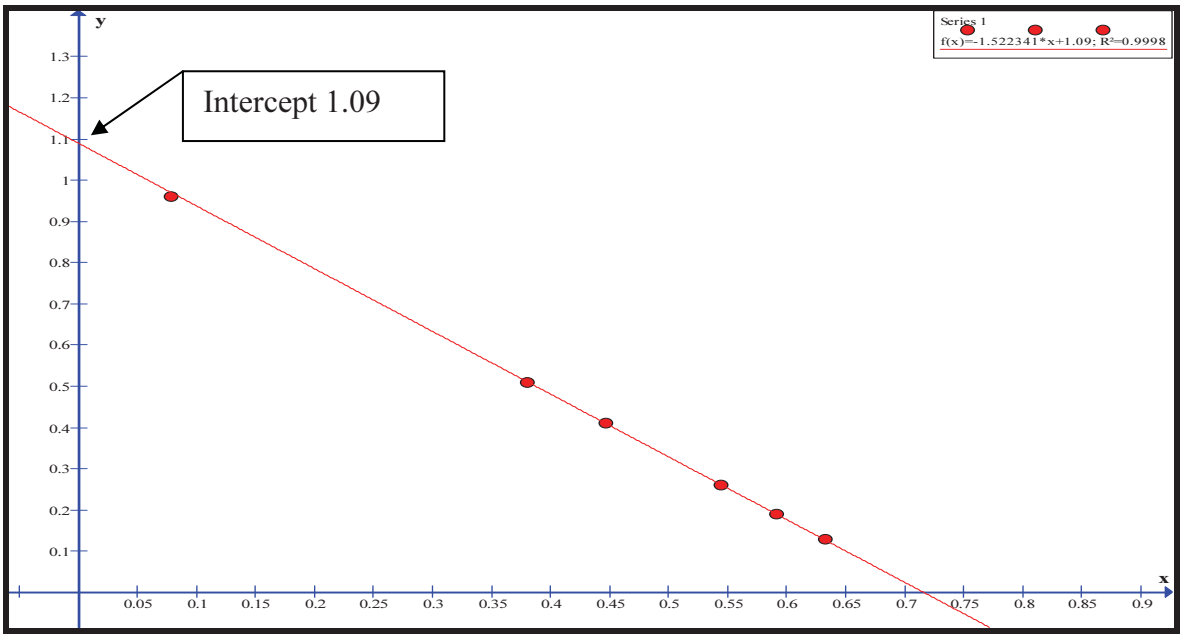


Figure 2.44 Linear relationship of log x and log y

(i) To Find a and b Graphically:

From the drawn graph: Intercept = 1.09 and Gradient = - 1.09/0.72 = - 1.51

Gradient: -1.51 = - a

Therefore: a = 1.51

Intercept: log b = 1.09

Therefore: b = 12.30

The non linear regression equation is: $yx^{1.51} = 12.3$

(ii) To Find a and b Statistically:

Alternatively, you can use the linear regression in Excel to find the values of y-intercept and the gradient followed by the calculation shown for graphical method above. Refer to analysis of data for single independent variable discussed earlier on how to get the gradient and y-intercept from regression output of Excel.

Remark: similar techniques can be applied for other non-linear relationships. You need to convert it into linear form. First, draw the scatter plot and check for non-linearity. If it is non-linear, linearize it. Refer to any mathematics books that have a chapter on techniques of linearization.

2.3 Hypothesis Testing

It is common among researchers of inferential statistics to test Hypothesis. There are many situations appropriate for a statement to be test in any research. It may be that they want to compare some attributes of a sample mean to a population mean for any number of reasons. There are many different hypothesis testing procedures – some involving a single sample mean, some based on two sample means, and still others that deal with three or more sample means. It is fairly straightforward to move from one type to the other. Our illustrative example focuses on tests concerning the difference between *two population m*eans or two *population proportions.* The procedure can be extended to other forms based on the needs of the researcher (Note: Refer to Appendix VI, page 185 and section 3.5 page 159 for the type of test and the situation it is applicable)

Directions:

1. Encode the following data in a worksheet: *(this is an example of small sample with paired value).*

2. Go to: **Data → Data analysis → t-test: Paired Two Sample for Means**

Table 2.8 Raw data on learning types

	A	B
1	Scores of Classroom-Based and Online Based students	
2		
3	Classroom Learning	On-line Learning
4	78	80
5	72	71
6	76	75
7	80	82
8	65	65
9	75	70
10	83	85
11	82	82
12	77	75

3. Fill-out the Input and Output Options as shown below:

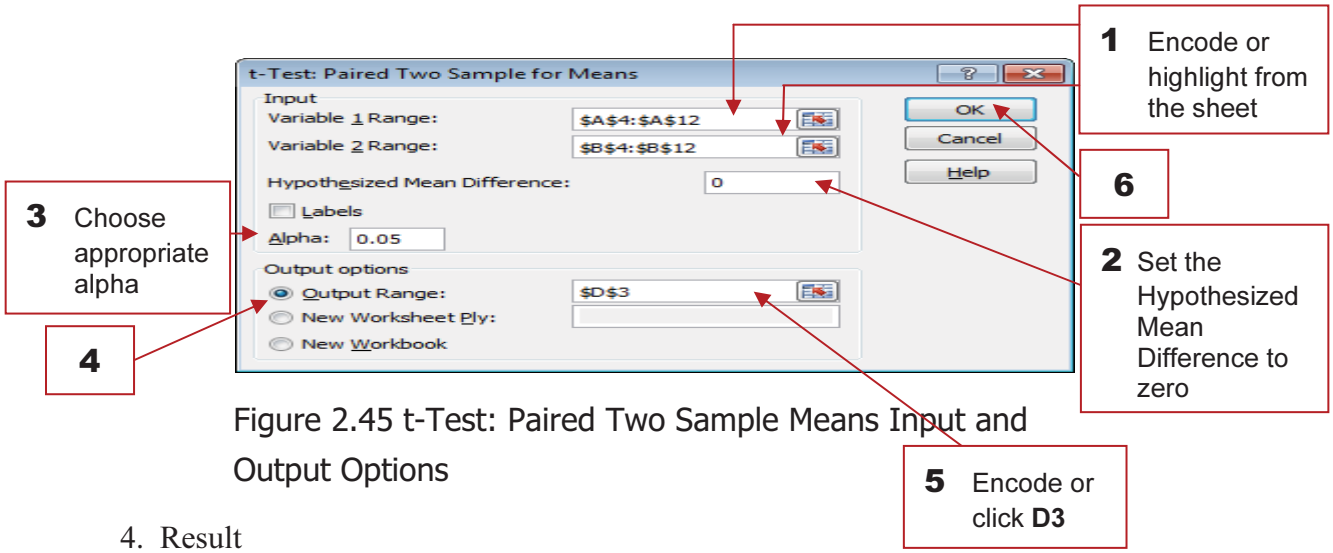


Figure 2.45 t-Test: Paired Two Sample Means Input and Output Options

4. Result

	A	B	C	D	E	F
1	**Scores of Classroom-Based and On-line Based students**					
2						
3	Classroom Learning	On-line Learning		t-Test: Paired Two Sample for Means		
4	78	80				
5	72	71			*Variable 1*	*Variable 2*
6	76	75		Mean	76.44444444	76.11111111
7	80	82		Variance	30.27777778	44.11111111
8	65	65		Observations	9	9
9	75	70		Pearson Correlation	0.945924345	
10	83	85		Hypothesized Mean Difference	0	
11	82	82		df	8	
12	77	75		t Stat	0.43643578	
13				P(T<=t) one-tail	0.337029824	
14				t Critical one-tail	1.859548038	
15				P(T<=t) two-tail	0.674059648	
16				t Critical two-tail	2.306004135	

Figure 2.46 Output for t-Test

Analysis of Output Data:

Reject Null Hypothesis if P-value is:

Less than 0.05 at 5% significance level. Or Less than 0.01 at 1% significance level.

For our output data:

Null Hypothesis H_0: $\mu_0 = \mu_1$

Alternate Hypothesis H_1: $\mu_0 \neq \mu_1$

Based on the P-value of 0.6740 *(we selected 2-tail)*, there is insufficient evidence to reject H_0 at the 5% significance level.

The P-value of the output data that is relevant to your analysis is based on your Alternative Hypothesis (whether it is one-tailed or two- tailed hypothesis).

For two sample t-test in Excel:

Data → Data Analysis → t-test: Two sample assuming equal variances

For *large-sample* test for two means: If sample size is both large ($n \geq 30$) we do not need to assume equal standard deviation.

Data → Data Analysis → t-test: Two sample assuming unequal variances

(Excel uses an approximation of t-distribution here rather than Normal Distribution).

Excel does not have an automatic procedure to carry out a *one-sample* t-test. We can obtain the result with some modification and using paired t-test.

(When you have only one sample, create another set of sample by putting in thevalue of the mean you are going to test as your entry for the second sample and the rest is the same as earlier procedure)

2.3.1 Understanding ANOVA and F-Ratio

An extension of differences of two-sample mean test to more than two samples is ANOVA (Analysis of Variance). We will focus on one-way analysis of variance as we are more interested in the relationship between one variable with another variable. ANOVA allows us to compare multiple samples in a single application. The procedure involves a comparison of different estimates of population variance.

ANOVA is tested through the use of F-Ratio. The F-Ratio compares the estimate of between-groups variance to estimate of within-groups variance. If the F-ratio calculated meets or exceeds the critical F-value, we will reject the null-hypothesis that there are no differences between the means.

The F-ratio is defined as:-

$$F_{\text{ratio}} = \frac{Estimate\ of\ between{-}groups\ variance}{Estimate\ of\ within{-}groups\ variance}$$

Figure 2.47a shows the distribution of three groups A, B and C for some attributes. The means are fairly close together, and there are substantial variations in the means within each

group. These factors cause the distributions to overlap. Figure 2.47b shows the means are well placed with no overlaps. It is the comparison between the groups' variation to within the groups' variation that tells us whether an assertion on an attribute is accepted or rejected.

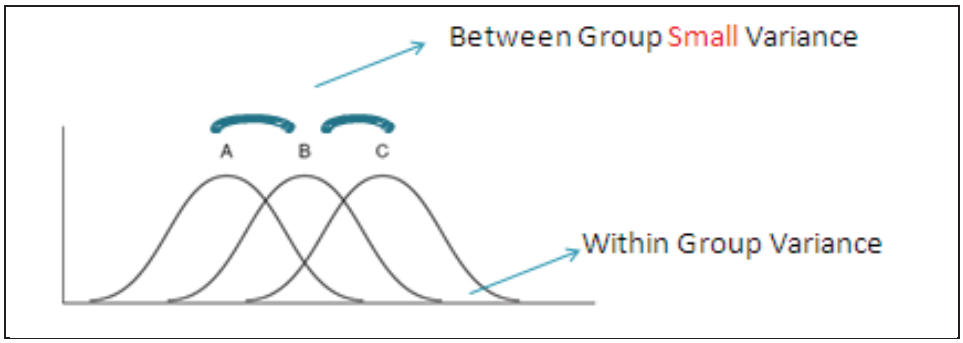


Figure 2.47a Small F-ratio

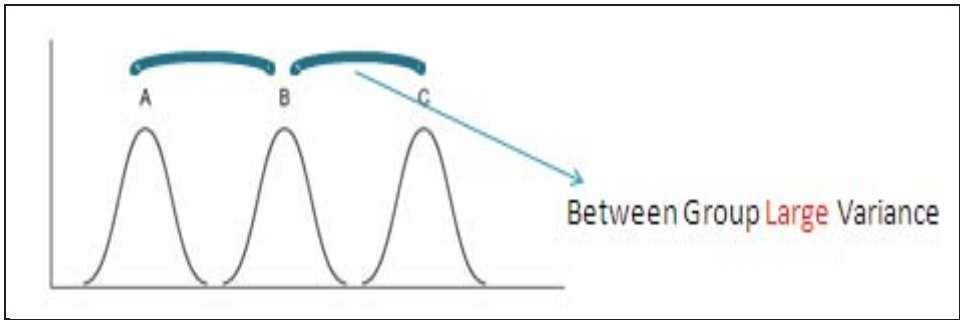


Figure 2.47b Large F-ratio

Illustration

You are given the time (in minutes) to complete a questionnaire by 4 different groups of students

You are to investigate whether there are any differences in the times to complete the questionnaire between the four groups.

Table 2.9

Group A	Group B	Group C	Group D
4	4	9	5
7	7	5	1
10	4	13	8
8	8	6	9
5	8	10	4
12	6	6	9
6	4	10	7
8	7	9	
9	5	12	
10	5	9	

Question Analysis:-

	Group A	Group B	Group C	Group D
Mean	7.9	5.8	8.9	6.1
Sample size	10	10	10	7
Overall mean (Grand mean)	7.2			

Hypothesis Tested: No difference between the groups' mean time to complete questionnaire.

Null Hypothesis, H_O: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

Within group Sum of Squares (SS_W)

Group A: $(4-7.9)^2 + (7-7.9)^2 + \ldots + (10-7.9)^2 = 54.90$

Group B: $(4-5.8)^2 + (7-5.8)^2 + \ldots + (5-5.8)^2 = 23.60$

Group C: $(9-8.9)^2 + (5-8.9)^2 + \ldots + (9-8.9)^2 = 60.90$

Group D: $(5-6.1)^2 + (1-6.1)^2 + \ldots + (7-6.1)^2 = 52.87$

SS_W for all groups = 54.90+23.60+60.90+52.87 = 91.97

Mean Square Within (MS_W) = $\frac{within-groups\ sum\ of\ squares\ (SSw)}{within-groups\ degrees\ of\ freedom\ (dfw)} = \frac{191.97}{33} = 5.82$

Between groups Sum of Squares (SS_B)

Group A: $10(7.9-7.2)^2 = 4.90$

Group B: $10(5.8-7.2)^2 = 19.60$

Group C: $10(8.9-7.2)^2 = 28.90$

Group D: $7(6.1-7.2)^2 = 8.47$

SS_B for all groups = 4.90+19.60+28.90+8.47 = 61.87

Mean Square Between (MS_B) = $\frac{between-groups\ sum\ of\ squares\ (SSB)}{between-groups\ degrees\ of\ freedom\ (dfB)} = \frac{61.87}{3} =$ 20.62

Calculation of F-Ratio

F-ratio = $\frac{MSB}{MSW} = \frac{20.62}{5.82} = 3.54$

Interpretation of the F-Ratio

From Table of F-distribution, Critical value = 2.92 (significant level 0.05, df between groups is 3, df within group is 33)

Test value is 3.54 > Critical value 2.92

Therefore, we Reject the Null Hypothesis at 0.05 level of significance

2.3.2 Post Hoc Test: Q-Test (Tukey's HSD)

When the null hypothesis is rejected, you may want to know which combination (groups) contributes significantly to variance. This is where the post hoc test is handy, the Q-test can be used to identify the variance between pair of the various combination.

Illustration

Continuing with the earlier illustration, the following pairs (or any other pairs) can be analyzed:-

Mean of group A and mean of sample B

Mean of group A and mean of sample C

Mean of group A and mean of sample D

Mean of group B and mean of sample C

Mean of group B and mean of sample D

Mean of group C and mean of sample D

The Calculation of Q-Statistic

When all sample sizes are equal use:

$$Q = \frac{|\overline{X1} - \overline{X2}|}{\sqrt{(MSw/n}}$$ where n is the mean sample size

When all sample sizes are unequal use:

$$Q = \frac{|\overline{X1} - \overline{X2}|}{\sqrt{MSw/\tilde{n}}}$$ where $\tilde{n}$ is the harmonic mean

Given as: $\tilde{n} = \frac{k}{\frac{1}{n1}+\frac{1}{n2}+\frac{1}{n3}+\frac{1}{n4}}$

We use Harmonic mean as our sample is of unequal size.

$$\tilde{n} = \frac{4}{\frac{1}{10}+\frac{1}{10}+\frac{1}{10}+\frac{1}{7}} = 9.03$$

The denominator of Q-Statistic is: $\sqrt{MSw/\tilde{n}} = \sqrt{\frac{5.82}{9.03}} = 0.80$

The Final Result:

Table 2.10

Possible Comparisons	$Q = \frac{\lvert\overline{X1} - \overline{X2}\rvert}{\sqrt{MSw/\tilde{n}}}$	Critical value for Q = 3.85 at 0.05 level of significance (From Q-Table)
Group A and Group B	|7.9 – 5.8| / 0.80 = 2.63	Not Significant
Group A and Group C	|7.9 – 8.9| /0.80 = 1.25	Not Significant
Group A and Group D	|7.9 – 6.1| /0.80 = 2.25	Not Significant
Group B and Group C	|5.8 – 8.9| / 0.80 = 3.88	Significant
Group B and Group D	|5.8 – 6.1| /0.80 = 0.38	Not Significant
Group C and Group D	|8.9 – 6.1| /0.80 = 3.5	Not Significant

From the Q-test, we know that Group B and Group C contributed significantly to between group variance and the cause of rejection of F-test in the first place.

Note: df is degree of freedom (For within groups: 37 – 4 = 33, For between group: 4 – 1 = 3

Usefulness of t-test:

1. to test whether two groups (categories) are different.
2. to test whether three or more groups (categories) are different- use of ANOVA.

2.3.3 Using Excel for ANOVA

Directions:

1. Use the same data as in **Table 2.9** and copy it into a new excel sheet
2. Go to: **Data → Data Analysis → Anova: Single Factor**

   Press OK
3. Copy the data and paste it in Input Range.

   Tick the column (as your data is arranged in column),

   Tick Alpha (select appropriate value),

   Tick Labels in First Row (if your input data range contains group names)

   Press OK

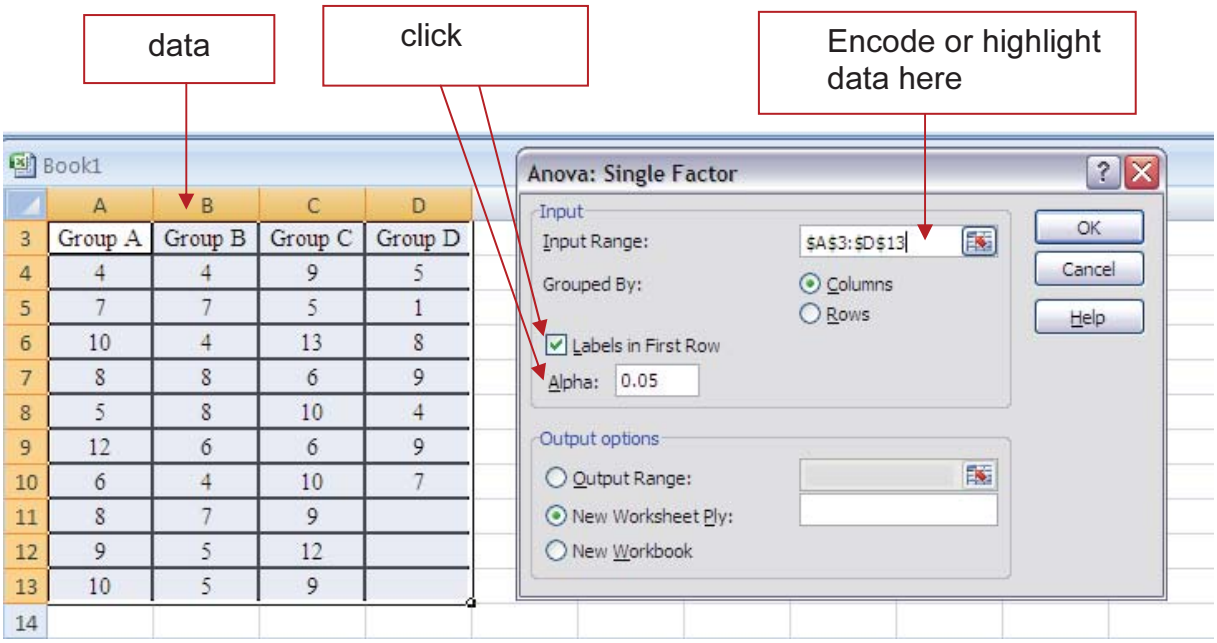
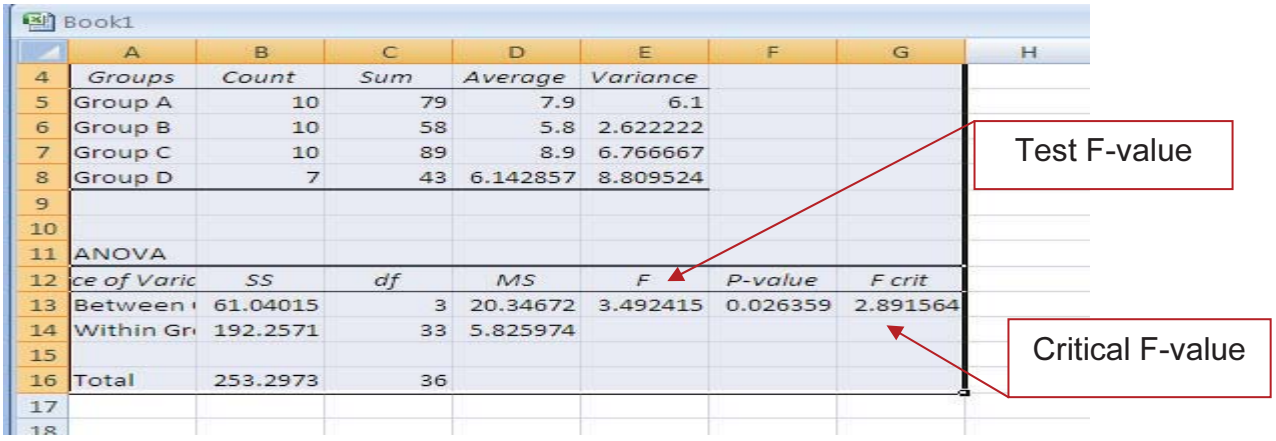


Figure 2.48 Anova: Input for Single Factor

4. The Final Output



Groups	Count	Sum	Average	Variance
Group A	10	79	7.9	6.1
Group B	10	58	5.8	2.622222
Group C	10	89	8.9	6.766667
Group D	7	43	6.142857	8.809524

ANOVA

ce of Varic	SS	df	MS	F	P-value	F crit
Between	61.04015	3	20.34672	3.492415	0.026359	2.891564
Within Gr	192.2571	33	5.825974			
Total	253.2973	36				

Figure 2.49 Anova: Output

Note 1: *the values are slightly different with calculation method due to rounding errors.*

Note 2: *If your relationship is between one variable with two variables, you can use the two-way analysis of variance, refer to statistical book for details.*

2.4 Chi-square Test

The chi-square test of independence is a hypothesis-testing procedure appropriate for categorical variables. It tests whether or not there is an association between two variables in some ways.

The *null hypothesis* (H_0) always starts with *no difference between* the variables and the *alternate hypothesis* (H_1) always starts with significant *difference* between the variables.

As a hypothesis-testing tool, we need to rely on the use of contingency table. A contingency table is a table that presents data (observed frequency) in terms of all combinations of two or more variables. For the calculation of chi-square test statistics, the expected frequencies need to be calculated.

This can be done in two ways depending on *observed frequency* or *probability* is given.

a. When observed frequency is given, the expected frequency for each Cell is given as:

$$\mathbf{E = (R \times C) / n}$$

Where R = row total, C = column total, E= expected frequency and n = sample size.

b. When observed probability is given, the Expected Frequency for each Cell is given as:

$$\mathbf{E = n \times p}$$

Where p = probability and n = sample size.

The Chi-square Statistics is defined as:

$$\chi^2 = \sum \{(0-E)^2 / E\}$$

Where O = observed frequency, E = expected frequency

Illustration: Computation of Expected Value for a 3 x 3 Contingency Table

First create a table with Age as category 1 (in this example we use 3 age groups) and Opinion as category 2 (in this example we use 3 satisfaction groups). Then fill the cells in the table with observed frequencies; a, b, c, d, e, f, g, h, and i.

	Satisfied	Not Satisfied	No Opinion	Total
Under 30	a	b	c	a+b+c
30-50	d	e	f	d+e+f
Over 50	g	h	i	g+h+i
Total	a+d+g	b+e+h	c+f+i	a+b+c+d+e+f +g+h+i

Figure 2.50a Contingency table of observed frequency

This is how the table will look with a set of sample frequencies. The bracketed numbers indicates expected frequencies whose calculation is shown below.

	Satisfied	Not Satisfied	No Opinion	Total
Under 30	226 (197.63)	182	123	531
30-50	117 (133.62)	153	89	359
Over 50	88	112	68	268
Total	431	447	280	1158

Figure 2.50b Contingency table of observed and expected frequency

Computation of the expected value for cell one.

$$\text{Value} = \frac{(row\ total\ for\ the\ cell) \times (column\ total\ for\ the\ cell)}{Grand\ total}$$

$$= (531 \times 431) / 1158$$

$$= 197.63$$

Repeat the steps for all the cells and type them as Expected Data in your Excel sheet

Statistical Analysis Using Excel Software

The use of software (Excel) helps us in determining the value of the Chi-square test statistic without the need for elaborate calculations. The following illustration will help you with the software and its interpretation.

Illustration

You are given the observed frequency of two different age groups and three satisfaction levels in using a product purchased below.

Table 2.11

	Satisfied	Not Satisfied	No Opinion
Under 30 years	10	30	30
Equal to or more than 30 years	0	10	10

You want to test whether the age groups (a category made up of different age intervals) is associated with satisfaction on product purchased (a category made up of different levels of satisfaction).

The following hypothesis can be tested.

H_0: there is no association between age groups and satisfaction on product purchased

H_1: there is association between age groups and satisfaction on product purchased

Directions:

1. Data: You need to calculate the Expected Frequency separately.

   Excel is not able to work out the Expected Frequency and the Chi-square value.

   (See table below - **Guide to Computation of Expected Value**)

   Encode this data:

Table 2.12 Frequency data

	A	B	C
1	Observed Frequency		
2	10	30	30
3	0	10	20
4			
5	Expected Frequency		
6	7	28	35
7	3	12	15

2. Enter the formula below:

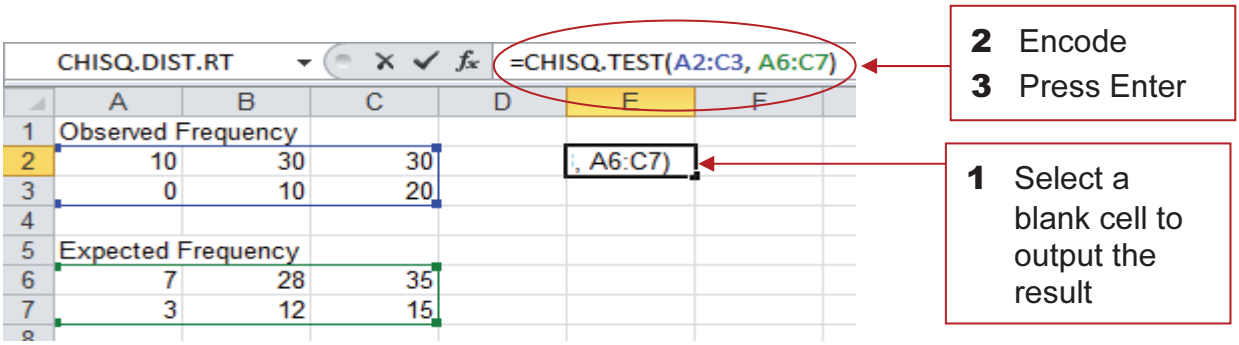


Figure 2.51a Encoding Chi-square data

For MS Excel version 2007 and below, use ***CHITEST***

=CHITEST(A2:C3, A6:C7)

3. Result

P-value = 0.028

	A	B	C	D	E	F
1	Observed Frequency					
2	10	30	30		0.028116	
3	0	10	20			
4						
5	Expected Frequency					
6	7	28	35			
7	3	12	15			

Figure 2.51b Chi-square output

Interpretation:

If P-value is < 0.05, the Null Hypothesis is rejected at 5% significant level.

For m the Chi-square output: 0.028< 0.05, we reject the null Hypothesis at 5% level

Therefore, we reject the null hypothesis H_0: there is no association between age groups and the satisfaction on product purchased at the 5% significant level.

That is we accept the alternative hypothesis H_1: there is association between age groups and the satisfaction on product purchased

Note: Be cautioned in using Chi-square distribution when the number of expected frequency is less than 5. The smaller the expected frequency, the less valid the chi-square distribution becomes. One solution to this problem is to combine two rows into one. *(Refer to statistical*

book, like Brase and Brase, Understanding Basic Statistics, 3rd edition, Chapter 12, Part 1 that deals with this problem or any other statistical books for more information).

A Guide to Finding Critical value and P-value for Chi-square

If you are keen in finding critical *value* for Chi-square distribution, you can follow the instruction given below.

Formulas → Insert Function → Select a Category (Statistical) → CHISQ.INV.RT

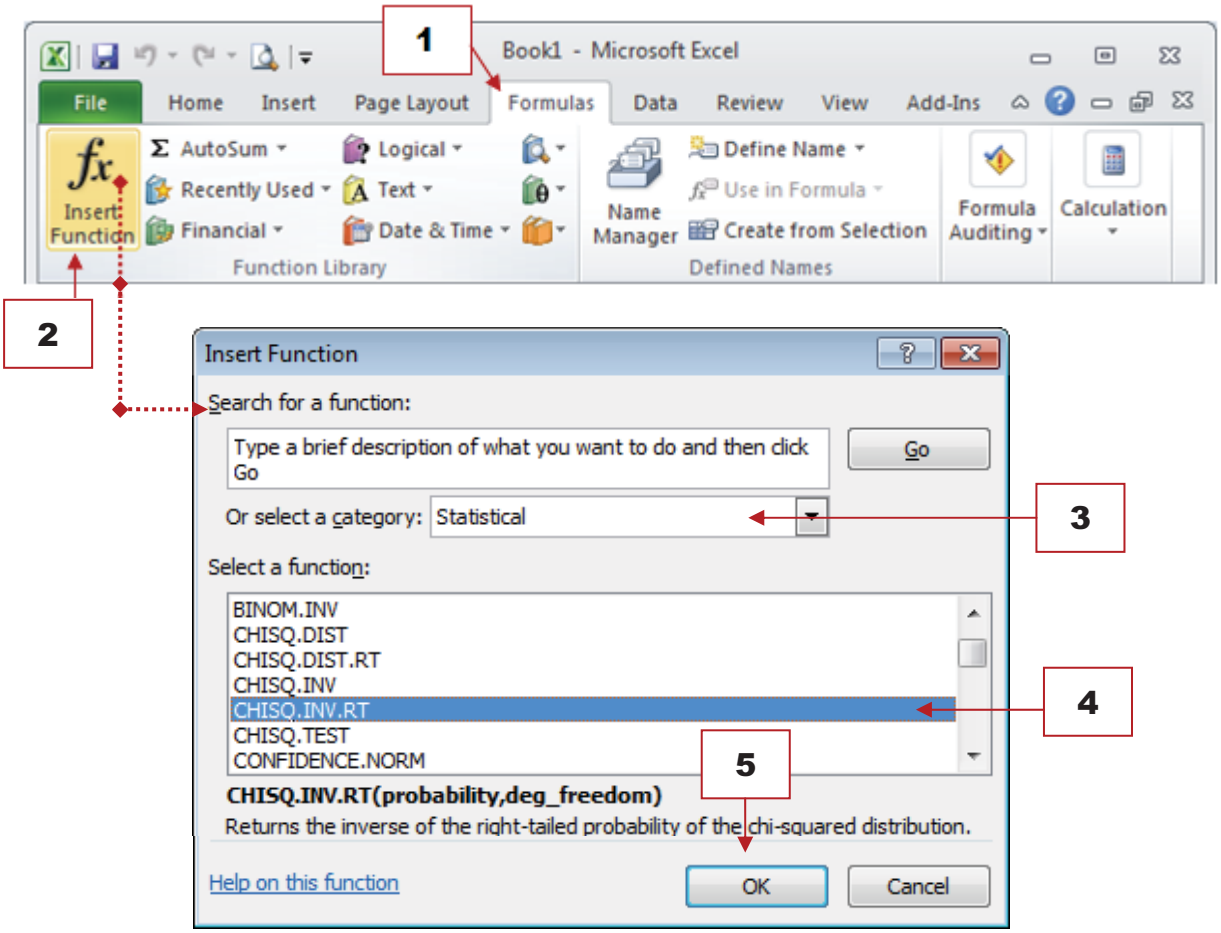


Figure 2.52 Insert Functions

Or, simply encode the formula at the formula bar:

For MS Excel version 2007 and below, use *CHIINV*

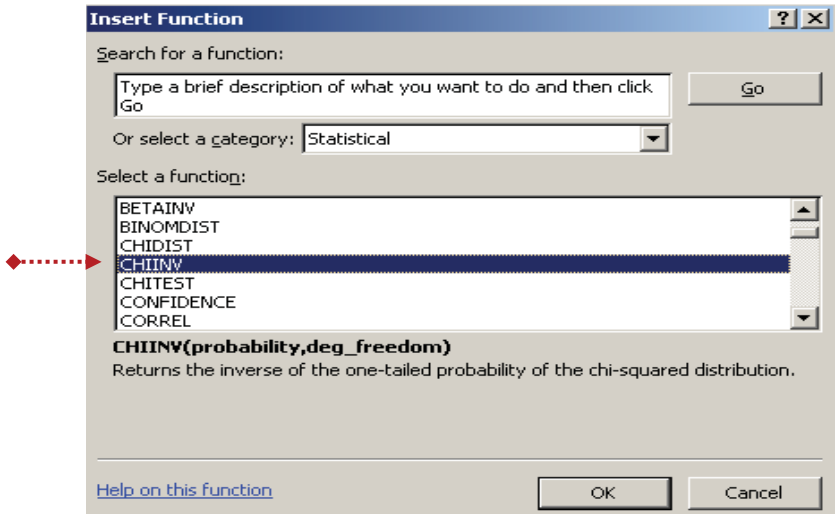
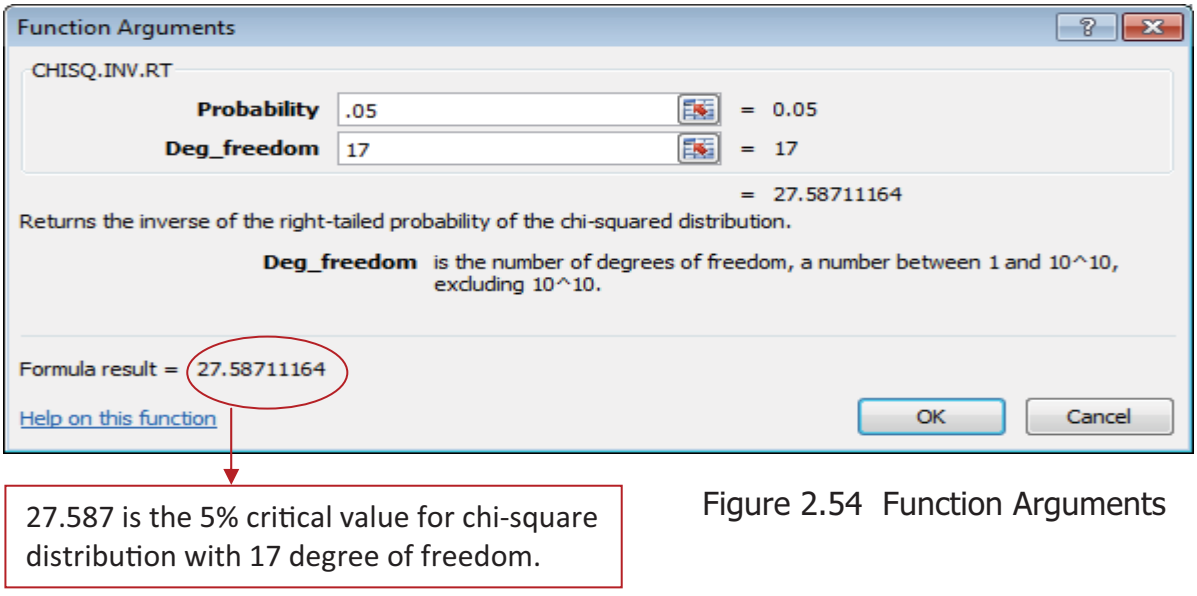


Figure 2.53 Insert functions

Select your critical value e.g. like 0.05 (that is 5%) or 0.01 (1%) and the degree of freedom given by (total rows – 1) x (total columns – 1) from the created data table.



27.587 is the 5% critical value for chi-square distribution with 17 degree of freedom.

Figure 2.54 Function Arguments

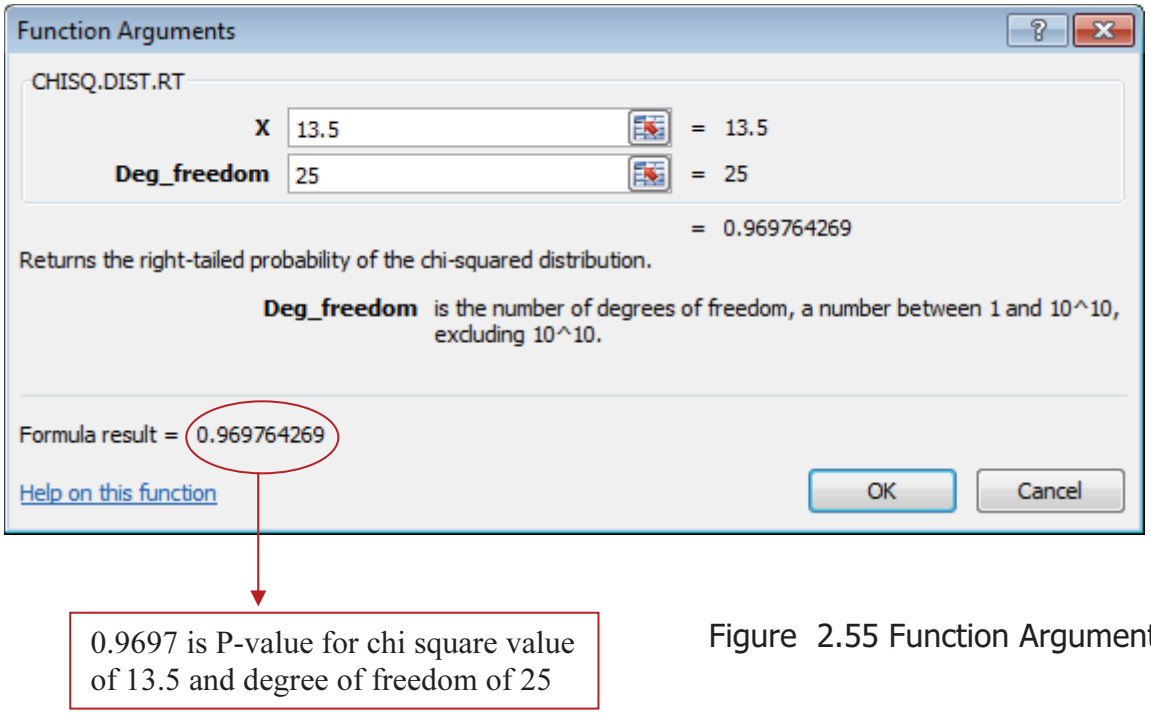
If you are keen in finding *P-value* for a computed Chi-square value, follow the given instructions:

Formulas → Insert Function → Select a Category (Statistical) → CHISQ.INV.RT

For MS Excel version 2007 and below, use ***CHIDIST***

Formulas → Insert Function → Select a Category (Statistical) → CHIDIST

Type in the Chi-square value and the degree of freedom:



0.9697 is P-value for chi square value of 13.5 and degree of freedom of 25

Figure 2.55 Function Arguments

Usefulness of Chi-Square Test:

1. Allows you to discover whether there is significant association between two variables.

   - it uses the observed frequency and the expected frequency.

   - remember that chi square indicates whether a statistically significant relationship exists but does not reveal information about the strength of that relationship.

2. Goodness of Fit

   When working with qualitative data, researchers may use a chi-square test to compare the results of observed date with data they expect to obtain according to specific hypothesis. Chi-square test examines the null hypothesis or the lack of correlation between expected and observed values.

   For example, when working with a market situation, you expected 50 percent of market share as your company's share. If the actual observed number of shares is other than that, however, you may want to determine the "goodness to fit" between the expected and observed result

3. For Likert Scale Response

   Chi square statistics can also be used to analyze Likert scale data. A statistical significance of a given hypothesis can be made between respondents' actual responses with expected answers. The greater the level of deviation between actual and expected responses, the higher the chi square statistic and, thus, the less well the results fit the hypothesis.

   You need to combine the response categories in your Likert scale into some categories before you use Chi square test.

For Example,

Category 1: agree (combine the agree and strongly agree responses)

Category 2: neither (neither agrees nor disagree)

Category 3: disagree (disagree and strongly disagree into another)

This gives you three categories of responses: agree, disagree, and neither where Chi square test can be used.

3 *ORGANIZING DATA Using SPSS*

We shall approach the use of SPSS through an example. The model discussed below can serve as a practical example and it ***can be duplicated*** or taken as a guide for other similar *Quantitative Research*.

We explain the model before embarking on *Data input, Data Output and Data Interpretation* using SPSS.

3.1 The MODEL

The Model contains three sections, namely;

Dependent Variable

(what you are investigating, i.e. *your Research Question/s*)

Independent Variables

(the variables that you think influence the dependent variable, i.e. the *Research Question is divided into a number of Objectives and each objective is one of your independent variables*)

Moderating Variable

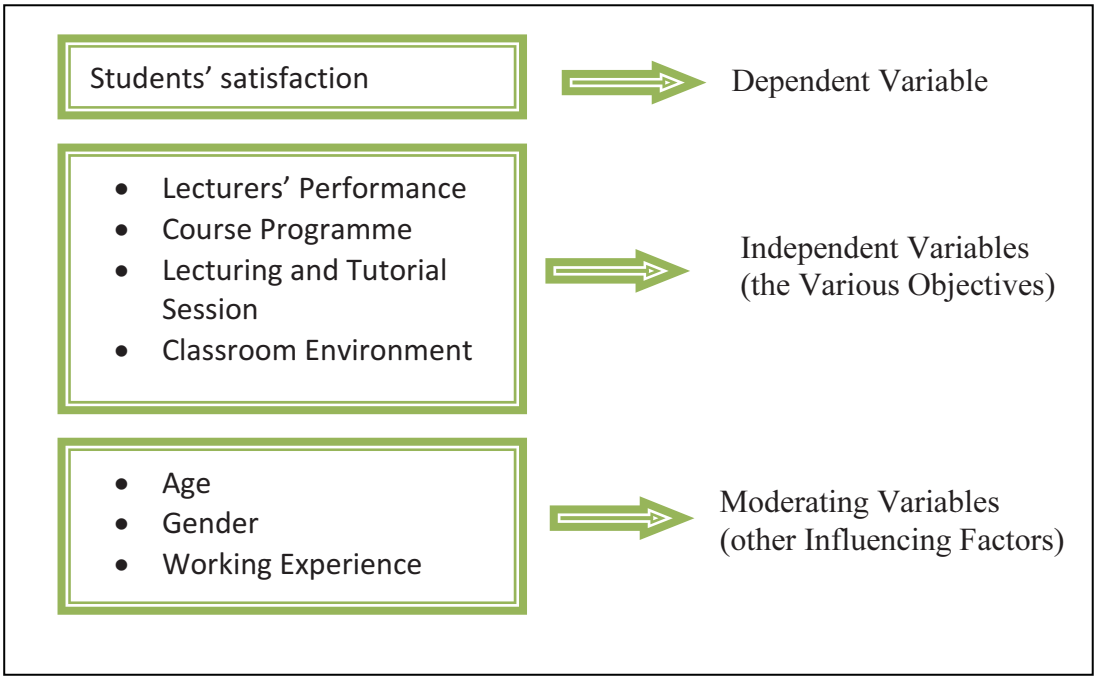
(i.e. *other variables* that have influence on research).

As an illustration, we will look at Students' Satisfaction (E.g. Research Question: *Are students satisfied with YIC College?*) in a college, the factors that influence the students' satisfaction and other factors that may have an influence on students' satisfaction.

Students' Satisfaction in a college

- Items that directly influence the Students' Satisfaction (the various objectives for your Research Question that can help you to answer the research question) can be:
    - *Lecturers' Performance*
    - *Course Programme*
    - *Lecturing and Tutorial Session*
    - *Classroom Environment*
    - (and any other factors that you think has an influence on students' satisfaction).
- Items that indirectly influence the Students' Satisfaction can be:
    - *Age*
    - *Gender*
    - *Working experience*
    - *Classroom Environment*
    - (and any other factors that you think has an indirect influence on students' satisfaction).

Figure 3.1 Types of variables



For each *Independent Variable*, generate sufficient questions to get the required response in that sub group.

- As an illustration, the questions to induce some response on *Lecturers' Performance* sub group can be as follows:

Lecturers' Performance

*The questions on lecturers' performance can be as follows (or any other questions you think are suitable with the use of 5-point **Likert** type scale):*

{Please tick the appropriate response based on the scale given below}

Strongly Disagree 1	Disagree 2	Neither Agree Nor Disagree 3	Agee 4	Strongly Agree 5

	1	2	3	4	5
My lecturers always take concern about my academic advancement.					
I found my lecturers helpful when I encounter difficulties in my subject area.					
My lecturers mark assignment work fairly and returned it after a reasonable period of days.					

- *Similarly*, device questions for the other Independent Variable *Course Programme.*
- *Similarly*, device questions for the other Independent Variable *Lecturing and Tutorial Session.*
- *Similarly*, device questions for the other Independent Variable *Classroom Environment.*

Note: Refer to Appendix IV, page 171 for the complete questionnaire.

The Flow Chart for the Model

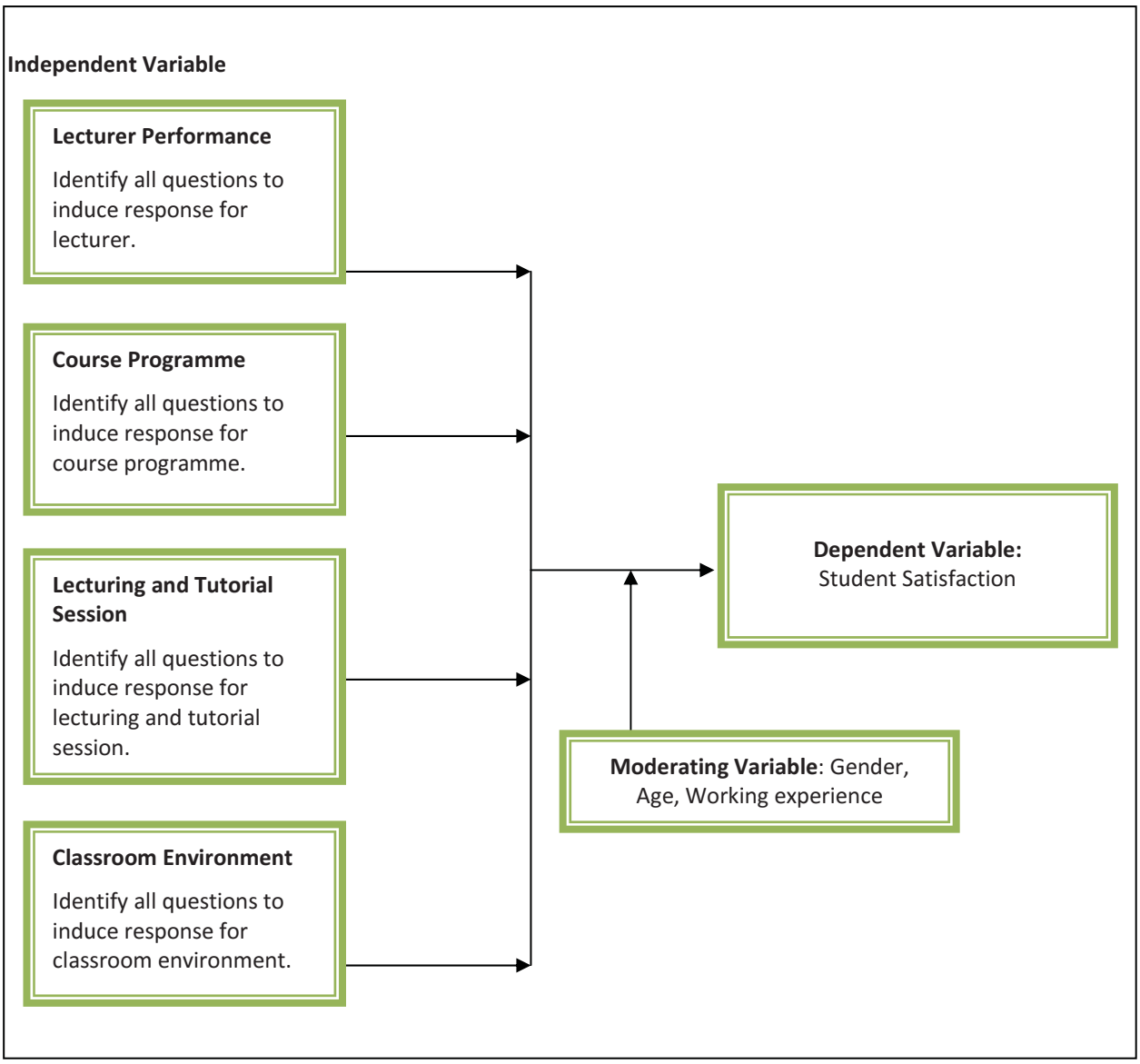


Figure 3.2 Model flow chart

What can you do with the Questionnaire Responses?

1. Find out the number of questionnaires distributed and the usable questionnaires received (i.e. the response rate).
2. You can summarize the data based on your **moderation variables**.
   - Summary of response based on *Gender*
   - Summary of response based on A*ge*
   - Summary of response based on W*orking Experience*

3. For the descriptive analysis, you can summarize and find the mean for each **Independent Variable** in the questionnaire.

Example: For questions on independent variable 'Lecturer' (we have devised 6 questions for this sub group, refer to the full questionnaire at the end):

- The average response rate for question 1.
- The average response rate for question 2.
- The average response rate for question 3.
- The average response rate for question 4.
- The average response rate for question 5.
- The average response rate for question 6.

Find the average response rate for lecturer by averaging Q1-Q6 above

Example: Summary for Lecturers

Question Number	Mean
1.	mean Q1 *(average of all responses for Q1)*
2.	mean Q2 *(average of all responses for Q2)*
3.	mean Q3 *(average of all responses for Q3)*
4.	mean Q4 *(average of all responses for Q4)*
5.	mean Q5 *(average of all responses for Q5)*
6.	mean Q6 *(average of all responses for Q6)*
	Mean of Q1-Q6 *(overall mean for Lecturers)*

- Repeat the steps above for independent variable *Course Programme*.
- Repeat the above steps for independent variable *Lecturing and Tutorial session*.
- Repeat the above steps for *Classroom Environment*.

4. All the above information can be summarized once again in an overview table if you wish.

Example:

	Objective	Mean
1	Summary of Lecturers	Overall mean of Lecturers
2	Summary of Course Programme	Overall mean of Course Programme
3	Summary of Lecturing and Tutorial	Overall mean of Lecturing and Tutorial
4	Summary of Classroom Environment	Overall mean of Classroom Environment
	Overall Students' Satisfaction	Overall mean of Objectives 1-4

5. Hypothesis can be tested

- Can be done on Moderating Variables *(Gender, Working Experience, Age)* and the Dependent Variable *(Student Satisfaction)*

Example:

a. on gender (male and female differ in different satisfaction level)

b. on working experience (working experience has an effect on satisfaction level)

c. on age of respondents (age has a bearing on satisfaction level)

6. Correlation Analysis and Testing (to test the existence of relationship between the independent variables)

Now you are ready to generate the above information using SPSS.

3.2 SPSS Environment

SPSS (**S**tatistical **P**ackage for the **S**ocial **S**ciences) is one of the most widely available and powerful statistical software packages. It covers a broad range of statistical procedures that allow you to;

- Summarize data (e.g., compute means and standard deviations),
- Determine whether there are significant differences between groups (e.g., t-tests, analysis of variance),
- Examine relationships among variables (e.g., correlation, multiple regression), and
- Graph results (e.g., bar charts, line graphs).
- And many other functions.

Guide to Using SPSS

First to create a new data, click on File menu, New, and Data.

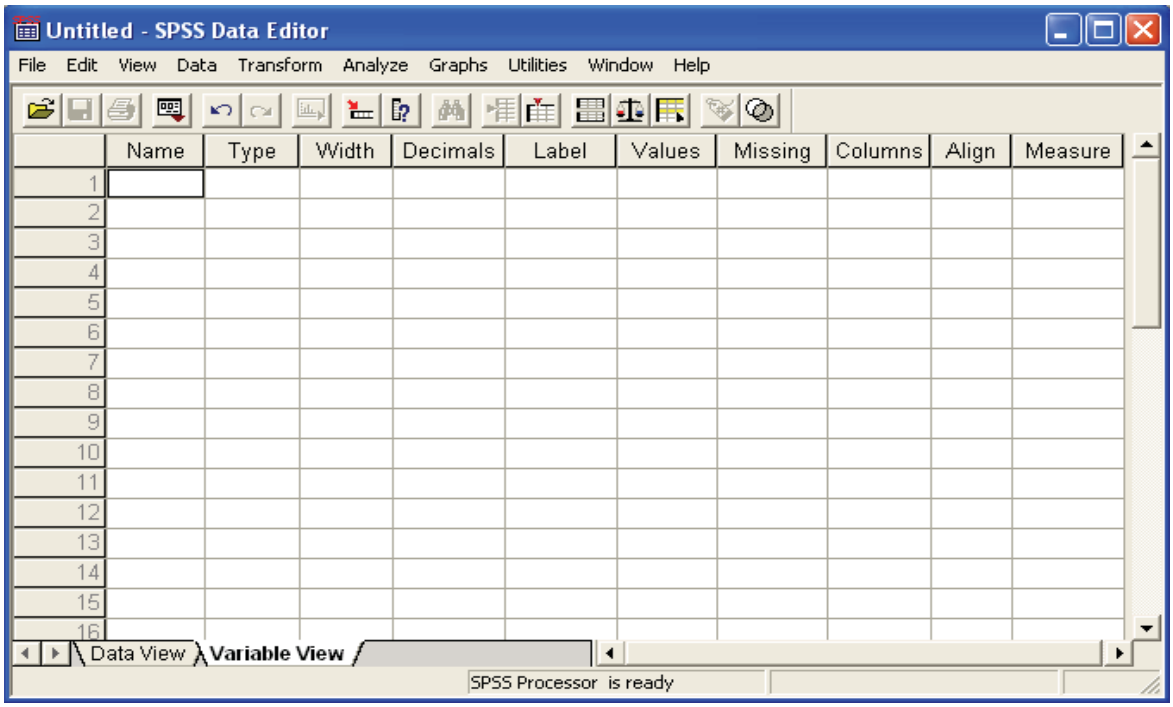


Figure 3.3 Data Editor (applicable for version 17 and lower, version 20 refer to appendix IX, page 193)

The various items and names in the above table are explained below:

3.2.1 Two Main Tabs

1. Data View

   → Consists of variables and values. These values are the data gathered from the survey questionnaire.

2. Variable View

   → Consists of all the variables being analyzed or studied. These variables are dependent, independent, or moderator.

Variable View Columns:

a. ***Name***. The variable name should start with a letter, and contain only letters and numbers *(no punctuation or spaces)*

b. ***Type***. The figure below shows the list of variable types.

Commonly used variable types:

i. Numeric - deals with a number. You can specify the number of decimal places that SPSS displays by typing the desired number in the Decimal Places text box.

ii. String - refers to a word or other alphabetic characters.

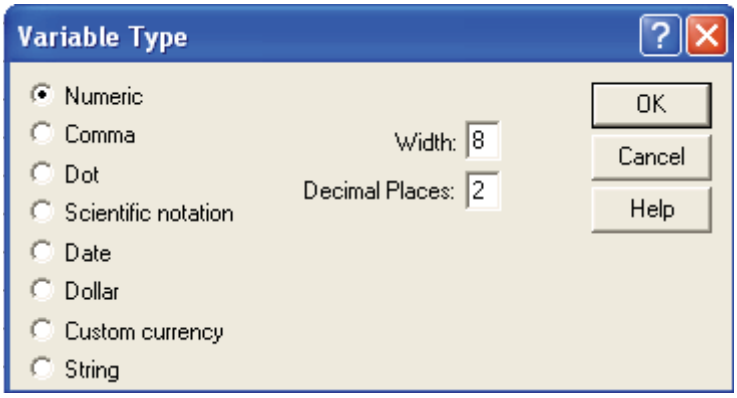


Figure 3.4 Variable Types

c. ***Width***. This refers to the total number of digits or length of the value.

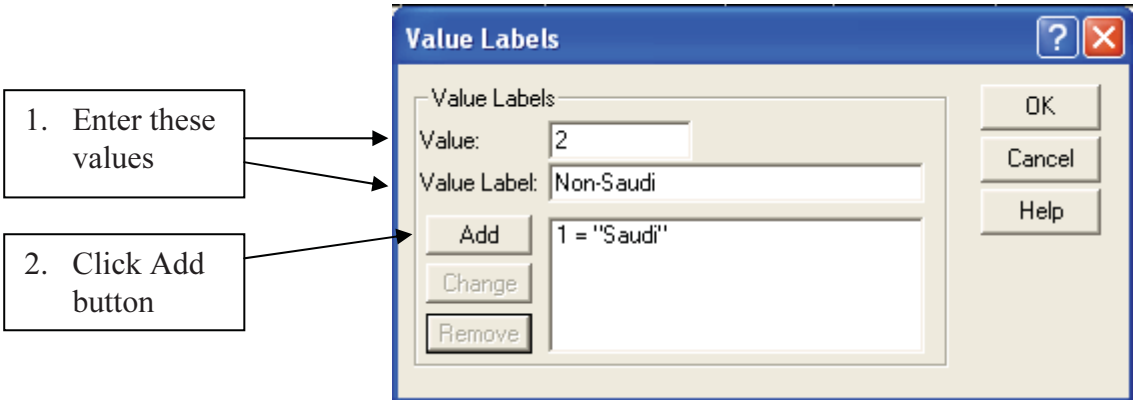
d. ***Decimals***. Number of decimal places.

e. ***Label***. This refers to the description of the variable.

f. ***Values***. This refers to the descriptive value labels for each value of a variable. This is useful if your data file uses numeric codes to represent non-numeric categories (for example, codes of 1 and 2 for male and female). Value labels can be up to 60 characters

long. Value labels are not available for long string variables (string variables longer than 8 characters).

Sample entries:



Result:

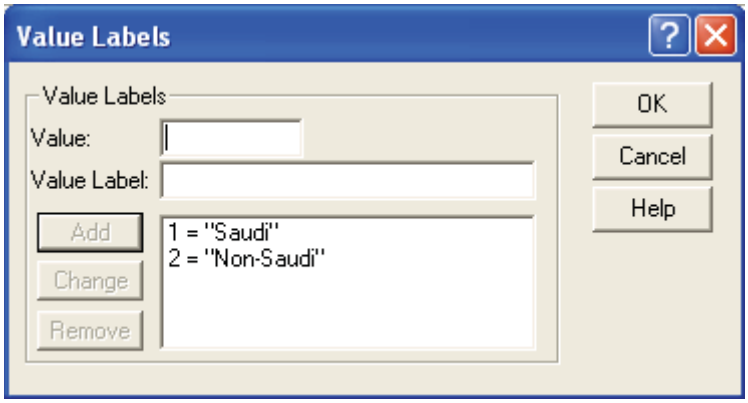


Figure 3.5 Variable Labels

g. ***Missing***. Missing Values defines specified data values as user-missing. It is often useful to know why information is missing. For example, you might want to distinguish between data missing because a respondent refused to answer and data missing because the question didn't apply to that respondent. Data values specified as user-missing are flagged for special treatment and are excluded from most calculations. To change the value of the missing column, the following window is used (can be opened through clicking the ellipsis button).

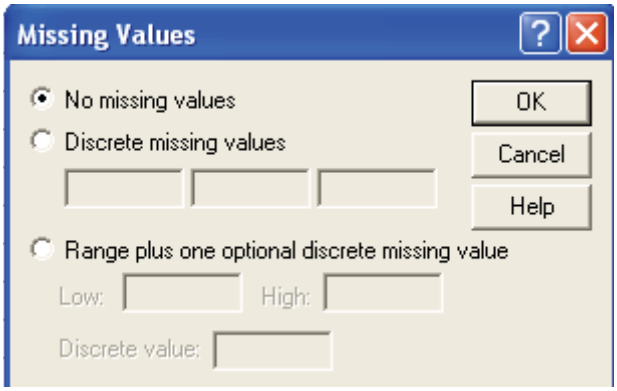


Figure 3.6 Missing values

- You can enter up to three discrete (individual) missing values, a range of missing values, or a range plus one discrete value.
- Ranges can be specified only for numeric variables.
- You cannot define missing values for long string variables (string variables longer than eight characters).

Missing values for string variables; all string values, including null or blank values, are considered valid values unless you explicitly define them as missing. To define null or blank values as missing for a string variable, enter a single space in one of the fields for Discrete missing values.

h. ***Columns***. Number of characters for the column width.

i. ***Align***. This refers to vertical alignment of values. Values may be aligned left, right or centered.

j. ***Measure.***

You can select one of three measurement levels:

Click on the down arrow and select the appropriate level of measure (sometimes you have to click on the down arrow twice to get the drop down menu to appear):

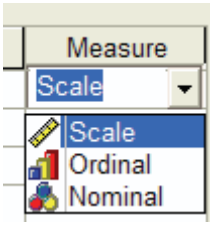


Figure 3.7 Scale types

i. *Levels of measurement*. Data values are numeric values on an interval or ratio scale (e.g., age, income). Scale variables must be numeric.

ii. *Ordinal*. Data values represent categories with some intrinsic order (e.g., low, medium, high; strongly agree, agree, disagree, strongly disagree). Ordinal variables can be either string (alphanumeric) or numeric values that represent distinct categories (e.g., 1=low, 2=medium, 3=high).

Note: for ordinal string variables, the alphabetic order of string values is assumed to reflect the true order of the categories. For example, for a string variable with the values of low, medium, high, the order of the categories is interpreted as high, low, medium -- which is not the correct order? In general, it is more reliable to use numeric codes to represent ordinal data.

iii. *Nominal*. Data values represent categories with no intrinsic order (e.g., job category or company division). Nominal variables can be either string (alphanumeric) or numeric values that represent distinct categories (e.g., 1=Male, 2=Female).

(Take note: For SPSS-format data files created in earlier versions (lower than version 17 of SPSS products, the following rules apply:

String (alphanumeric) variables are set to nominal.

String and numeric variables with defined value labels are set to ordinal.

Numeric variables without defined value labels but less than a specified number of unique values are set to ordinal.

Numeric variables without defined value labels and more than a specified number of unique values are set to scale.)

Illustration: As practice, enter the following taken from Section A and part of Section Two *(Note: see sample questionnaire appendix IV, page 171)* in a new SPSS data editor in the **Variable View**. Save this as `Sample.sav`

A Portion of Sample Questionnaire

SECTION ONE: PERSONAL DATA

1. Your Gender: □ Male □ Female
2. Your Age □ Less than 22 years
   □ 22 to 25 years
   □ More than 25 years
3. Your Residency □ Saudi
   □ Non-Saudi
4. Name of Program Taken _____
5. Program Duration _____
6. Do you have any past working experience? □ Yes □ No

Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agee	Strongly Agree
1	2	3	4	5

Section Two: Lecturer

	1	2	3	4	5
7. Have your lecturers always take concern about your academic advancement?					
8. I found my lecturers helpful if I encountered difficulties in my subject area.					

Upon entry of the above Information, you're **Variable View** should look like this:

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
Gender	String	1	0	Gender	f = "female" m = "male"	None	1	Left	Nominal
Age	Numeric	1	0	Age	1 = "< 22" 2 = "22-25" 3 = ">25"	None	1	Right	Scale
Race	Numeric	1	0	Race	1.00= "Saudi" 2.00 = "Non-Saudi"	None	1	Right	Nominal
Program	String	25	0	Programme taken	None	None	25	Left	Nominal
Duration	String	20	0	Programme duration	None	None	20	Left	Nominal
work_exp	Numeric	1	0	Work experience	0 = "No" 1 = "Yes"	None	1	Right	Nominal
q7	Numeric	1	0	Lecturer Q-1	1= "Strongly Disagree" 2= "Disagree" 3= "Neither Agree Nor Disagree" 4= "Agree" 5= "Strongly Agree"	None	1	Right	Ordinal
q8	Numeric	1	0	Lecturer Q-2	same as q7	None	1	Right	Ordinal
Repeat the process for the remaining variables – change only the label of each variable.									
q30									

Figure 3.8 The variable view

The data entered here by the readers will serve as input data for cases 1, 2 and 3 discussed next.

3.2.2 Sample Cases

Once you are familiar with entry of Information in **Variable View**, you should go to **Data View**. Data View consists of variables and values. These values are the data gathered from the survey questionnaire and you should enter each of the responses in Data View.

Based on the data entered in Data View, we can create summary of responses (frequency responses) based on *Moderating Variables*; Gender, Age and Working Experience. The following sections will illustrates how you go about with analyzing the data.

Case 1: Frequencies

Problem #1. Suppose we like to create a *summary of responses* based on GENDER. Do the following SPSS steps:

1. Open the SPSS data file created above, `Sample.sav`.
2. From the main menu, click

Analyze → Descriptive Statistics → Frequencies…

This window will appear…

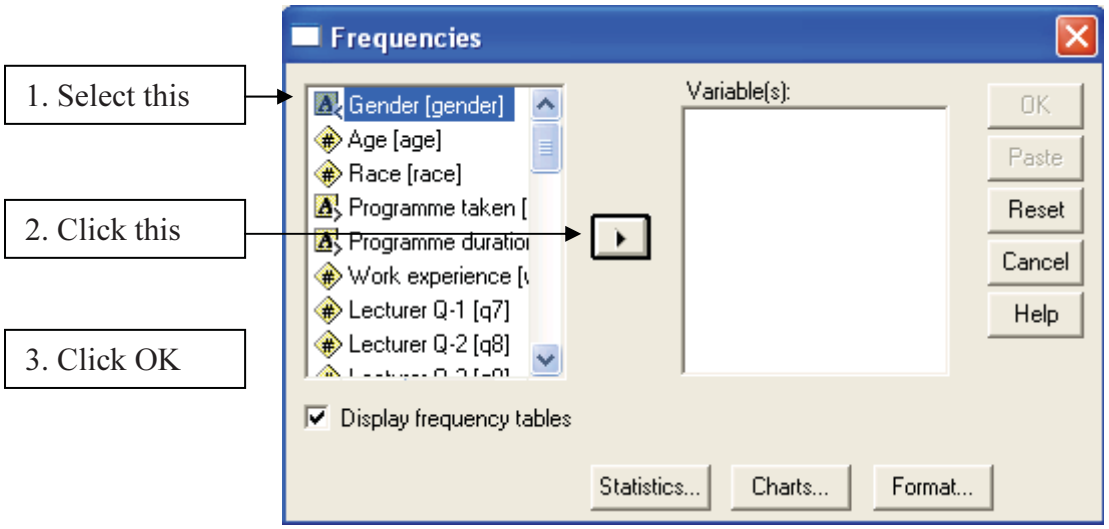


Figure 3.9 Window view of frequencies

The result is:

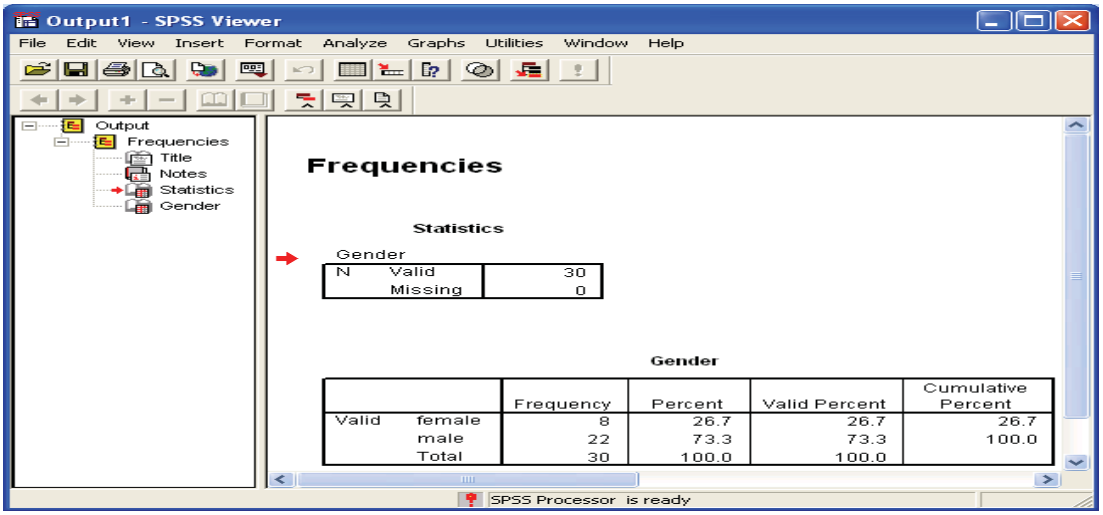


Figure 3.10 Frequencies output table

Problem #2. Create *summary of responses* based on AGE.

Do the same steps as the previous example on Gender. Only change the variable to 'AGE'.

Result:

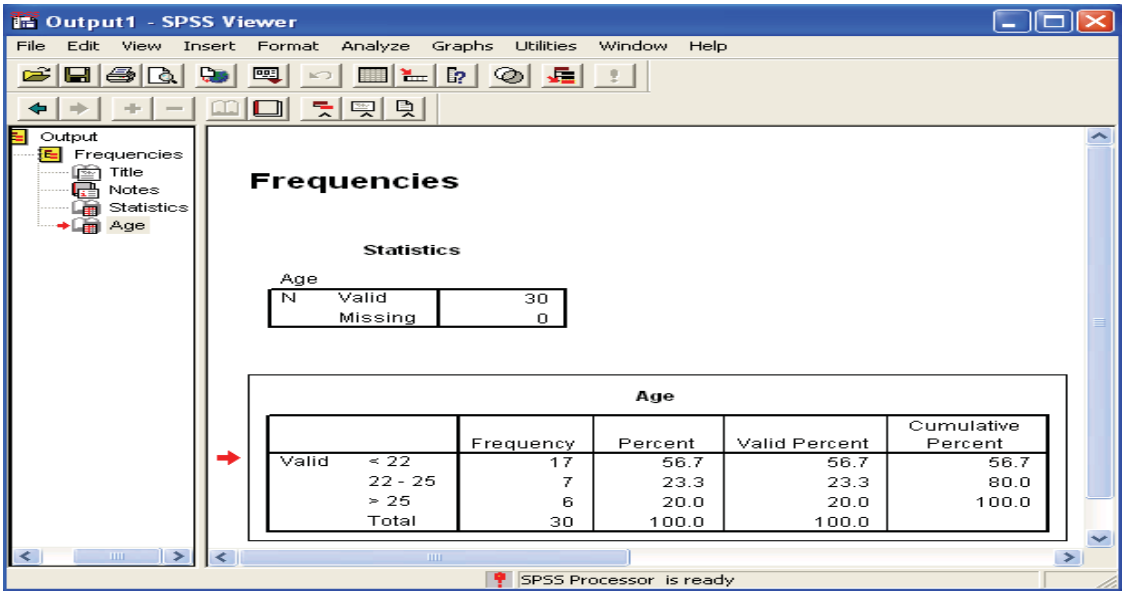


Figure 3.11 Frequencies output table

Figures 3.10 and 3.11shows how frequency can be displayed, for figure 3.10 it is shown along Gender and figure 3.11 along Age components in the rows.

Problem #3. Create *summary of responses* based on WORKING EXPERIENCE.

Applying the steps above, you should get the same output with the following result.

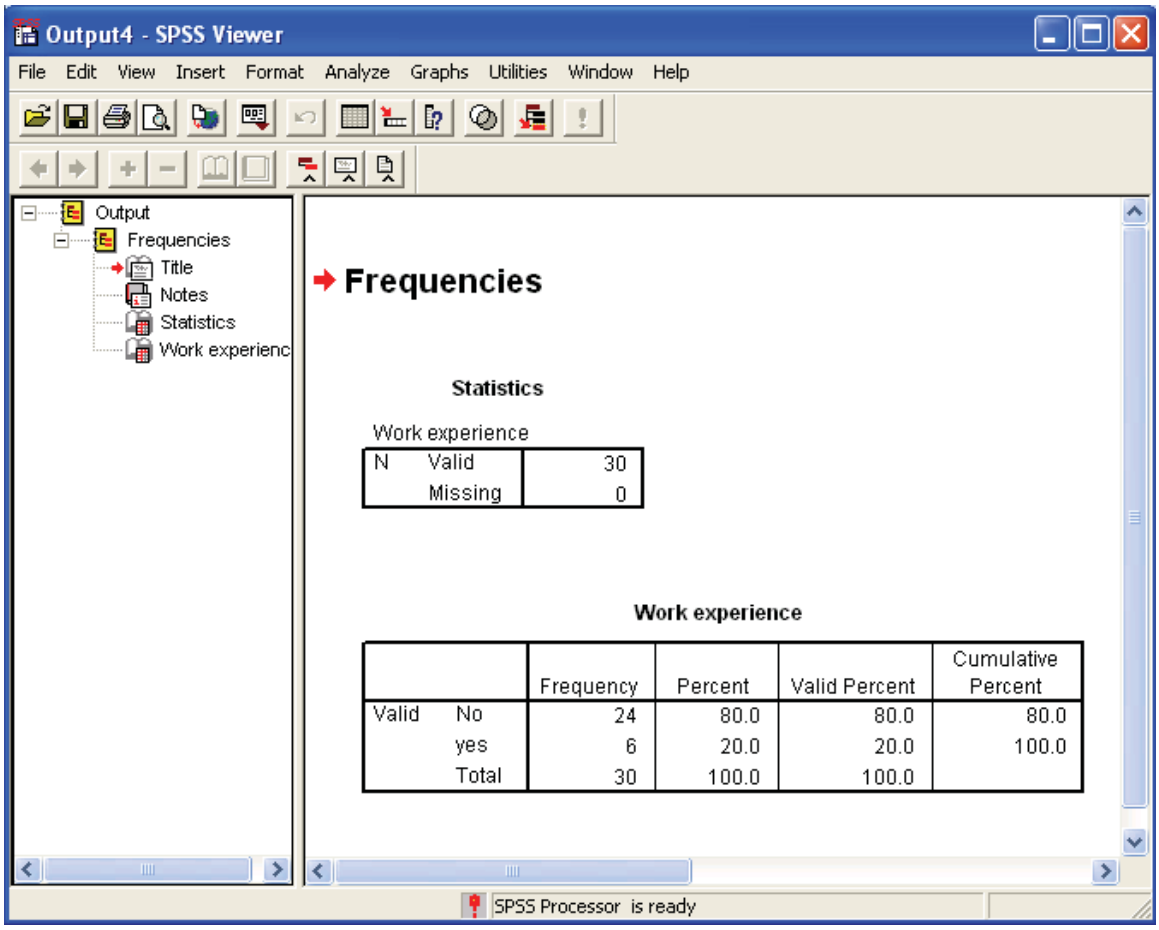


Figure 3.12 Frequencies output table

Figure 3.12 shows how frequency can be displayed along Yes and No response.

The rows indicate these categories.

What You Can Do with Frequency Data

Explain the Patterns Observed in Data:

a. Begin with Exploratory Analysis;

i. Show specific values with the use of table to summarize the data.

ii. Where possible group data into categories thorough the use of charts; bar chart, histogram and pie chart.

b. Go for Comparing Variables;

i. Show specific values, interdependence with the use of cross-tabulation and contingency table.

ii. Compare highest and lowest with the use of multiple bar charts.

iii. Compare proportion showing percentage component bar chart or two or more pie charts.

iv. Compare trends- multiple line chart

v. Compare totals- stacked bar chart.

vi. Compare proportions and totals- comparative proportional pie charts.

vii. Show relationship between cases for variables-use scatter plots

Note: If your research involves just analyses of data without going to specific hypotheses testing, frequency analysis is sufficient. Student can do data analyses, charts, cross charts etc. and write-up based on the frequencies.

For others with a Research Question and hypotheses testing in mind, they need to go beyond.

Going Beyond Exploratory Analysis

Beyond Exploratory Analysis we can summarize responses for the *Independent Variables*. This is done by looking at each subgroup.

In the case of Lecturers Performance we have six questions (refer to sample questionnaire). We can find the mean responses by summarizing the responses for question 1 from all the respondents. Then repeat the process for question 2 through 6.

Finally, find the mean of all the individual question means. This will give the *overall mean response value for Lecturer Performance* (summary of all six questions' means).

We repeat the above steps for all the independent subgroups namely Lecturers' Performance, Course Programme, Lecturing and Tutorial Session and Classroom Environment

The following steps will illustrate.

Case 2: Descriptive Analysis – Method 1 (no grand mean)

This method creates a summary of mean for every question only. The GRAND MEAN will not be shown in the final report. From this technique, the researcher may compute separately the grand mean (i.e. averaging all individual question means) when doing the actual word processing of documents.

Note: Case 3 Method will show the computation of mean including the grand mean.

Problem #1. Compute the `mean` of all responses for questions under 'Lecturer' category.

SPSS Steps:

1. From the main menu, click

Analyze → Descriptive Statistics → Descriptives…

2. Select questions 7 to 12. Click OK.

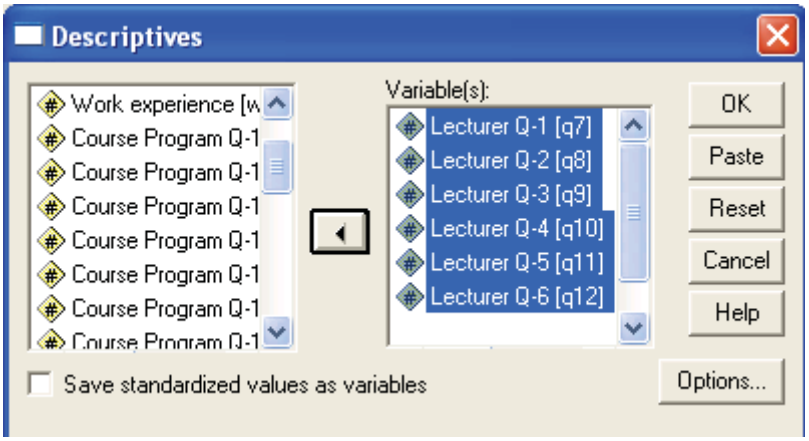


Figure 3.13 Descriptives window

Result:

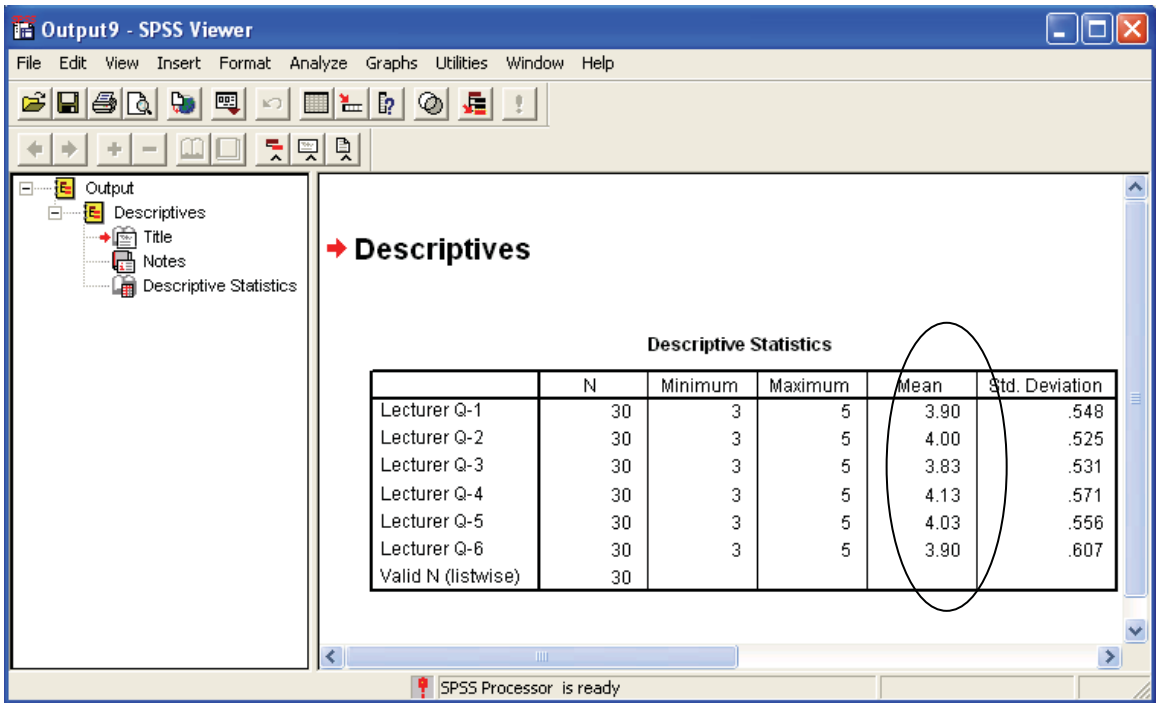


Figure 3.14 Output for Descriptives data

Figures 3.13 and 3.14 display how individual question responses can be summarized.

The rows in Figure 3.14 show means and standard deviations for Q1 to Q6.

Based on the Original Model this is what you have done so far for one of the objectives:-

Lecturer Performance

Summary of Average responses

Q1: mean = 3.90
Q2: mean = 4.00
Q3: mean = 3.83
Q4: mean = 4.13
Q5: mean = 4.03
Q6: mean = 3.90

Mean of Q1-Q6 = 3.97

What you can do with the above Information for Objective 1: Lecturer Performance

Comment based on the scale you have used (1-strongly disagree, 2- disagree, 3- neither agree nor disagree, 4- agree and 5-strongly agree)

Write a commentary on average responses for Q1

Write a commentary on average responses for Q2

Write a commentary on average responses for Q3

Write a commentary on average responses for Q4

Write a commentary on average responses for Q5

Write a commentary on average responses for Q6

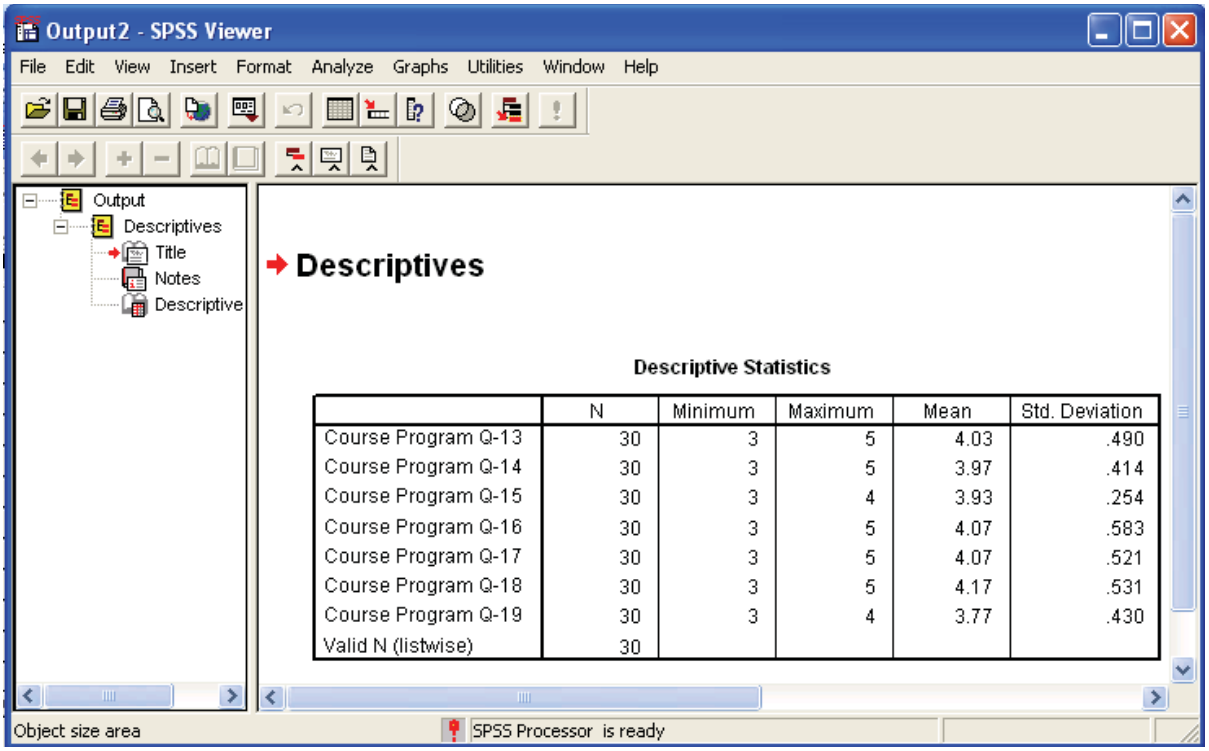
And finally an Overall Commentary for the averages for Q1-Q6

Problem #2. Compute the `mean` of all responses for questions under 'Course Program' category.

To create new group of questions, click the `Reset` button.

Do the same steps as above. Select ONLY questions 13 to 19.

Result:



	N	Minimum	Maximum	Mean	Std. Deviation
Course Program Q-13	30	3	5	4.03	.490
Course Program Q-14	30	3	5	3.97	.414
Course Program Q-15	30	3	4	3.93	.254
Course Program Q-16	30	3	5	4.07	.583
Course Program Q-17	30	3	5	4.07	.521
Course Program Q-18	30	3	5	4.17	.531
Course Program Q-19	30	3	4	3.77	.430
Valid N (listwise)	30				

Figure 3.15 Output for Descriptives data

Figure 3.15 is a repeat of process in figure 13 and figure 3.14, it shows how individual question responses can be summarized.

What you can do with the above Information for Objective 2: Course Program

Repeat the same procedure as for Objective 1

Problem #3. Compute the mean of all responses for questions under 'Lecturing and Tutorial Sessions' category.

To create new group of questions, click the Reset button.

Do the same steps as above. Select ONLY questions 20 to 26.

Result:

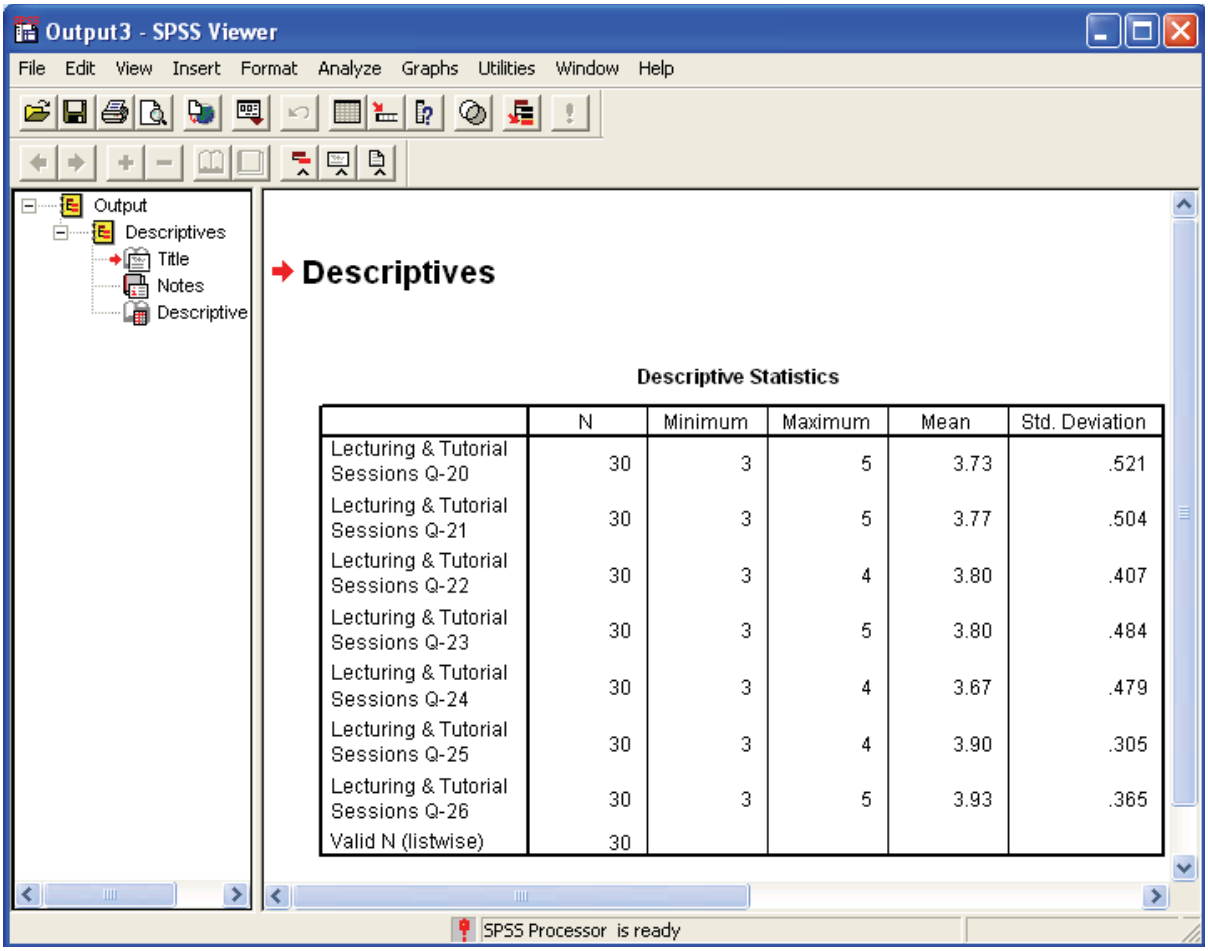


Figure 3.16 Output for Descriptives data

Figure 3.16 is a repeat of process in figures 3.13 and 3.14, it shows how individual question responses can be summarized.

What you can do with the above Information for Objective 3: Lecturing & Tutorial Sessions

Repeat the same procedure as for Objective 1

Problem #4. Compute the mean of all responses for questions under 'Classroom Environment' category.

To create new group of questions, click the `Reset` button.

Do the same steps as above. Select ONLY questions 27 to 30.

Result:

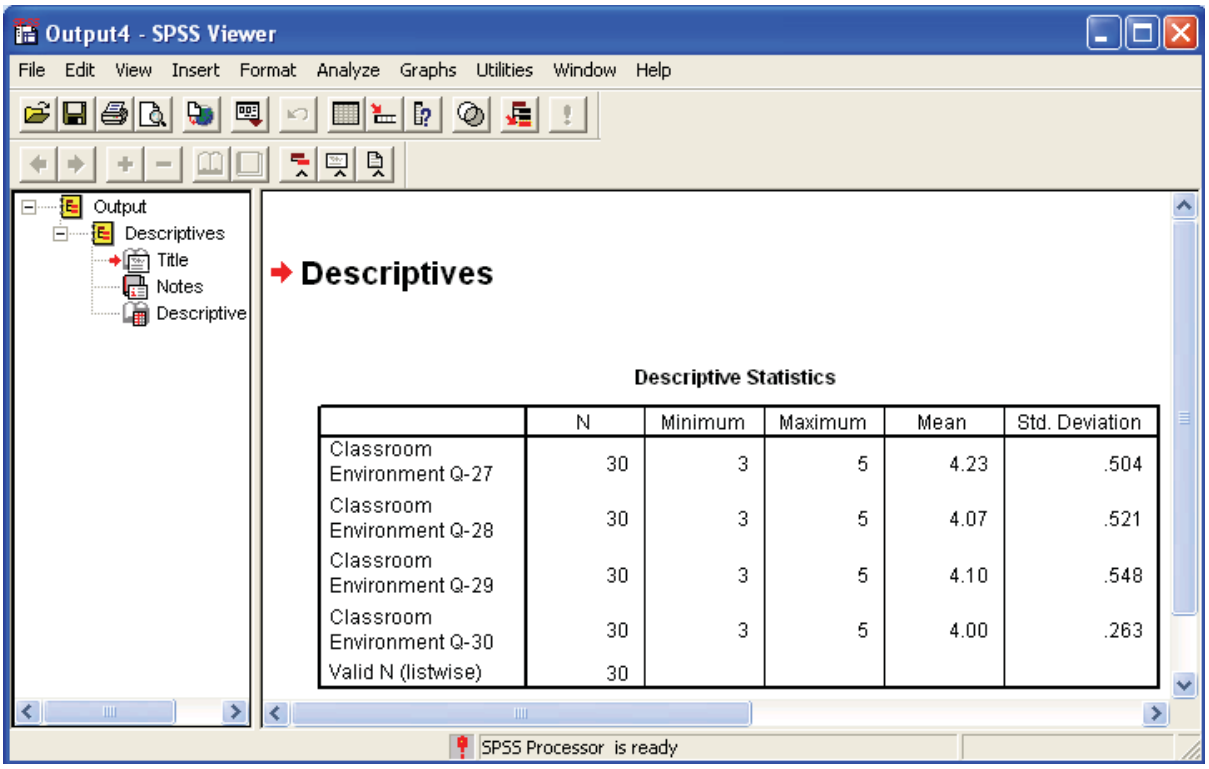


Figure 3.17 Output for Descriptives data

Figure 3.17 is again a repeat of process in figures 3.13 and 3.14, it shows how individual question responses can be summarized for specific objective.

What you can do with the above Information for Objective 4: Classroom Environment

Repeat the same procedure as for Objective 1

Case 3: Descriptive Analysis – Method 2 (with grand mean)

This method is an enhanced creation of reports from the examples given in Case 2.

Problem #1. To create a report that includes the grand `mean` of all the responses for questions under 'Lecturer' category.

A. TRANSFORM method:

1. Open the file (transposed), `Sample.sav`
2. From the main menu, click `Transform` → `Compute`...

3. Enter the data specified below:

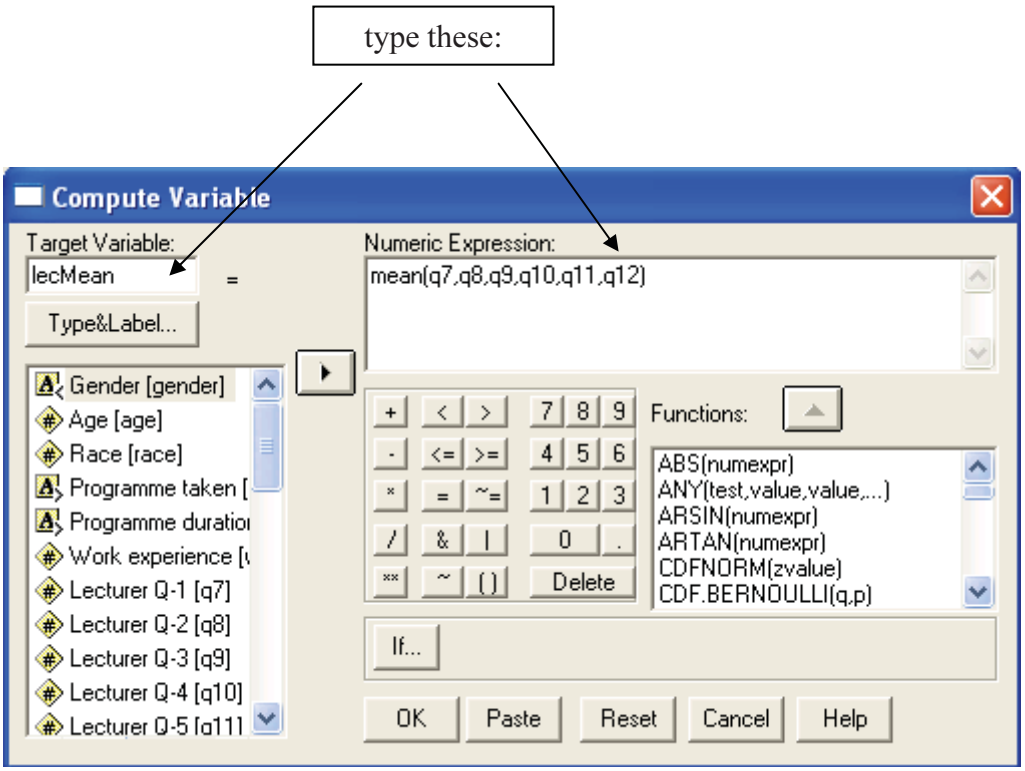
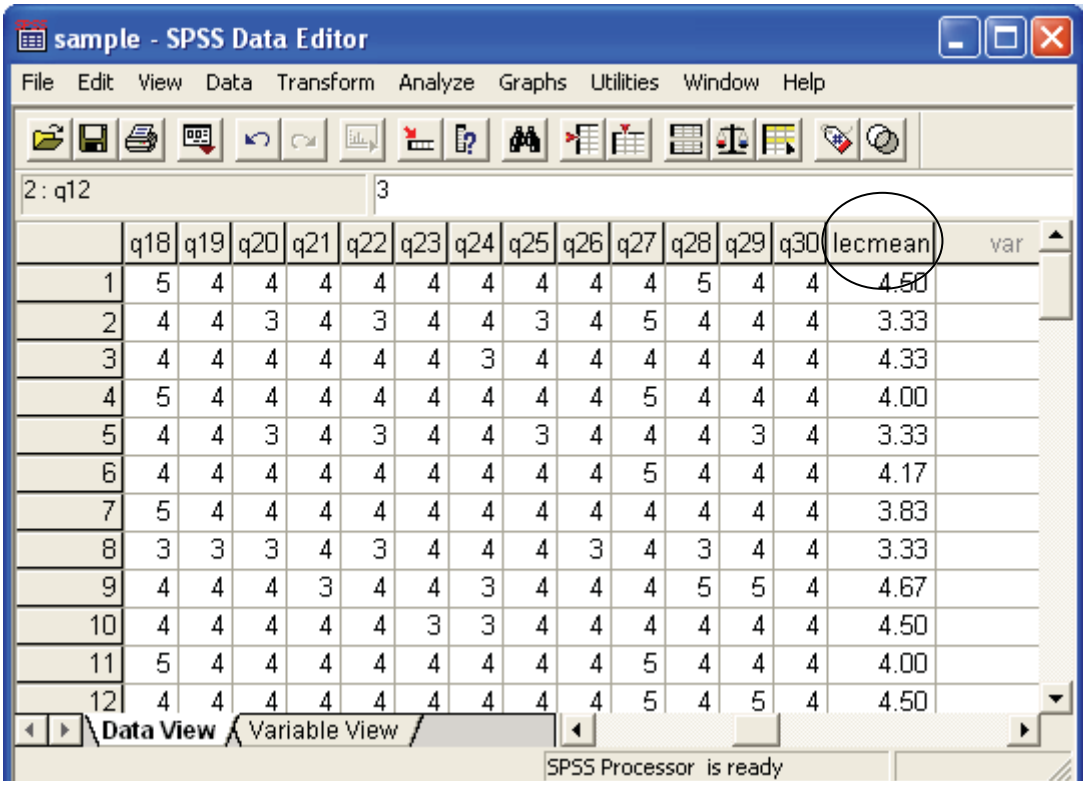


Figure 3.18 Computation of variables

From the Numeric Expression, variables `q7..q12` corresponds to question 7 to question 12 in the questionnaire which are under the 'lecturer' category. The `mean ()` function computes the mean of the variables enclosed and stores the value in target variable `lecMean`.

4. Click OK button. Save the file.

New variable (column) is added, **lecMean** – this variable *(or any suitable name of your choice)* holds the mean of all responses for each question.



	q18	q19	q20	q21	q22	q23	q24	q25	q26	q27	q28	q29	q30	lecmean	var
1	5	4	4	4	4	4	4	4	4	4	5	4	4	4.50	
2	4	4	3	4	3	4	4	3	4	5	4	4	4	3.33	
3	4	4	4	4	4	4	3	4	4	4	4	4	4	4.33	
4	5	4	4	4	4	4	4	4	4	5	4	4	4	4.00	
5	4	4	3	4	3	4	4	3	4	4	4	3	4	3.33	
6	4	4	4	4	4	4	4	4	4	5	4	4	4	4.17	
7	5	4	4	4	4	4	4	4	4	4	4	4	4	3.83	
8	3	3	3	4	3	4	4	4	3	4	3	4	4	3.33	
9	4	4	4	3	4	4	3	4	4	4	5	5	4	4.67	
10	4	4	4	4	4	3	3	4	4	4	4	4	4	4.50	
11	5	4	4	4	4	4	4	4	4	5	4	4	4	4.00	
12	4	4	4	4	4	4	4	4	4	5	4	5	4	4.50	

Figure 3.19 Output for computation of variables

Figure 3.19 shows the output with the uses of 'Compute Variable' to find overall mean of questions Q7 through Q12.

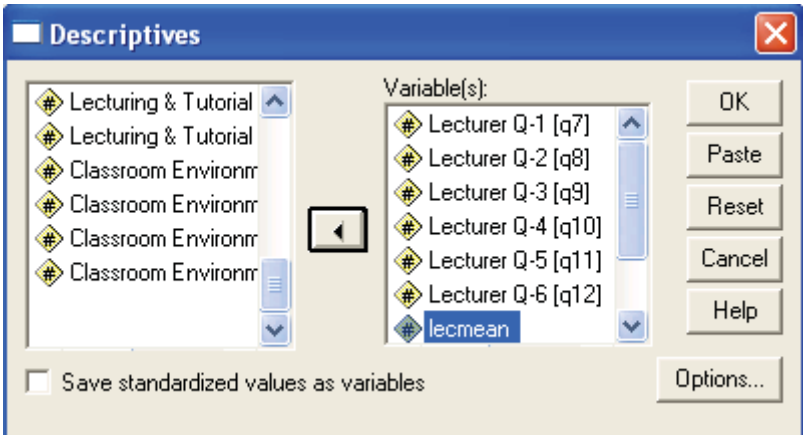
B. CREATION OF REPORTS:

To view the formatted output or report of the above data, you can apply the previous technique on **Descriptives**.

For instance, we would like to create a report on the grand mean of all the means we computed above, do the following steps:

Analyze → Descriptive Statistics → Descriptives…

Select the variables shown below, in that order, then click OK.



Result:

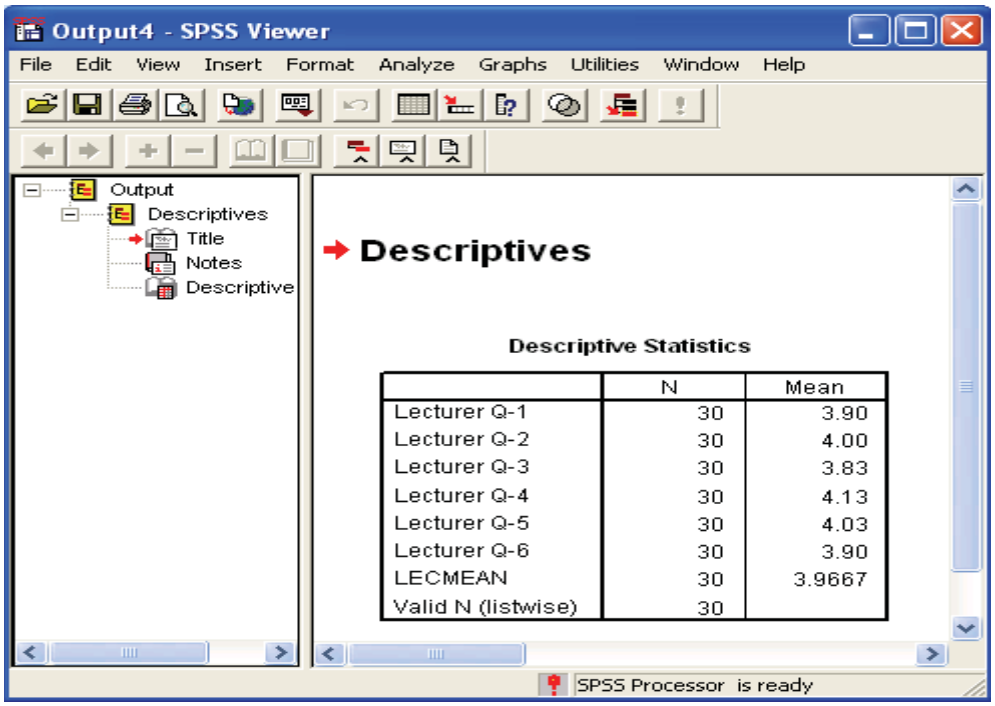


Figure 3.20 Descriptive variables with output viewer

Note: You can change the `options` to modify the information shown in the table, save this file, edit the table's text, copy-paste the table on to other application software like MS Word, etc.

Self-Assessment:

Apply the previous steps to other categories or groups of questions. These are the `course program` (q13 to q19), `lecturing and tutorial` (q20 to q26), and `classroom environment` (q27 to q30).

A recap of All Summary Data:

Descriptive Statistics

	N	Mean
Course Program Q-13	30	4.03
Course Program Q-14	30	3.97
Course Program Q-15	30	3.93
Course Program Q-16	30	4.07
Course Program Q-17	30	4.07
Course Program Q-18	30	4.17
Course Program Q-19	30	3.77
COUMEAN	30	4.0000
Valid N (listwise)	30	

Descriptive Statistics

	N	Mean
Lecturing & Tutorial Sessions Q-20	30	3.73
Lecturing & Tutorial Sessions Q-21	30	3.77
Lecturing & Tutorial Sessions Q-22	30	3.80
Lecturing & Tutorial Sessions Q-23	30	3.80
Lecturing & Tutorial Sessions Q-24	30	3.67
Lecturing & Tutorial Sessions Q-25	30	3.90
Lecturing & Tutorial Sessions Q-26	30	3.93
TUTMEAN	30	3.8000
Valid N (listwise)	30	

Descriptive Statistics

	N	Mean
Classroom Environment Q-27	30	4.23
Classroom Environment Q-28	30	4.07
Classroom Environment Q-29	30	4.10
Classroom Environment Q-30	30	4.00
CLAMEAN	30	4.1000
Valid N (listwise)	30	

Summary of the 4 means computed above:

Do Case 3 Transform Method to create a **ss** (student satisfaction) variable that computes the mean of variables `lecmean`, `coumean`, `tutmean`, and `clamean`. The figure below shows the entries of these variables.

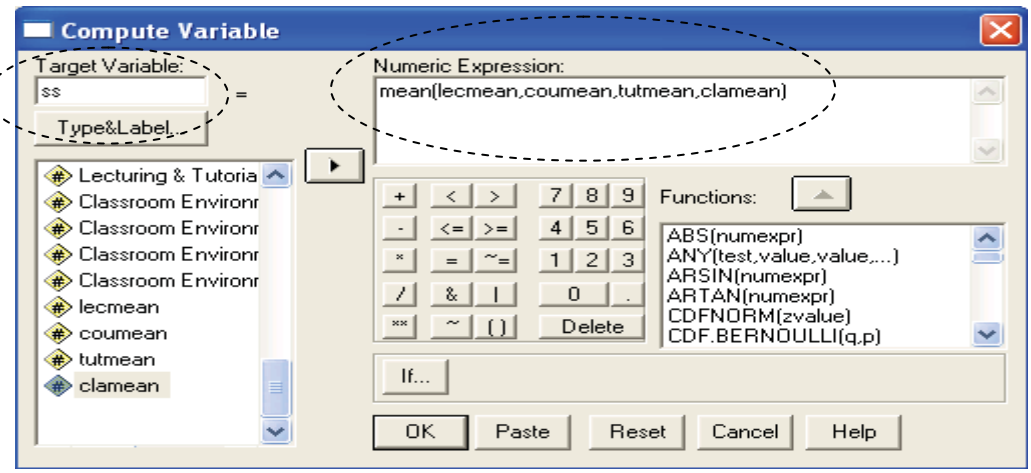


Figure 3.21 Computation of variable

Then, do Case 3 – Creation of Reports

Select the variables as shown below.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
LECMEAN	30	3.33	4.67	3.9667	.39246
COUMEAN	30	3.14	4.57	4.0000	.29540
TUTMEAN	30	3.57	4.29	3.8000	.21420
CLAMEAN	30	3.50	4.75	4.1000	.30513
SS	30	3.45	4.32	3.9667	.25050
Valid N (listwise)	30				

Grand Mean

We have the following results. Mean *Lecturer Performance* (LECMEAN) 3.9667, mean *Course Programme* (COUMEAN) 4.0000, mean *Lecturing and Tutorial Session* (TUTMEAN) 3.8000, mean *Classroom Environment* (CLAMEAN) 4.1000 and *Students Satisfaction* (SS) means 3.9667. These are the output with the use of 'Compute Variable' on all the objectives.

Mean SS = mean {LECMEAN + COUMEAN + TUTMEAN + CLAMEAN}

Based on the Model this is what you have done so far

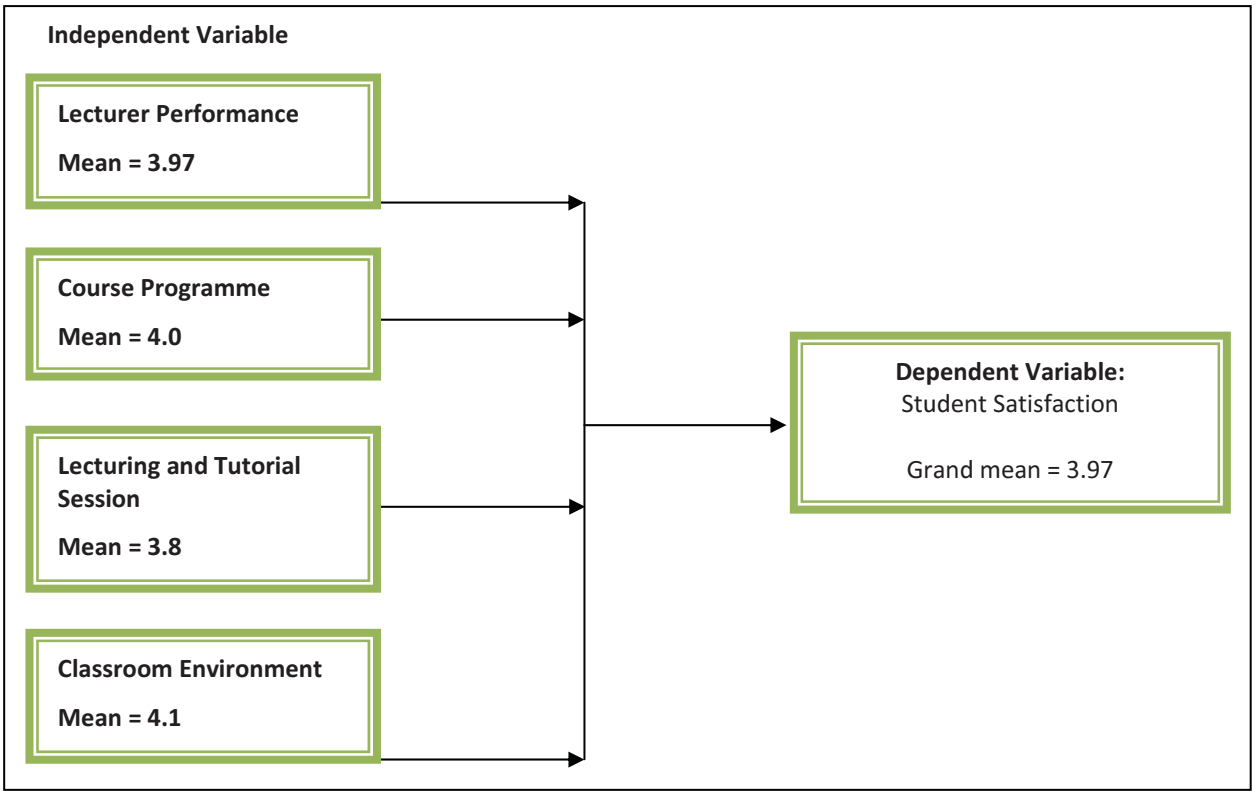


Figure 3.22 Research model

What can you do with the Grand Mean Summary Data?

The overall mean (Grand Mean) is **3.97**. Now you can comment on the **Research Question** i.e. Students' Satisfaction based on the Likert Scale used earlier and shown below.

Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1	2	3	4	5

Case 4: Summary of the Overview of All Variables

A survey is not complete without a table summary of all the variables involved in the research. We will explain how to obtain the summary for Students Satisfaction first.

The dependent variable "Students' Satisfaction" can be categorized in different levels like; Below average, Average, Above average (you devise your scale for these 3 levels based on your research requirements). We then compare the response frequency for Students Satisfaction with the various moderating variables like Gender, Age and Work Experience.

Similarly, you can categorize the independent variables and moderating variables and add them into the same table.

Illustration:

Suppose we like to create a *summary of the overview of all variables* that may be of the same table presentation as shown below:

Student Satisfaction	**Frequency**	**Gender**		**Age**			**Work Experience**		**Mean**	**Std dev**
		M	**F**	**<22**	**22-25**	**>25**	**Yes**	**No**		
Below Average										
Average										
Above Average										
Total										

As the example above we have selected 1 through 2.5 scales (from the Likert type Scale) to represent Below Average, more than 2.5 to 3.5 as Average and more than 3.5 to 5 as Above Average. Or alternatively use the Original Scale Rating as used in Grand Mean Summary Data above.

A. Steps to compute for the frequency of student satisfaction under *below average category*:

1. From the menu, click **Transform → Count…**
2. Enter the following data:

a. Target Variable: `ss_below`

Note: You can use any other suitable name here.

b. Numeric Variables: `ss`

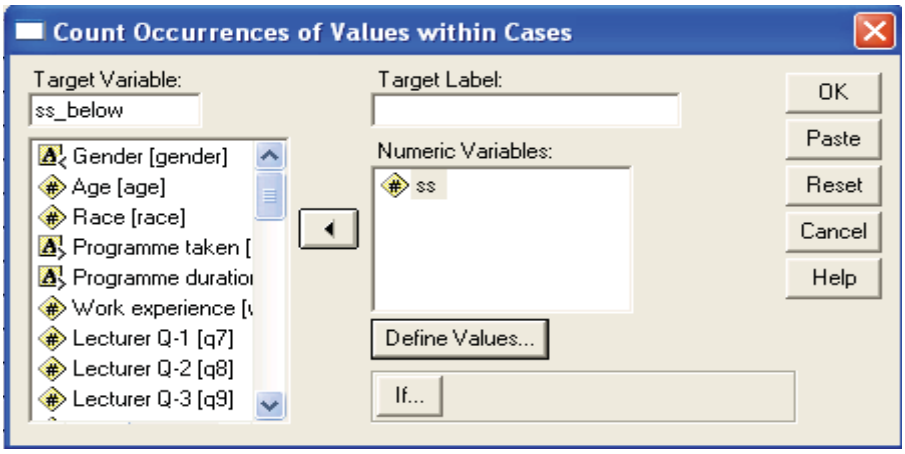


Figure 3.23

c. Define Values

i. Enter values 1 through 2.5 in the **range**.

ii. Click Add button.

1 thru 2.5 will be displayed in the right box, Values to Count.

iii. Click Continue button.

This window will be closed, and will return to the above window.

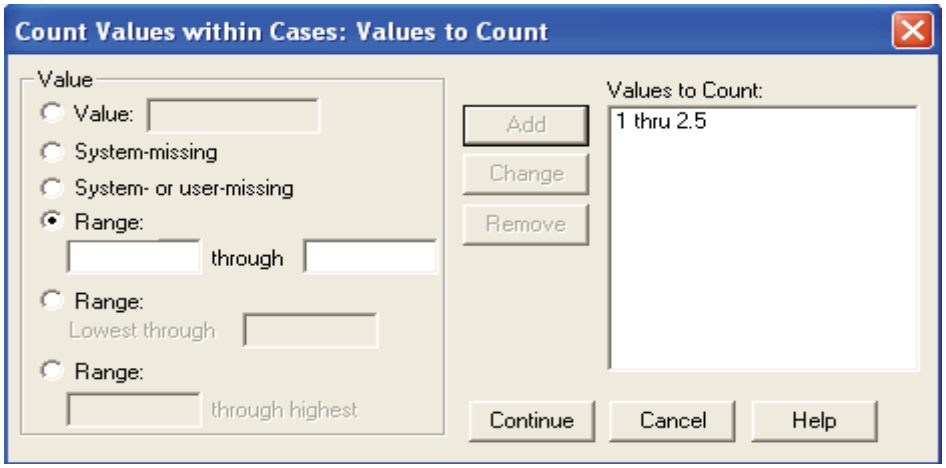


Figure 3.24

d. Click OK button.

Check out in the Data View that new column 'ss_below' is added.

B. To compute for the frequency of student satisfaction under *average category*, do the same steps as above but change the following values:

1. Target Variable: `ss_ave`
2. Numeric Variables: `ss`
3. Define Values:

   Range: `2.6` through 3.5

C. To compute for the frequency of student satisfaction under *above average category*, do the same steps as above but change the following values:

1. Target Variable: ss_above
2. Numeric Variables: ss
3. Define Values:

   Range: **`3.6`** through 5

D. Output of the three (new) variables.

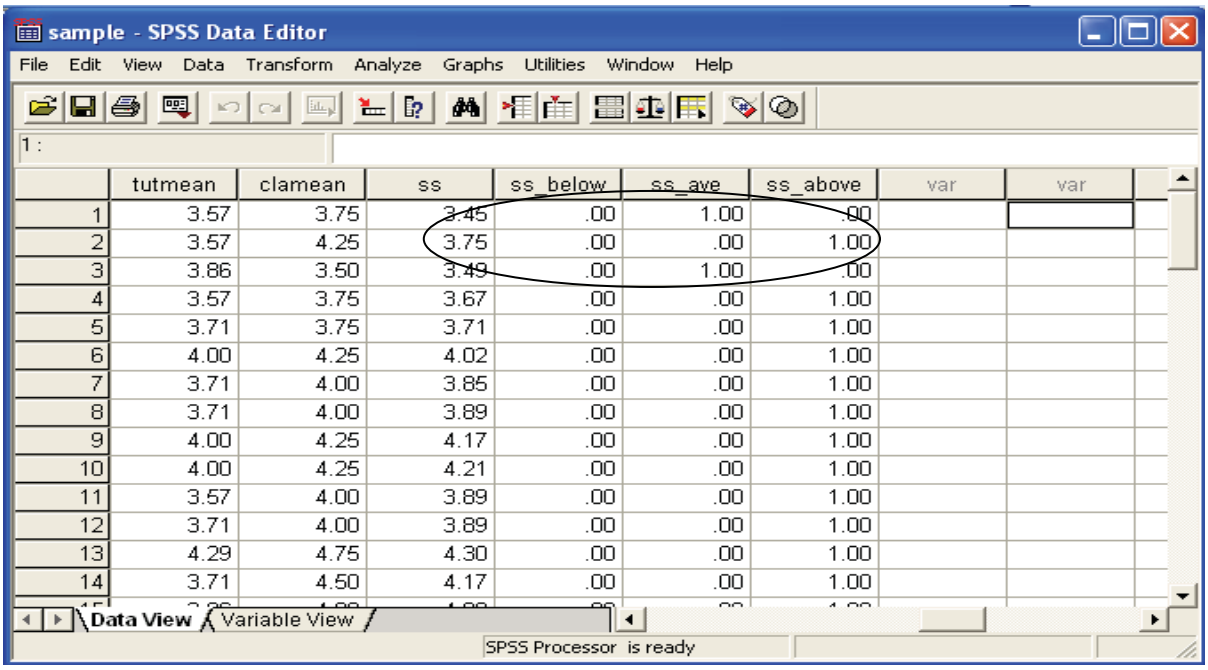


Figure 3.25 Data output in data editor

Figure 3.25 shows how with the use of 'Count Values within Cases' the overall satisfaction can becategorized into different levels.

E. To compute for the `total` frequencies of `ss_below`, `ss_ave`, and `ss_above`:

Analyze → Descriptive Statistics → Frequencies

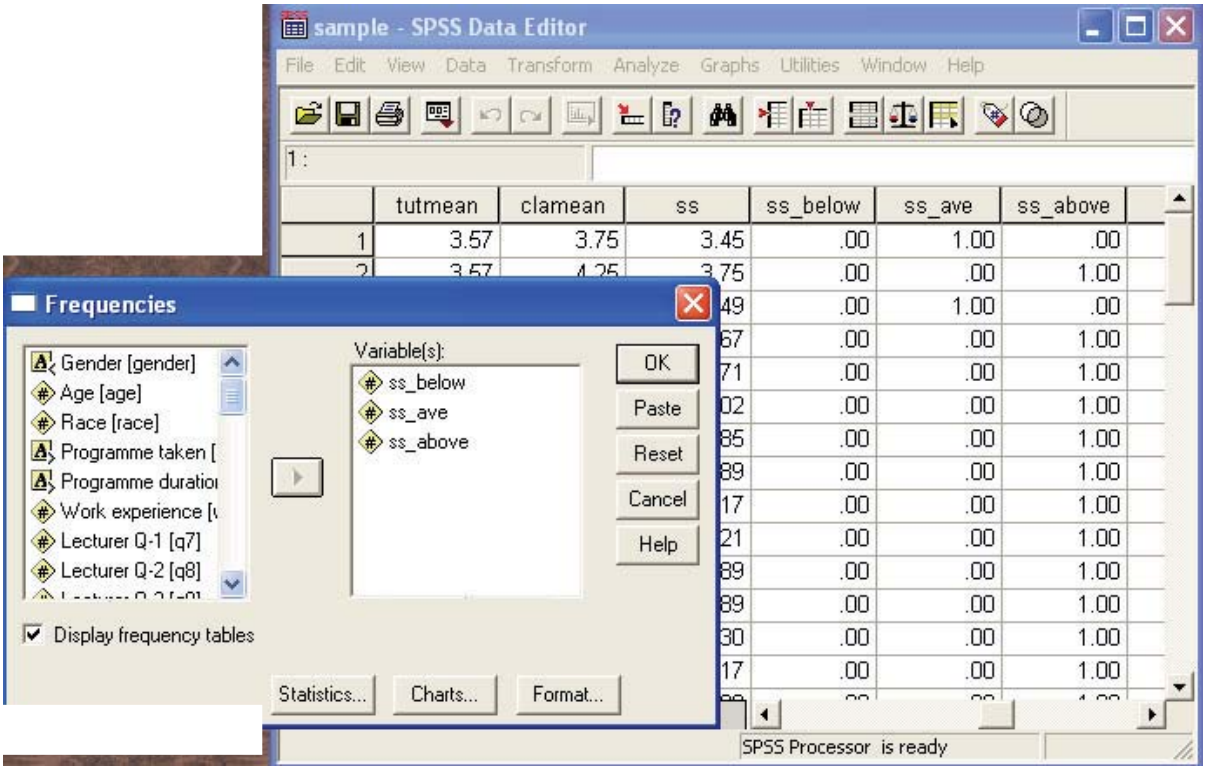


Figure 3.26

F. Output of the breakdown of frequencies.

Frequencies

Statistics

		SS_BELOW	SS_AVE	SS_ABOVE
N	Valid	30	30	30
	Missing	0	0	0
Sum		.00	2.00	28.00

Now we have the frequencies for Below Average as 0, Average is 2 and Above Average is 28. We, fill the table below with these entries.

To fill the individual entries in the table, multiply the frequency with the percentage for each of the entries.

From Summary statistics, we know the percentage of Male (M) and Female (F) respondents. Refer to earlier summary table on Gender for this.

To calculate Students Satisfaction frequencies:

Below Average and Male frequency = 0 x % male = 0 x 73.3% = 0

Average and Male frequency = 2 x % male = 2 x 73.3% = 1 (rounded to whole number)

Above Average and Male frequency = 28 x % male = 28 x 73.3% = 21 (rounded)

Similarly, repeat for all other entries using earlier Summary Statistics for the percentage in each group.

Student Satisfaction	**Frequency**	**Gender**		**Age**			**Work Experience**		**Mean**	**Std dev**
		M	**F**	**<22**	**22-25**	**>25**	**Yes**	**No**		
Below Average	**0**	0								
Average	**2**	1								
Above Average	**28**	2 1								
Total	**30**									

For summaries on other variables (Lecturer, Course Programme, Lecturing and Tutorial Session Classroom Environment) repeat the same procedure as above except replace Students Satisfaction with the variable you are interested.

The complete table format is given next.

	Gender		Age			Work Experience		Mean	Std dev
	M	**F**	**< 22**	**22 – 25**	**> 25**	**Yes**	**No**		
1. Student Satisfaction Below Average									
Average									
Above Average									
Total									
2. Lecturer Below Average									
Average									
Above Average									
Total									
3. Course Programme Below Average									
Average									
Above Average									
Total									
4. Lecturing and Tutorial Session Below Average									
Average									
Above Average									
Total									
5. Classroom Environment Below Average									
Average									
Above Average									
Total									

What can you do with this table?

You can write up on Students' Satisfaction on various sub groups based on the tabulated data.

3.3 Hypotheses of Differences

Hypotheses Test can be done on: Moderating Variables and Dependent Variable

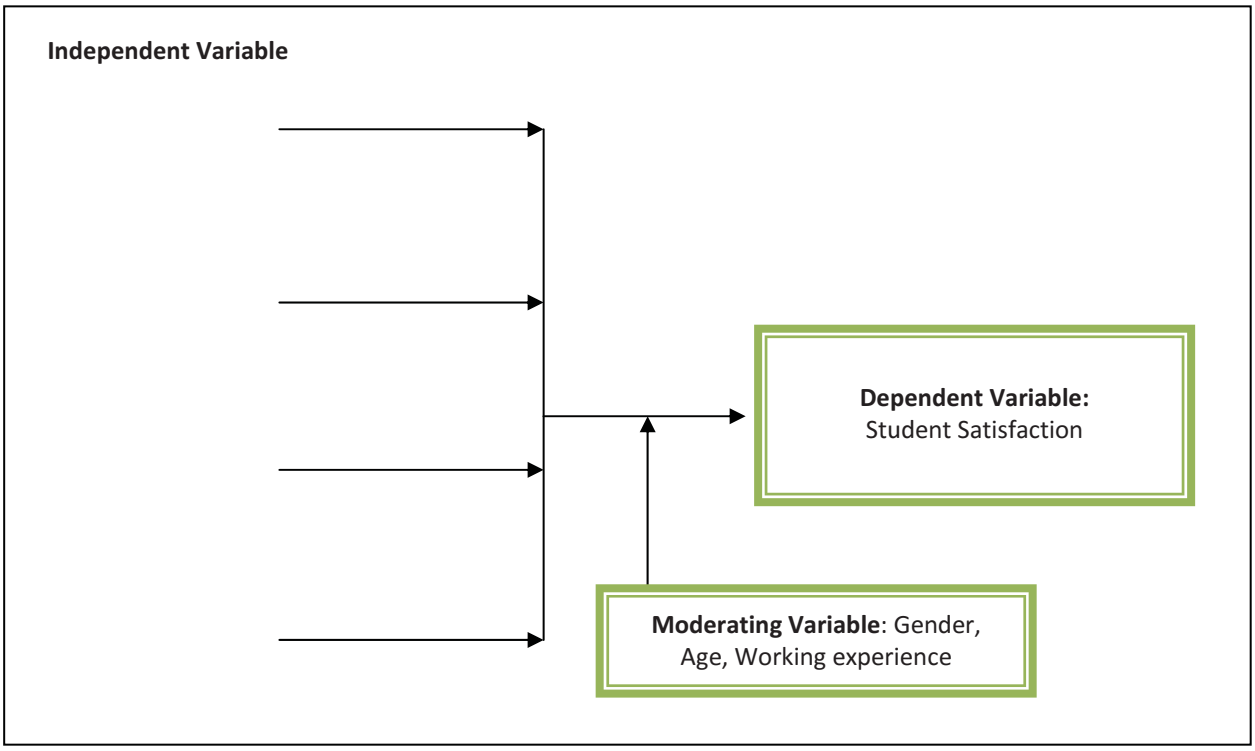
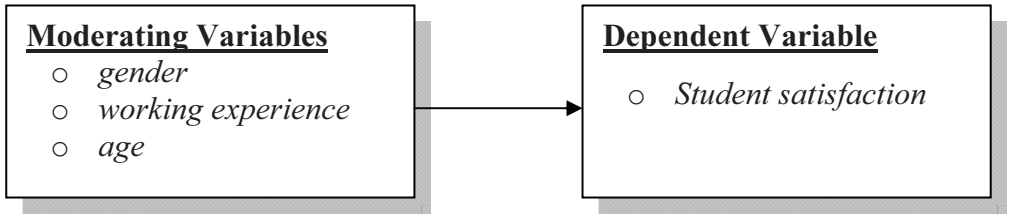


Figure 3.27



Some Research Hypotheses you can raise based on these variables.

Example:

Does Gender have any effects on Students Satisfaction?

Does working experience influence Students Satisfaction?

Does age has a bearing on Students Satisfaction etc.?

1) *Test on Gender and Students Satisfaction.*

H_o: $\mu_1 = \mu_2$ No difference between gender (male or female) and student satisfaction

H_1: $\mu_1 \neq \mu_2$ Significant difference between gender (male or female) and student Satisfaction.

(You can include inequalities here depending on your Hypothesis objectives)

Steps:

a. Analyze → Compare Means → Independent-Samples T Test…

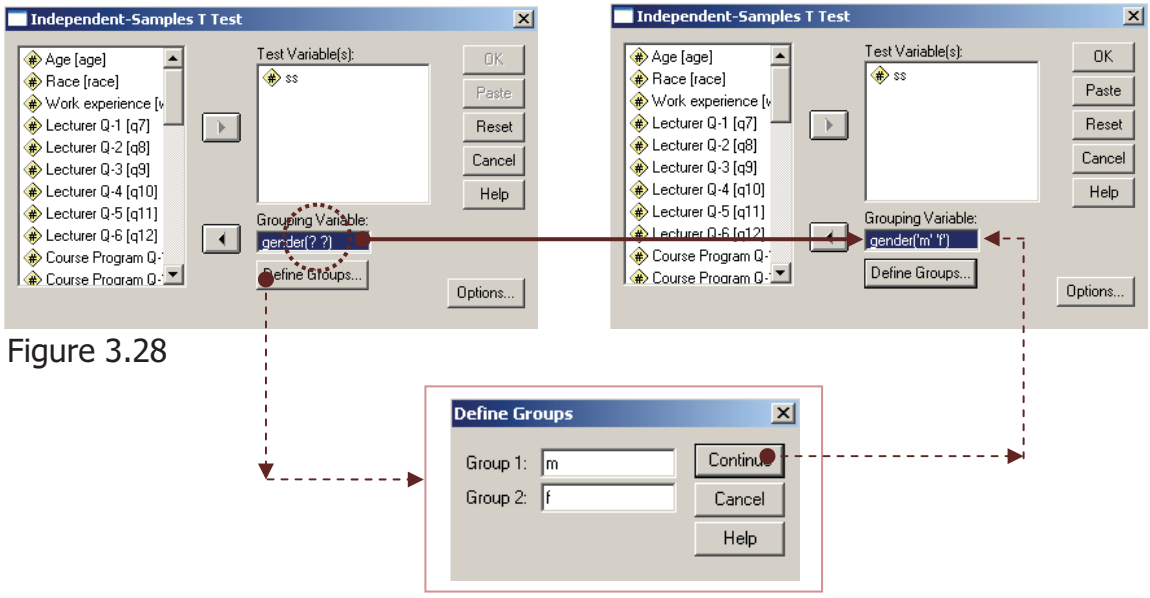


Figure 3.28

b. Output

T-Test

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
SS	male	22	3.9784	.23972	.05111
	female	8	3.9345	.29319	.10366

Significant level for Levene's Test

P-value for equal variances

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
SS	Equal variances assumed	.544	.467	.418	28	.679	.0438	.10493	-.17110	.25876
	Equal variances not assumed			.379	10.608	.712	.0438	.11557	-.21169	.29935

Figure 3.29

Interpretation of the above Hypothesis:

Compare P value (sig 2-tailed, from the above table) with Critical value 5% (95% Confidence Interval).

From the above table P= 0.679 (the value due to the test, look under sig. 2-tailed, equal variances assumed) and this means P>0.005 (the value based on 5% critical value).

Therefore H_1: $\mu_1 \neq \mu_2$ is not substantiated, we fail to reject H_o: $\mu_1 = \mu_2$. That is gender has no effect on students satisfaction at the 95% confidence level.

Assumption of Homogeneity of Variances (Equal variances assumed)

The independent t-test assumes the variances of the two groups we are measuring to be equal. If the variances are unequal then this can affect the Type I error rate, (Note: refer to Appendix VI, page 185). The assumption of homogeneity of variance can be tested using Levene's Test of Equality of Variances, which is produced in SPSS when running the independent t-test.

This test for homogeneity of variance provides an *F* statistic and a significance value (*P*-value). We are primarily concerned with the significance level – if it is greater than 0.05 then our group variances can be treated as equal. However, if P < 0.05, then we have unequal variances and we have violated the assumption of homogeneity of variance. In our test case it is greater than 0.05.

If the Levene's Test for Equality of Variances is statistically significant and, therefore, indicates unequal variances, we can correct for this violation, but an understanding is beyond the scope of this book.

2) *Test on Working Experience and Students Satisfaction.*

H_o: $\mu_1 = \mu_2$ No difference between working experience and student satisfaction

H_1: $\mu_1 \neq \mu_2$ Significant difference between working experience and student satisfaction)

(You can include inequalities here depending on your needs)

Steps:

a. Analyze → Compare Means → Independent-Samples T Test…

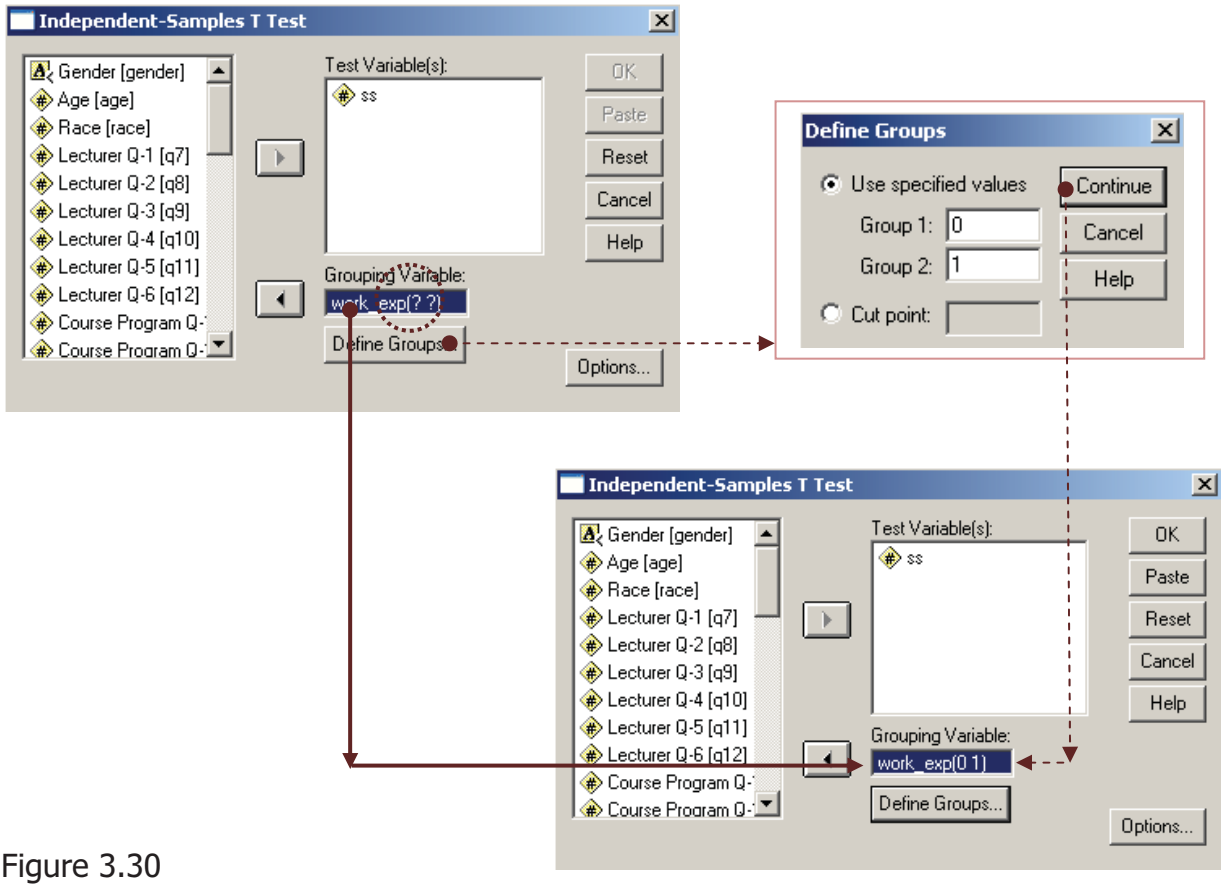


Figure 3.30

b. Output

T-Test

Group Statistics

	Work experience	N	Mean	Std. Deviation	Std. Error Mean
SS	No	24	3.9184	.23966	.04892
	yes	6	4.1597	.20939	.08548

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
SS	Equal variances assumed	.220	.642	-2.254	28	.032	-.2413	.10705	-.46061	-.02203
	Equal variances not assumed			-2.450	8.610	.038	-.2413	.09849	-.46567	-.01697

Figure 3.31

Interpretation of the above Hypothesis:

Compare P value (sig 2-tailed, from the above table) with Critical value 5% (95% Confidence Interval).

From the above table P= 0.032 and this means $P < 0.05$.

Therefore H_1: $\mu_1 \neq \mu_2$ is substantiated, we reject H_o: $\mu_1 = \mu$.That is work experience has effect on students satisfaction.

3) *Test on Age Groups and Students Satisfaction.*

(*ANOVA is useful when you want to test multiple means*)

H_o: No difference between age groups and student satisfaction

H_1: Significant differences between age groups and student satisfaction

Steps:

a. Analyze → Compare Means → One-Way ANOVA …

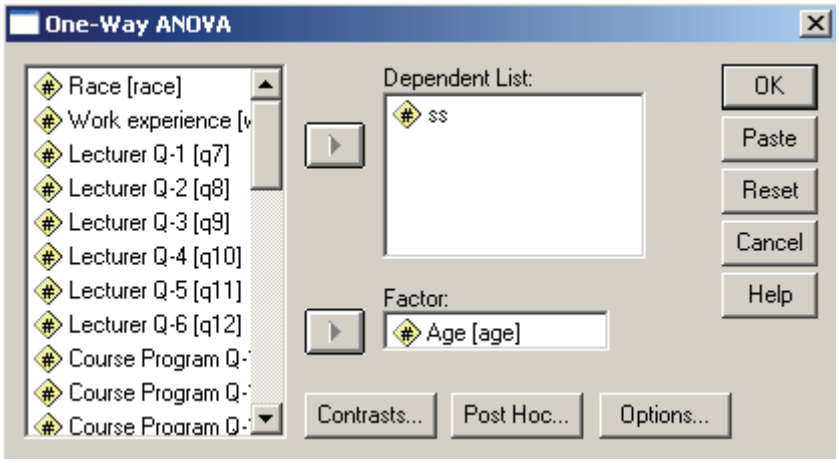


Figure 3.32

b.Output

Oneway

ANOVA

SS

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.365	4	.091	1.569	.214
Within Groups	1.455	25	.058		
Total	1.820	29			

Interpretation:

F value at 1.569 (test value) or the sig. F value at .214 is too small to be significant as compare to critical value for F statistics at given testing level of confidence.

Therefore, we do not reject the null hypothesis H_o: $\mu_1 = \mu_2 = \mu_3$. We, conclude that there are no differences between age groups and student satisfaction level.

Correlation Analysis and Hypotheses

The correlation analysis is used to analyze the relationship between the dependent variable and the independent variables. Various Hypotheses can be generated here. As an example, are lecturers performance and students satisfaction correlated? Use the table on correlation values given below and formulate other combinations for your hypotheses.

1. Click Analyze → Correlate → Bivariate…

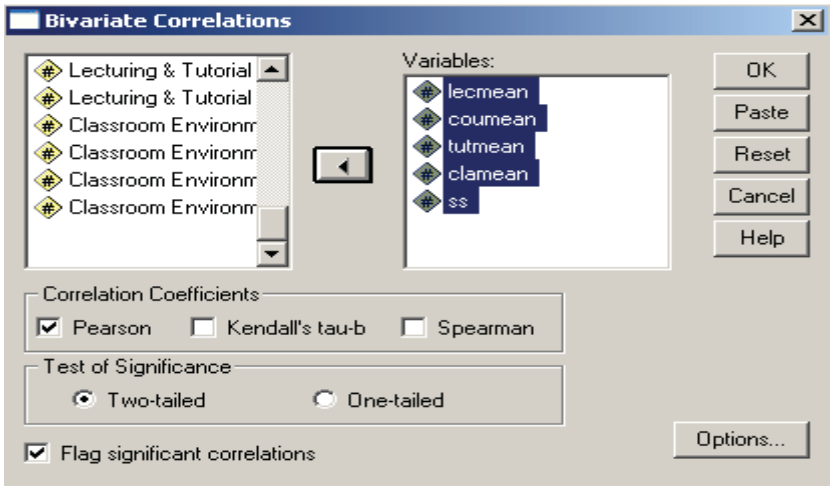


Figure 3.33

2. Output

Correlations

		LECMEAN	COUMEAN	TUTMEAN	CLAMEAN	SS
LECMEAN	Pearson Correlation	1	.701**	.455*	.605**	.880**
	Sig. (2-tailed)	.	.000	.012	.000	.000
	N	30	30	30	30	30
COUMEAN	Pearson Correlation	.701**	1	.478**	.574**	.846**
	Sig. (2-tailed)	.000	.	.008	.001	.000
	N	30	30	30	30	30
TUTMEAN	Pearson Correlation	.455*	.478**	1	.580**	.710**
	Sig. (2-tailed)	.012	.008	.	.001	.000
	N	30	30	30	30	30
CLAMEAN	Pearson Correlation	.605**	.574**	.580**	1	.835**
	Sig. (2-tailed)	.000	.001	.001	.	.000
	N	30	30	30	30	30
SS	Pearson Correlation	.880**	.846**	.710**	.835**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.
	N	30	30	30	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

To read the table, take a variable from horizontal and vertical variables. Run down and along these two variables and at the point of intersection, we obtain the correlation value.

As an illustration, we select SS from the horizontal and LECMEAN from the vertical and they cross at .880 and this is the correlation coefficient between Students Satisfaction and Lecturer Performance.

Correlation can be set Between Independent Variables and Dependent variable

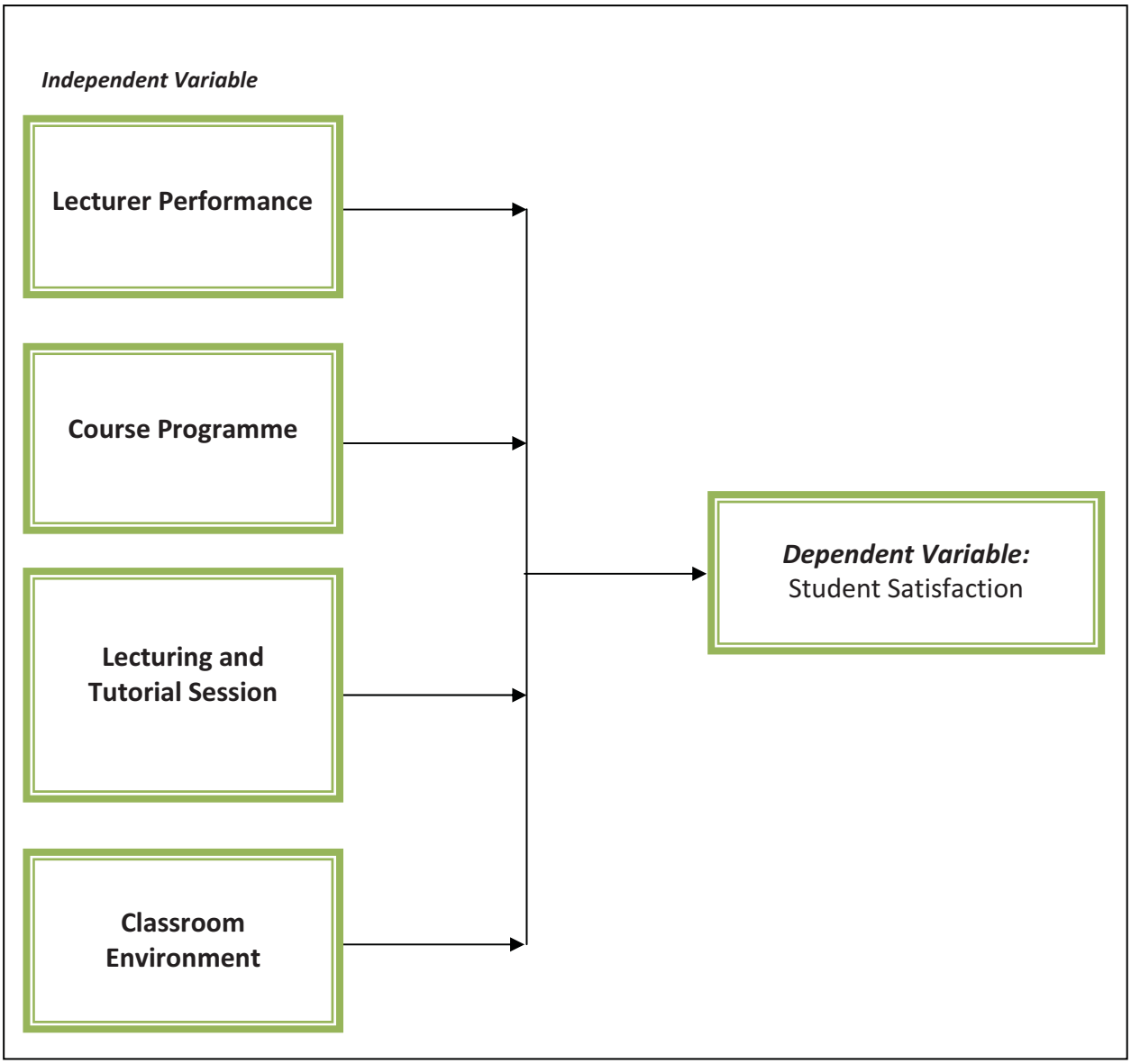


Figure 3.34

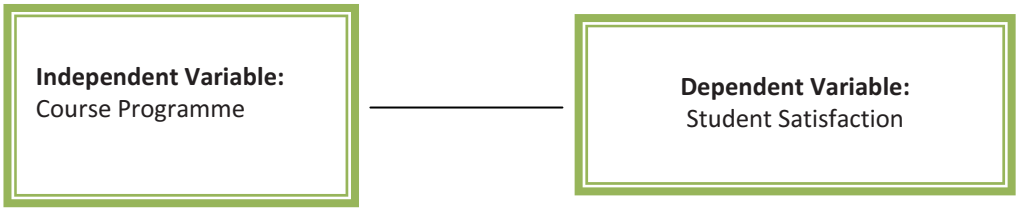
Interpretation of the above table and Hypotheses to be tested:

Illustration on Hypothesis to be tested:

1. ***Hypothesis to be tested:*** Is there any significant relationship between Students Satisfaction and Lecturer Performance?

The reading of .880 between SS (student's satisfaction) and LECMEAN (lecturer performance) shows a very good positive relationship between these two variables (88%).

Explain this observation in details in your research analysis and findings.



2. ***Hypothesis to be tested:*** Is there any significant relationship between Student's Satisfaction and Course Programme?

The reading of .846 between SS (student's satisfaction) and COUMEAN (Course Programme) shows a very good positive relationship between these two variables (84.6%).

Explain this observation in details in your research analysis and findings.

3. ***Hypothesis to be tested:*** Is there any significant relationship between Student's Satisfaction and Lecturing and Tutorial Session?

The reading of .710 between SS (student's satisfaction) and TUTMEAN (lecturing and tutorial session) shows a very good positive relationship between these two variables (71.0%).

Explain this observation in details in your research analysis and findings.

4. ***Hypothesis to be tested:*** Is there any significant relationship between Student's Satisfaction and Classroom Environment?

The reading of .835 between SS (student's satisfaction) and CLASMEAN (Classroom Environment) shows a very good positive relationship between these two variables (83.5%).

Explain this observation in details in your research analysis and findings.

3.4 Other Models

Linear Model

The Model we have used is called a simple linear model where we add the averages of individual objectives to create a grand mean i.e.

Mean of Students' Satisfaction = Mean { Lecturer Performance + Course Programme + Lecturing and Tutorial Session + Classroom Environment}

Replication of Model

This linear model can be replicated to many other similar Research Questions. All you need to do is replace the Dependent Variable with your Research Aim (aim is based on your Research Question) and Independent Variables are replaced with your Research Objectives. Then you can follow steps given in our example.

Other Models

If you want a model other than a Linear Model, you can use similar approach, albeit, with some mathematical treatments and make it into a Linear Model. Just to illustrate,

$$Z = A \times B$$ (Z is dependent variable, A, B are independent variables)

In this model the A values and B values are multiplied and not added.

The mathematical treatment can be in the form of Log:

$$\log Z = \log (A \times B)$$

Simplifies to $\log Z = \log A + \log B$

Here we have converted the model into an additive model with the use of logarithm. For further details refer to mathematical book that has a chapter on linearization.

Note: Refer to instructions in section 2.2.3 pages 91-92.

3.5 Guide to Choosing Correct Hypothesis-testing Procedure

You can follow the guidelines given below in deciding the appropriate test to be applied based on the information given below.

For One Population Mean

If:

Normal Population----------Use one-sample t-test

Normal Population---------and Symmetric Population----------use Wilcoxon signed-rank test

Not Normal Population----Not Symmetric Population----Large sample----use One-sample z-test

For Two Population Means

and Paired Samples

If:

Normal differences-----use Paired t-test

Non Normal differences-----with Symmetrical differences----use Paired W-test

Non Normal differences-----Non Symmetrical differences-----Large sample-----use paired Z-test

For UnPaired Samples

If:

Normal populations------with Equal std dev.-----use Pooled t-test

Normal populations-----Non equal std dev.-------use Nonpooled t-test

Non Normal population-----but Same Shape------use Mann-Whitney test

Non Normal population-----Non Same Shape-----but Large samples------use Two-sample z-test

Note: Refer to Appendix VI, pages 182 for further information on this.

Terms Used in the Book

Abstract – a summary of an article, a book or the whole research content (الملخص)

Analysis – the process of breaking down data to look for patterns and relationships (تحليل البيانات)

Analysis of Variance – a statistical concept used to compare grouped data for differences (تحليل التباين)

Appendix – additional material that a reader is interested in but not included in the article (الملاحق)

Attributes – characteristics of respondents (الخواص)

Assumptions – everything that is assumed, for an investigation to succeed (الافتراضات)

Bibliography – an alphabetical list of all articles consulted and used for the research article (قائمة المراجع)

Boolean logic – the logical sequence of the working of search engine (المنطق البولينى)

Box plot – also called whisker plot, a pictorial way of representing some summary data (طريقة الصندوق لتمثيل العلاقة بين المتغيرات)

Causal relationship – the relationship between variables, where changes in one affects the other (علاقة سببية)

Census – collection and analysis of data for the entire population (إحصاء شامل)

Central tendency measure – summary of statistics that refer to mean, median and mode (مقاييس النزعة المركزية)

Chi-square test – a non parametric statistical test to discover the variance between observed and expected frequency (اختبار مربع كاى)

Cluster sampling – a sampling approach where the sample is grouped into some traits before the sample is selected (العينة العنقودية)

Correlation - the relationship between two variables (الارتباط)

Cover letter – a letter attached to a questionnaire that explains the purpose of the survey (الرسالة التوضيحية)

Critical literature review – an analysis of articles for merits, faults, and the inclusion of own views on the article (مراجعة دقيقة حول ما كتب فى موضوع معين)

Cross tabulation – summary of data in table form involving two or more variables (الجداول المتقاطعة)

Data – statistics, facts, and opinions collected for analysis (البيانات)

Dependent variable – variable that changes in response to changes in other variables (المتغير التابع)

Descriptive research – research that focuses on representation of persons, events, and situations (البحث الوصفى)

Dissertation – research project undertaken by students in bachelor and masters programs, and for an academic audience (الاطروحة)

Exploratory data analysis – data analysis that focuses on diagrams to explore and understand data (تحليل البيانات الاستطلاعي)

Frequency distribution – summary table for data (التوزيع التكرارى)

Gantt chart – a chart of activities and time that helps in time management of project (مخطط غانت)

Hypothesis – testable proposition (statement) on relationship between events and concepts (الفرضية)

Independent variable – variable that causes changes to dependent variable(s) (متغير مستقل)

Likert scale – a numbered scale that allows a respondent to indicate a choice for a given statement (مقياس ليكرت)

Methodology - the theory of how a research is undertaken (منهج البحث)

Nominal data – descriptive data (البيانات الاسمية)

Open question – a question allowing respondents to give their own answers (السؤال المفتوح)

Pilot test – a small scale study of questionnaire responses to minimize answering and recording problems (الاختبار التجريبي)

Population – all members of interest in a research, the complete set (مجتمع البحث)

Preliminary search – a search of literature for research ideas (البحث التمهيدي)

Primary data – data collected specifically for the research (البيانات الأساسية)

Qualitative data – non numerical data (البيانات النوعية)

Questionnaire - a set of questions in a predetermined order used to collect data from respondents who answer the same set of questions (الاستفتاء)

Regression analysis – the method of calculating the regression equation (تحليل الانحدار)

Representative sampling – sample that represents exactly the population (العينات الممثلة)

Research objectives – specific statements that identify what the researcher wishes to accomplish from the research (أهداف البحث)

Research questions – the key questions that the research process will address (اسئلة البحث)

Respondent – the person who answers the questions (المستجيب)

Sample – a subset of the population (عينة)

Sampling frame – the population of interest to the researcher (إطار العينات)

Secondary data – data other than primary and collected by other researchers for their research (البيانات الثانوية)

Simple random sampling – a sampling method that gives an equal chance of being selected into the sample (العينة العشوائية البسيطة)

Stratified random sampling – a sampling method where a population is divided into strata before a random sample is selected (العينة العشوائية الطبقية)

Survey – collection of data from a sizeable population through structured questionnaire (المسح البحثى)

Thesis – the name for research projects at PhD or Mphil levels (الاطروحة/الرسالة)

References

Academic Skills Program. (2010). *Writing a literature Review*. University of Canberra. http://www.canberra.edu.au/studyskills/writing/literature, accessed on 03/11/2010

Barnett, V. (1991). Sample Survey Principles and Method, London, Edward Arnold.

Bell, J. (1993). *Doing your research project: A guide for First-Time researchers in education and social science 2nd Ed.* Buckingham: Open University Press.

Bruce, C. S. (1994). Studies in Higher Education. *Research student's early experiences of the dissertation literature review'* (Vol. 19, No. 2, pp. 217-229).

Cochran, W. G. (1977). *Sampling Techniques, 3nd Ed.* New York: John Wiley and Sons, Inc.

Froese, F. J and S. Xiao. (2012), Work values, job satisfaction and organizational commitment in China, *International Journal of Human Resource Management*, 23, No.9-10, 2144-2162.

Hart, C. (1998). *Doing a literature review.* London: Sage Publication.

Hear, J.M., and Spell. C.S. (2009). How does distributive justice affect work attitudes? The moderating effects of autonomy, *International Journal of Human Resource Management*, 20, 1827-1842.

Neil A. Weiss. (2005). *Introductory Statistics*, 6th Ed., Pearson Addison Wesley.

Plotnick, J. (2002). *Paraphrase and Summary: Writing Workshop.* http://www.utoronto.ca/ucwriting/handouts.html/pdf/paraphrase.pdf, accessed on January 04, 2011

Purdue University Online Writing Lab. (2011). *Paraphrase: Write it in Your Own Words*. http://owl.english.purdue.edu/owl/resource/619/01/ accessed January 06, 2011

Saunders, L. and Saunders, T. (2007). *Research Methods in Business 4th Edition.* Pearson Educational Limited.

Saunders, M., Lewis, P. and Thornhill, A. (2003). *Research Methods for Business Students,* 3rd Edition, Pearson Educational Limited.

Sally Caldwell. (2007). *Statistics Unplugged*, Second Edition., Wadsworth, Cengage Learning.

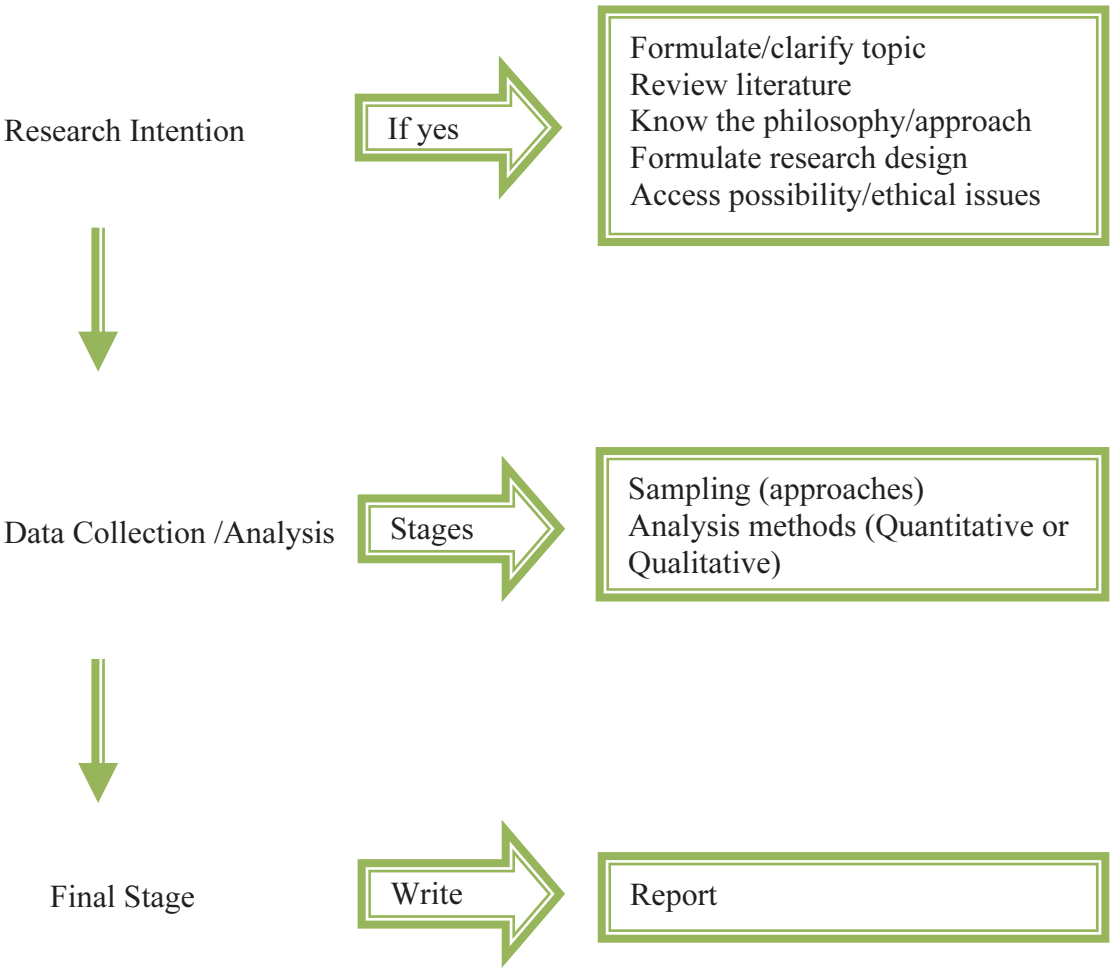
The OWL at Purdue University. (2011). *The Writing Lab*, http://owl.english.purdue.edu/owl/resource/619/01/, accessed on 06/01/2011.

Writing Handouts. (2006). *Steps to Effective Paraphrasing and to Effective Summarizing.* http://www.ndu.edu.lb/academics/fhum/writingcenter/writing.htm, accessed January 06, 2006.

Yamane, T. (1967). *Statistics, An Introductory Analysis, 2nd Ed.* New York: Harper and Row.

APPENDIX I

The Research Process



Adapted from:

Saunders, Lewis and Thornhill

Research Methods in Business

4th Edition, Pearson Educational Limited 2007

APPENDIX II

Contents of Abstract

Put some effort in writing the abstract, it is an important part of your research. Most readers of your research may have time for only that.

Generally, **Abstract** should contain the following items in summary form and at most two pages, but generally 300 to 500 words are sufficient.

a. Purpose/aim

b. Design/methods

c. Findings

d. Limits

e. Practical Importance

f. Values

APPENDIX III

Other Citation Styles

I. APA

The APA citation style was established by the American Psychological Association for publishing written materials in the fields of psychology, education, and other social science.

A. Reference

Book

Sequence:

Author(s)	Date	Title of Book	Place of Publication	Publisher

Example:

Vaswani, V. (2004). *The Complete Reference MySQL: MySQL and PHP*. California USA: McGraw-Hill/Osborne.

Encyclopedia

Sequence:

Author(s)	Year	Topic	Title of Encyclopedia	Volume, Pages	Place of Publication	Publisher

Example:

Lima, M. and Fonseca, N. (2008). Active Queue Management. In *Encyclopedia of Internet Technologies and Applications* (pp. 1-6). USA: IGI Global.

Journal

Sequence:

Author(s)	Date	Title of Article	Name of the Journal	Volume	Pages

Example:

Al-Sharari, S. (2008). *Hajj: The Lifetime Journey. Al-Khafji, 39, 52-65*.

Newspaper

Sequence:

Author(s)	Date	Title of Article	Name of the Newspaper	Pages

Example:

Abu-Nasr, D. (2009, January 31). Festival sees a dazzling array of Arabian horses. *Saudi Gazette*, p.3.

B. Parenthetical Referencing

The sources that you use should be cited in the text of your paper, either in parentheses or as part of the text itself:

> LaGrega et al. (2001) listed the Code of Ethics of Engineers adopted by the Accreditation Board of Engineering and Technology (ABET).
>
> Or
>
> The Code of Ethics of Engineers adopted by the Accreditation Board of Engineering and Technology (ABET) (LaGrega et al., 2001).

The references above refer to the entire source in a general way. If you are referring to a specific part of the source – or quoting exactly – include the specific page number(s) of that part:

(Vaswani, 2004, p. 98)

(Lima & Fonseca, 2008, pp. 126-127)

Sources with three through five authors:

(Conklin, Williams, White, Davis, & Cothren, 2004, p. 300) [for the first time you cite it]

(Conkil et al., 2004, p. 250) [for all other times you cite it]

II. MLA

MLA citation style refers to the rules and conventions established by the Modern Language Association for acknowledging sources used in a research paper.

A. Reference

Book

Sequence:

Author(s)	Title of Book	Publisher	Place of Publication	Year

Example:

Vaswani, Vikram. The Complete Reference MySQL: MySQL and PHP. McGraw-Hill/Osborne: California USA, 2004.

Encyclopedia

Sequence:

Author(s)	Topic	Title of Encyclopedia	Edition	Year

Example:

Lima, Michele and Fonseca, Nelson. "Active Queue Management." Encyclopedia of Internet Technologies and Applications. 1st ed. 2008.

Dictionary

Sequence:

Word(s)	Title of Dictionary	Edition	Place of Publication	Publisher	Year

Example:

"Arab Fund for Economic and Social Development (AFESD)." Dictionary of Travel, Tourism & Hospitality. 3rd ed. UK: Butterworth-Heinemann, 2003.

Journal

Sequence:

Author(s)	Title of Article	Name of the Journal	Volume	Date	Pages

Example:

Al-Sharari, S. "Hajj: The Lifetime Journey." Al-Khafji 39.1 (2008):1.

Newspaper

Sequence:

Author(s)	Title of Article	Name of the Newspaper	Date	Pages

Example:

Abu-Nasr, D. “Festival sees a dazzling array of Arabian horses.” Saudi Gazette 31 January 2009: 3.

B. Parenthetical Referencing

Examples:

In Conklin, et al “Principles of Computer Security,” stated that an interesting paradox when speaking of social engineering attacks is that people are not only the biggest problem and security risk but they are also the best tool in defending against a social engineering attack (69).

The interesting paradox when speaking of social engineering attacks is that people are not only the biggest problem and security risk but they are also the best tool in defending against a social engineering attack (Conklin, et al. 69).

APPENDIX IV

QUESTIONNAIRE

SECTION ONE: PERSONAL DATA

1. Your Gender: □ Male □ Female
2. Your Age □ Less than 22 years
   □ 22 to 25 years
   □ More than 25 years
3. Your Residency □ Saudi
   □ Non-Saudi
4. Name of Program Taken _____
5. Program Duration _____
6. Do you have any past working experience? □ Yes □ No

Strongly Disagree 1	Disagree 2	Neither Agree Nor Disagree 3	Agee 4	Strongly Agree 5

Section Two: Lecturer

	1	2	3	4	5
7. Have your lecturers always take concern about your academic advancement?					
8. I found my lecturers helpful if I encountered difficulties in my subject area.					
9. Have your lecturers marked assignment work fairly and returned it in reasonable time?					
10. Have your lecturers always give encouragement or motivation towards your study?					
11. My lecturers have good knowledge and skills in lecturing their subjects well.					
12. There is ample cooperation and information provided by lecturer in doing my assignment.					

Section Three: Course Programme

My Course Programme:	1	2	3	4	5
13. is very good in helping with my future career.					
14. is essentially long in the programme duration.					
15. is reasonable in term of the course fees.					
16. is able to increase my knowledge and skill.					
17. is widely recognizable in the job market.					
18. is provided with sufficient resources and references in helping my assigned coursework.					
19. is well organized and allocated the appropriate lecturers in the specific subject.					

Section Four: Lecturing and Tutorial Sessions

	1	2	3	4	5
20. I always give response in the lecturing and tutorial sessions.					
21. I have gained good understanding of the concepts covered during lectures.					
22. The lecturing is well organized and very interesting.					
23. I enjoyed the short discussion during lecturing and tutorial sessions.					
24. I feel bored during the lecturing and tutorial sessions.					
25. It is clear what I am expected to learn in each class.					
26. Have your lecturers shown good teaching skills and communicate effectively with the students while lecturing?					

Section Five: Classroom Environment

	1	2	3	4	5
27. The number of students in the classroom has no adverse influence in my studies and concentration.					
28. I have a good relationship with my colleague.					
29. I feel comfortable in my classroom environment.					
30. The classroom is fulfilling with the necessary facilities that are required for lecturing and tutorial purposes.					

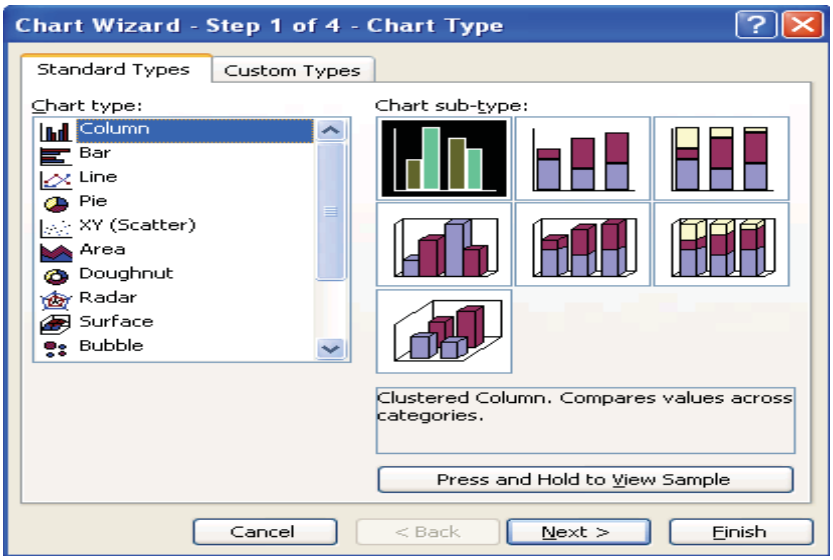
THANK YOU VERY MUCH FOR YOUR TIME AND COOPERATION

Appendix V

MS Excel 2003

Graphs and Diagrams

Open the Excel file, Click on **Insert** icon at the top, select **Chart** for the Chart Wizard. The following window will pop-up:



Examples:

(1) Bar Graph

a. from the Chart Wizard Window, select **Bar** as Chart Type. *Refer to the following figure.*

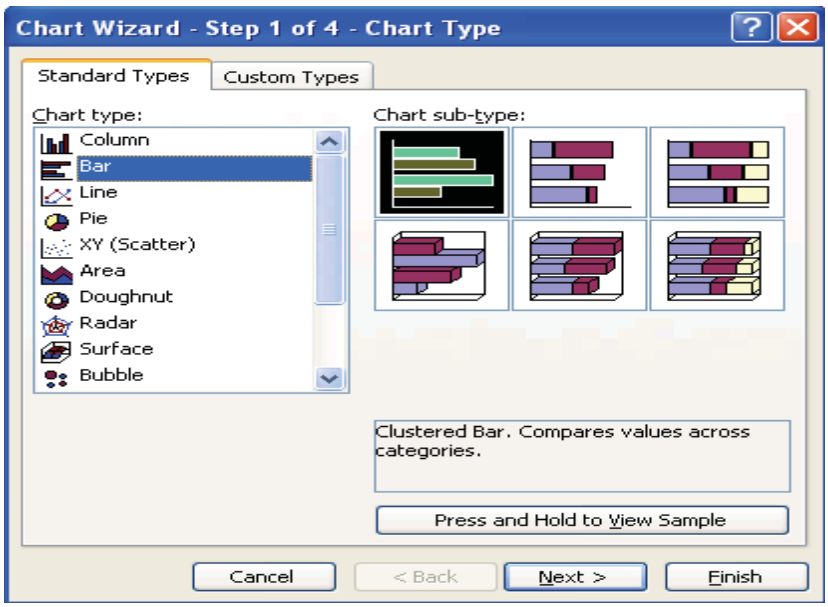


Chart Wizard Step 1 of 4.

b. Select the **first option** in Chart sub-type. Click **Next Button**. *The following window will appear.*

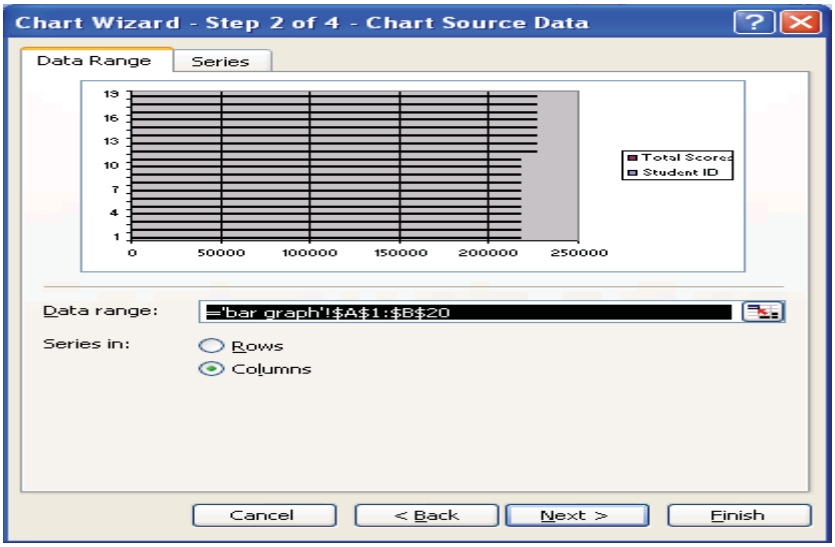
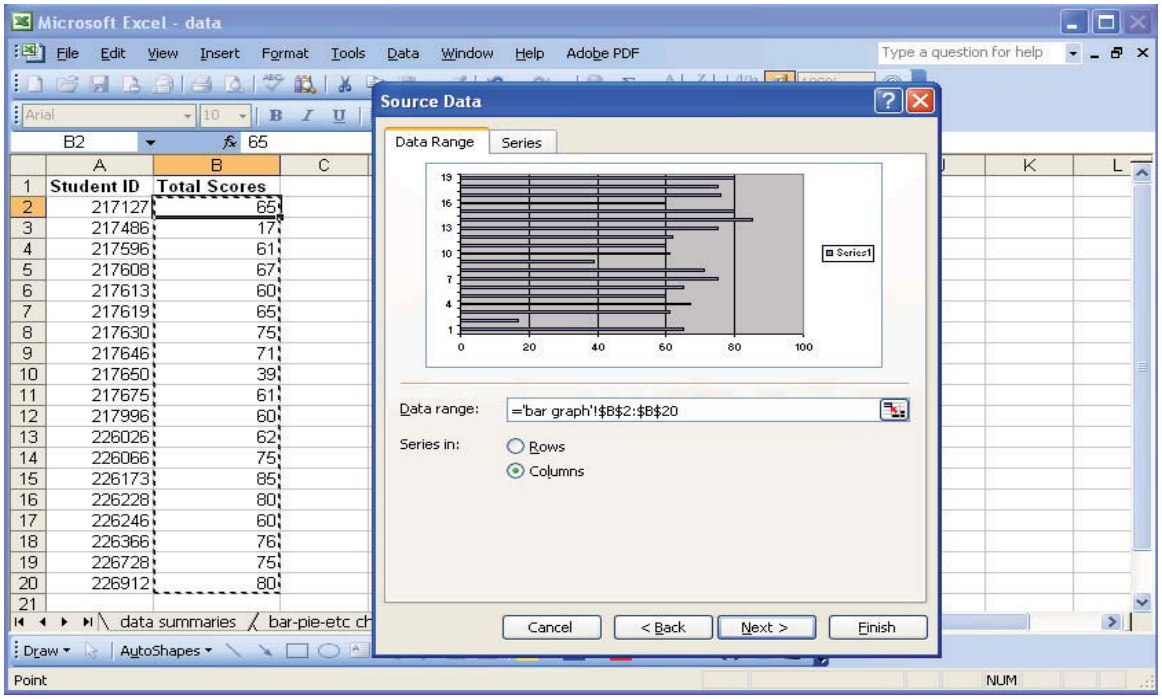


Chart Wizard 2 of 4

From the previous window, the Data range is active; then, highlight the **Data range** from the worksheet as shown below. The Data range will be changed as you selected the data.



c. Now that Data range has been set, the Series in will be the same as Column. Click **Next button**, Step 3 of 4 window will appear.

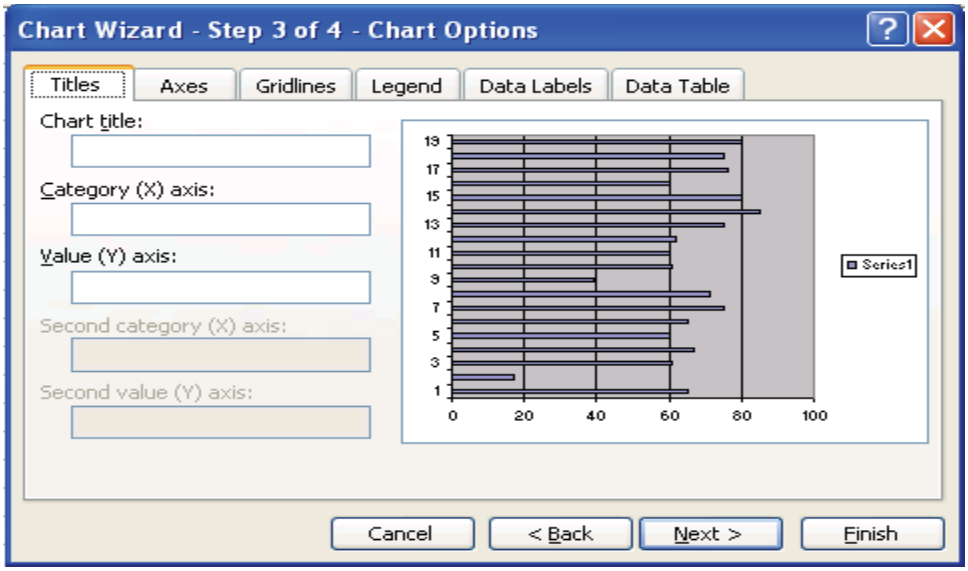


Chart Wizard window step 3 of 4.

d. Check out the **tabs**: Titles, Axes, Gridlines, Legend, Data Labels, and Data Table

- for Chart title type, Student Performance.
- for Category (X) axis, type, Students
- for Value (Y) axis, type, Total Scores

e. The result will look like the following figure. Click **Next** button.

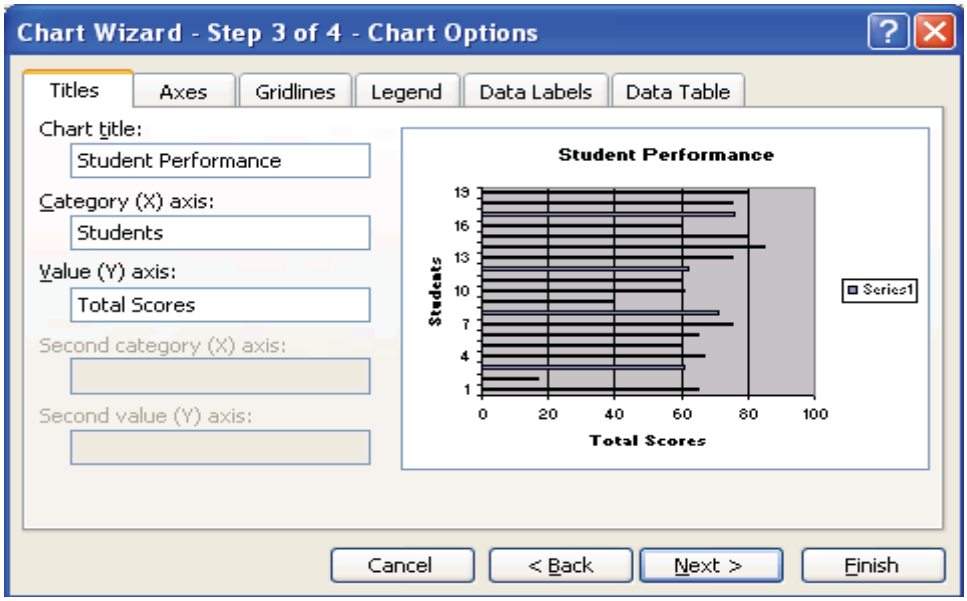


Chart Wizard Step 3 of 4.

f. The following figure will appear from the previous step. This time, no change with the default options.

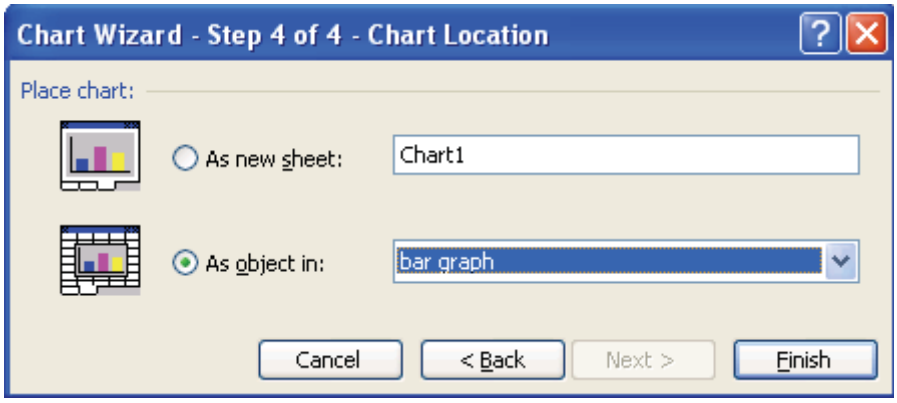
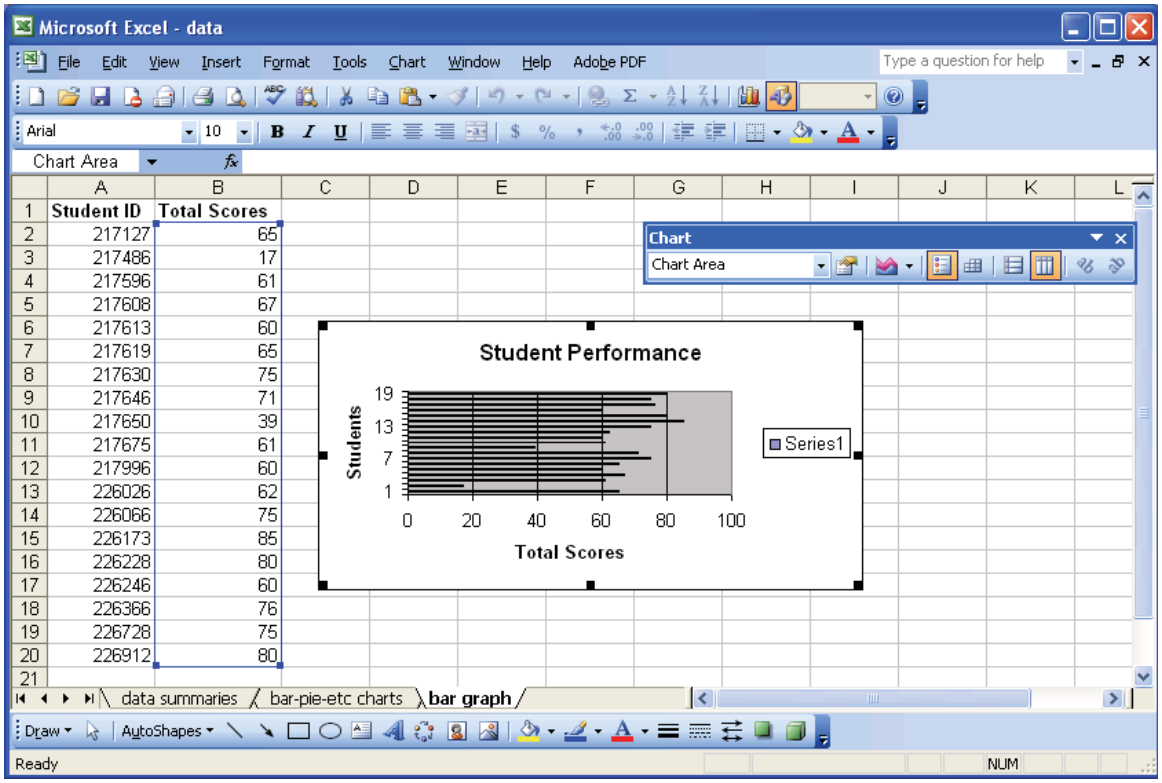


Chart Wizard Step 4 of 4.

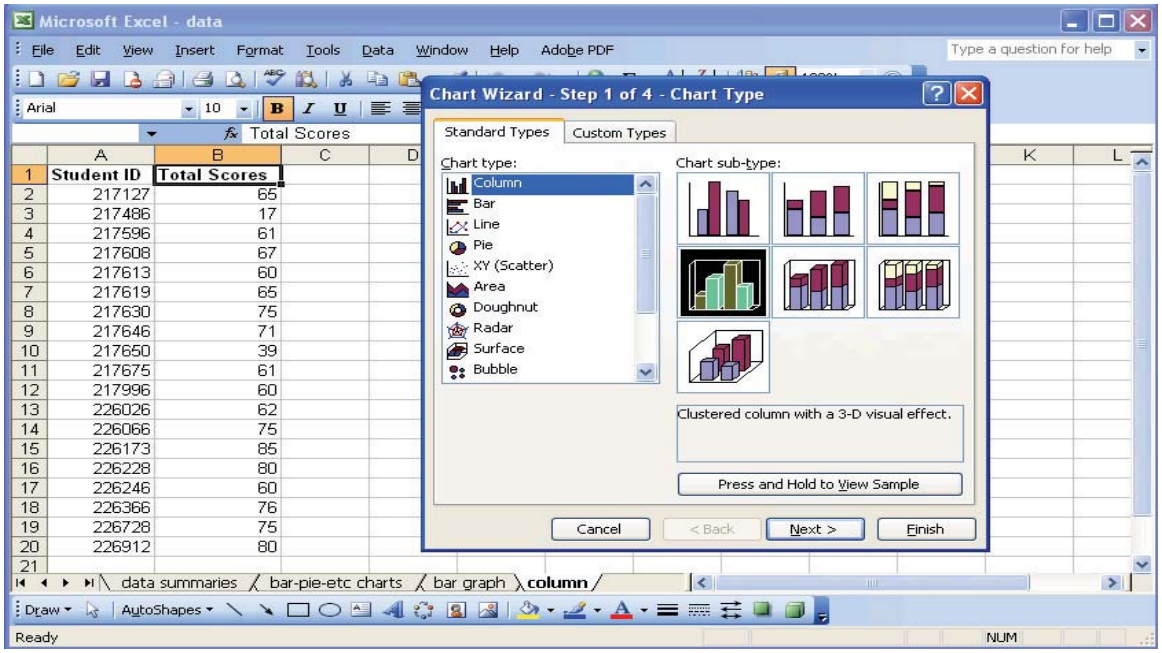
g. Click **Finish** button. The result will look like the following figure:



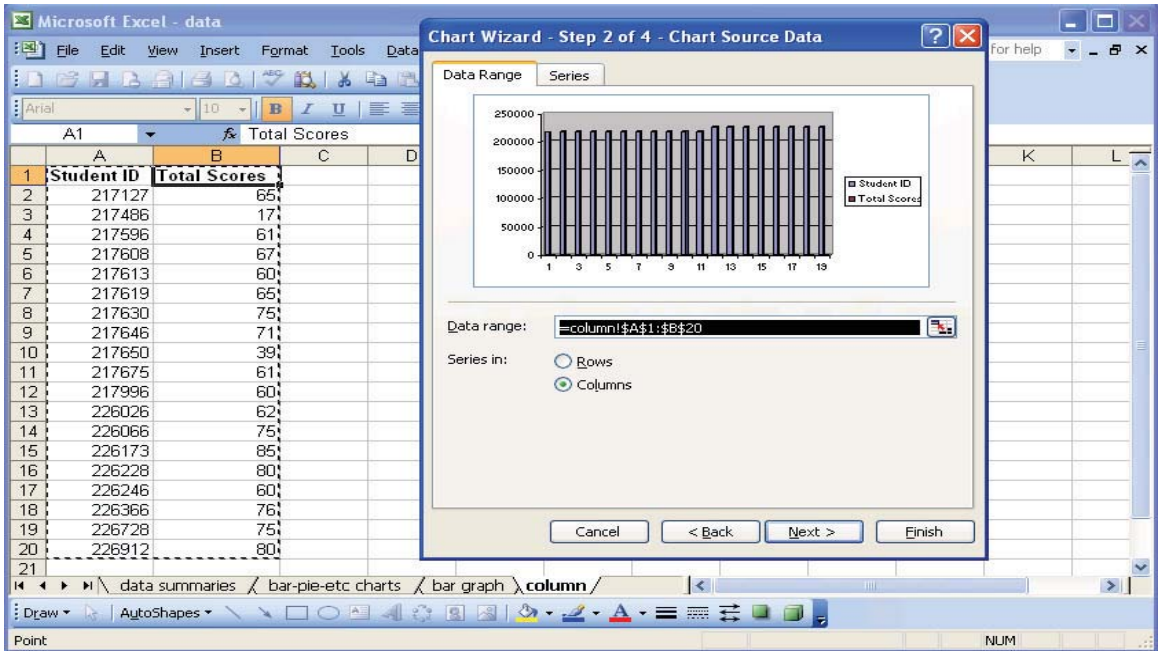
Result of the Chart wizard on Bar Graph

(2) Column Chart Type

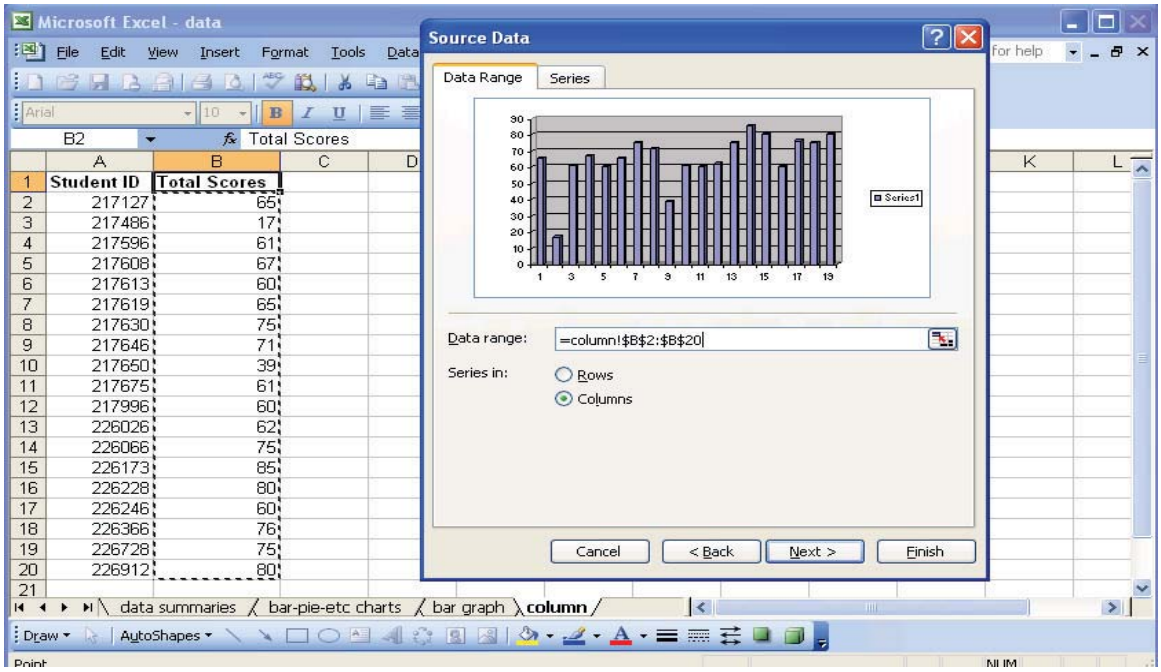
a. Go to Chart Wizard icon on the top menu and select it. Select “Chart type Column” and the following dialogue box will appear. *Using earlier data.*



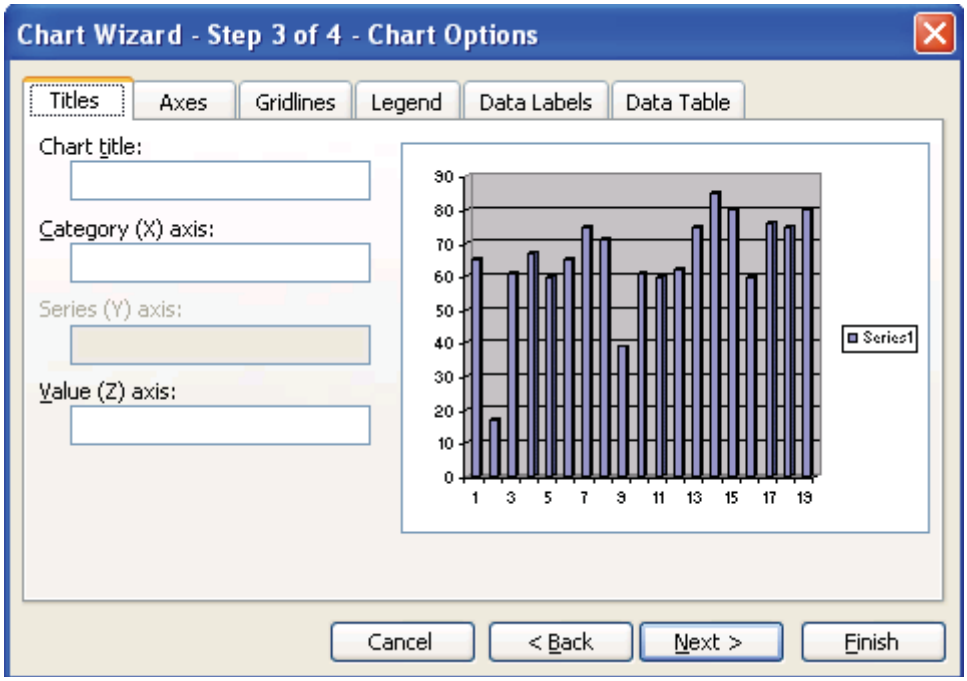
b. Follow the steps. From the above figure, select the Chart sub-type as shown, i.e., clustered column with a 3-D visual effect. Click Next button. The following figure will appear.



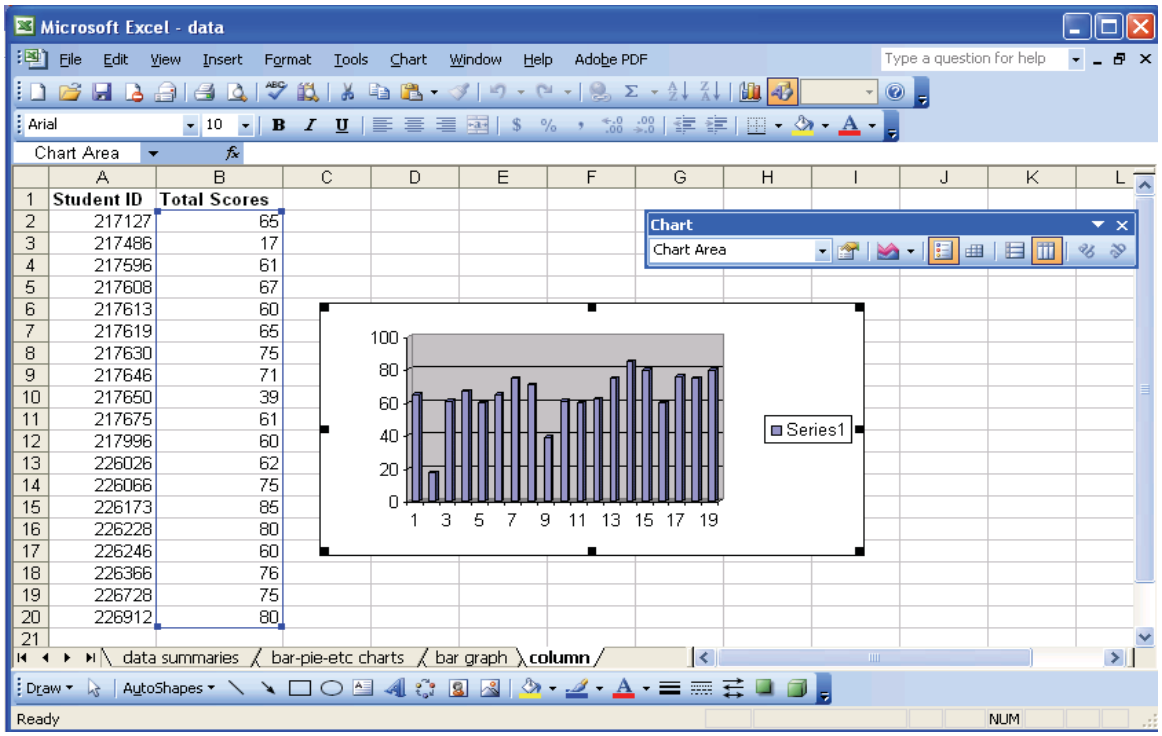
c. To change the Data range of the previous figure, highlight on data range from cells B2 to B20. Notice that the Data range is automatically changed. The result will look like the following figure. Click Next button.



d. The chart can be edited to add titles etc as follows.



e. Click Next button for the final step. Then, Finish button. Result of these steps is as follows:

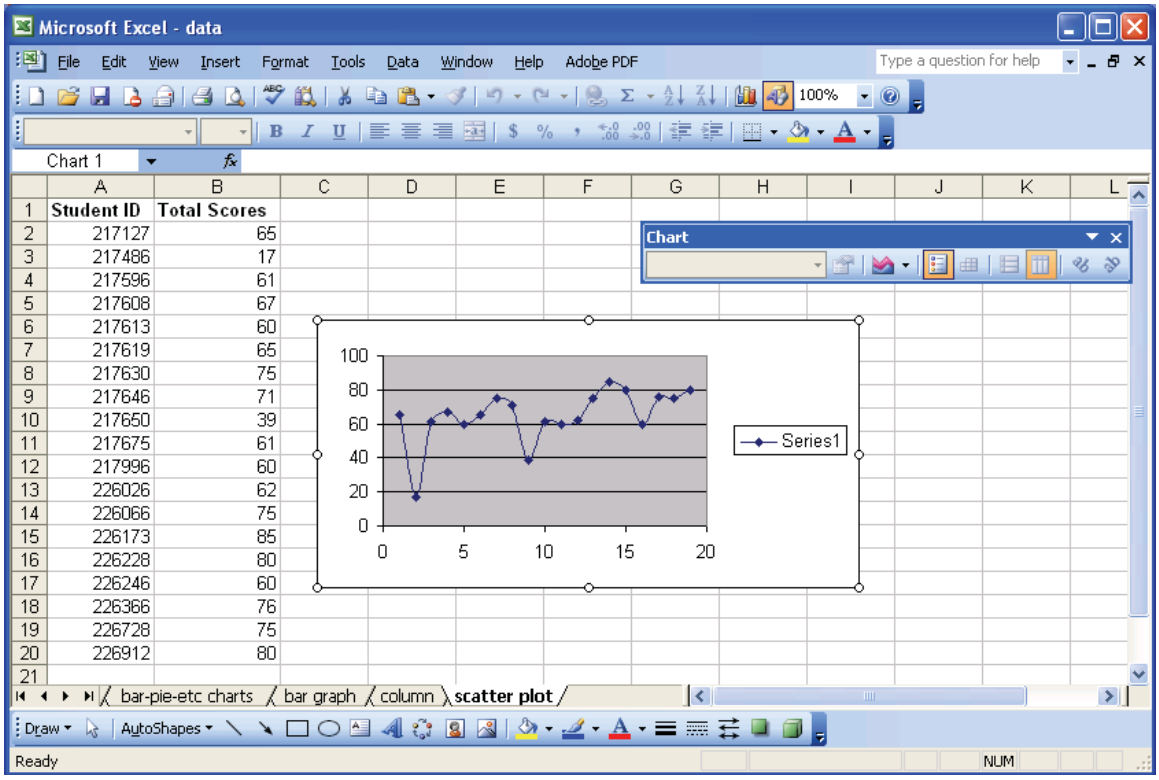


	A	B
1	Student ID	Total Scores
2	217127	65
3	217486	17
4	217596	61
5	217608	67
6	217613	60
7	217619	65
8	217630	75
9	217646	71
10	217650	39
11	217675	61
12	217996	60
13	226026	62
14	226066	75
15	226173	85
16	226228	80
17	226246	60
18	226366	76
19	226728	75
20	226912	80
21		

Note: You can edit the size, design, style, or properties of the Chart by right clicking on the chart itself, then select a particular property you want to modify.

(3) Scatter Plots

For Scatter Plot, select **Chart Wizard** from the icons on the menu bar, select **XY** (Scatter). Follow the same steps as the examples above.



	A	B
1	Student ID	Total Scores
2	217127	65
3	217486	17
4	217596	61
5	217608	67
6	217613	60
7	217619	65
8	217630	75
9	217646	71
10	217650	39
11	217675	61
12	217996	60
13	226026	62
14	226066	75
15	226173	85
16	226228	80
17	226246	60
18	226366	76
19	226728	75
20	226912	80

(4) BOX Plot (also known as Box-and-Whisker plot)

An indirect guide to plotting Box plot:

- For minimum and maximum values:

  Tools → Data Analysis → Descriptive Statistics (same steps as earlier)

  For the Quartiles:

  Insert → Function (click on the menu button on the menu bar.)

  Select Function category: **Statistical**
  For Function name select **QUARTILE**
  Highlight Array to include the Data,
  Use value **1 to give first quartile (Q1), value 2 for second quartile (Q2) (median) and value 3 for third quartile (Q3).**

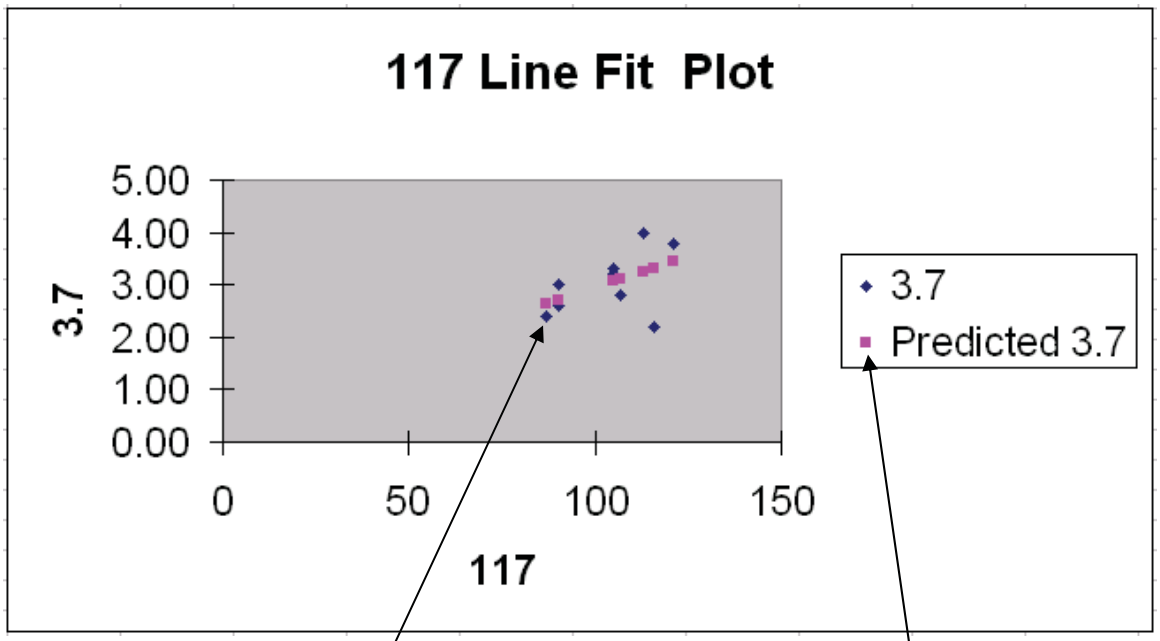
(5) Correlation

For correlation coefficient ‘r’, go to: **Tools → Data analysis → Correlation**

(6) Regression

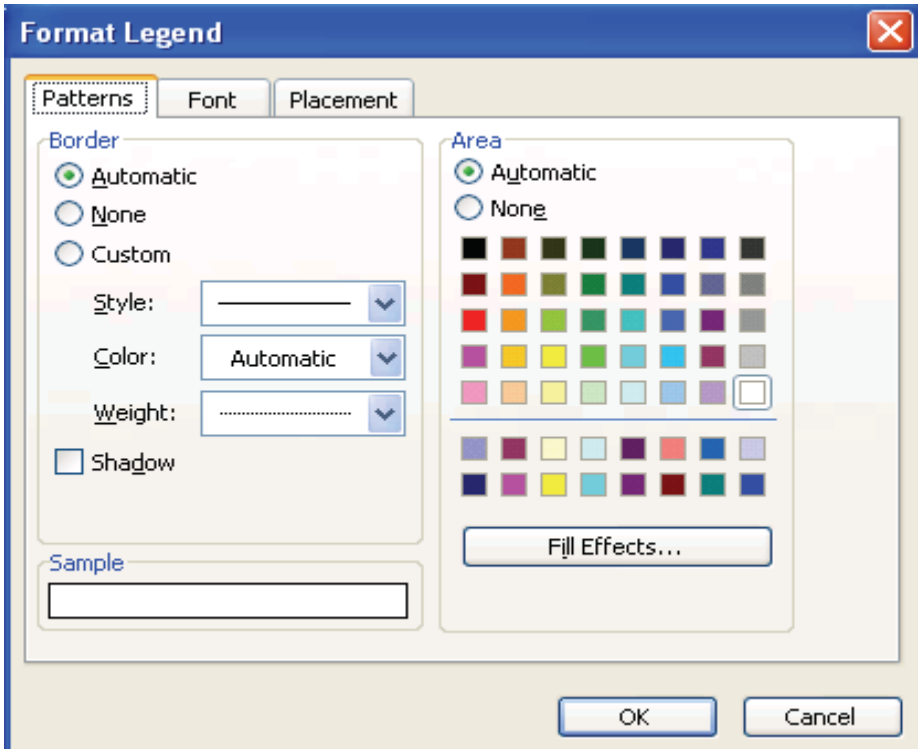
For regression with fitted line, go to: **Tools → Data Analysis → Regression**

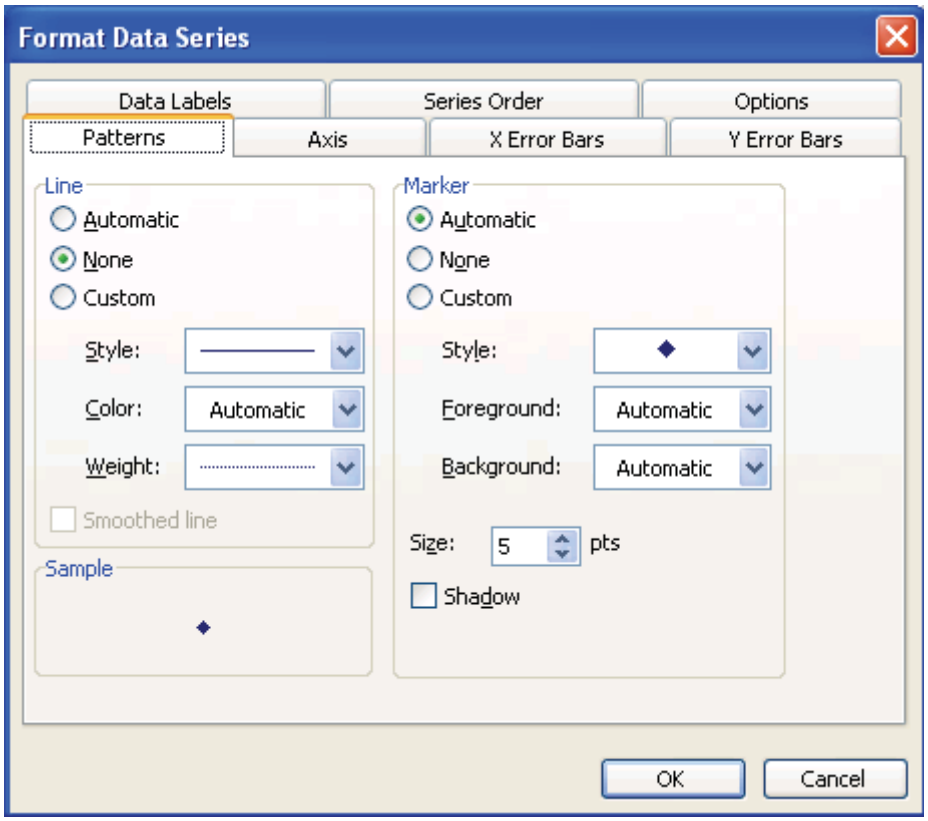
Refine the graph or plot by Edit---Clear---All



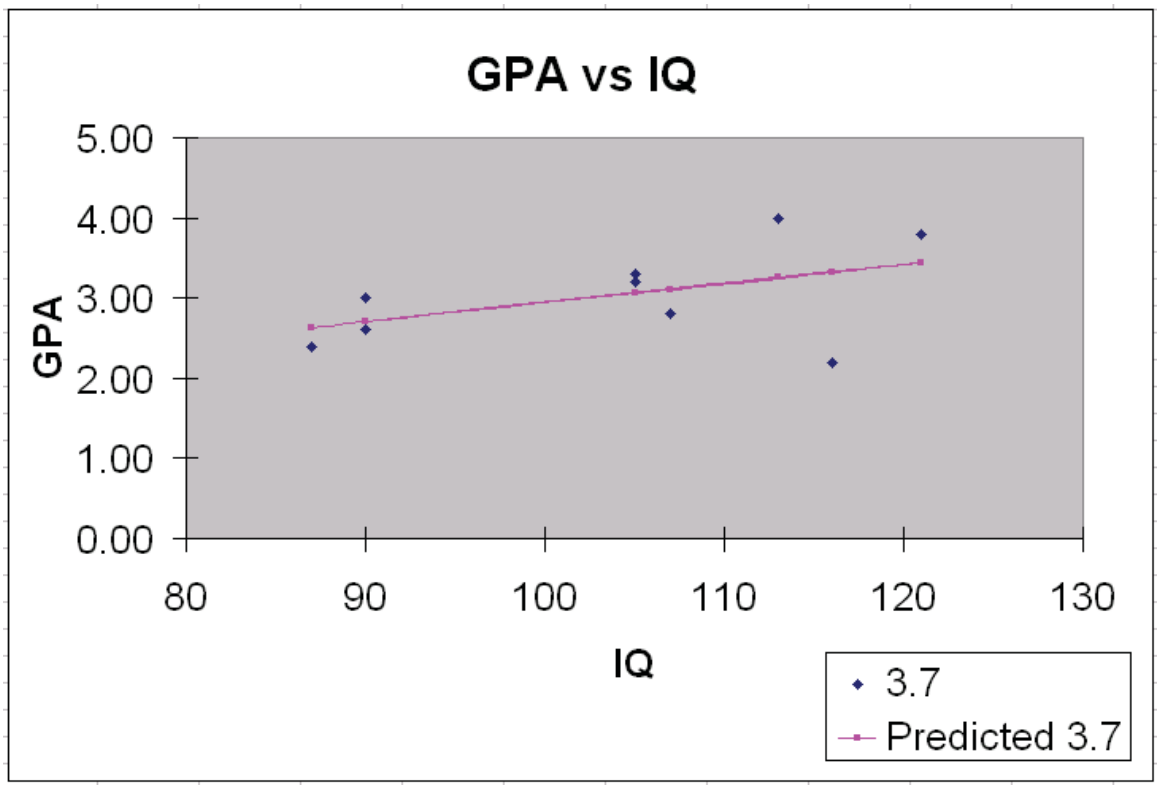
Double click on the plot points (solid square or diamond shape) to edit.

Double click on, square points in the graph to edit the fitted line.


Format Legend
Patterns
Font
Placement
Border
Automatic
None
Custom
Style:
Color:
Automatic
Weight:
Shadow
Sample
Area
Automatic
None
Fill Effects...
OK
Cancel


Format Data Series
Data Labels
Series Order
Options
Patterns
Axis
X Error Bars
Y Error Bars
Line
Automatic
None
Custom
Style:
Color:
Automatic
Weight:
Smoothed line
Sample
Marker
Automatic
None
Custom
Style:
Foreground:
Automatic
Background:
Automatic
Size:
5
pts
Shadow
OK
Cancel

The Output gives the following information.



(7) Regression of More than One Independent Variable (Multiple Regressions)

For regression data, go to: **Tools → Data Analysis → Regression**

(8) Hypothesis Testing

- For tests concerning the difference between *two population m*eans or two *population proportions*:

  Tools → Data Analysis → t-test: Paired Two Sample for Means

- For two sample t-test:

  Tools → Data analysis → t-test: Two sample assuming equal variances

- For *large-sample* test for two means. If sample size is both large ($n \geq 30$) we do not need to assume equal standard deviation:

  Tools → Data analysis → t-test: Two sample assuming unequal variances

(9) Chi-square Distribution and Hypothesis Tests

- A Guide to Finding Critical value and P-value

Insert → Function → Statistical → CHIINV

- If you are keen in finding P-value for a computed Chi-square value, follow the given instructions.

Insert → Function → Statistical → CHIDIST

Appendix VI

Some Common Statistical Terms and Concepts

A. The Use of P-Value in Statistical Testing

1. All tests of statistical significance leads to a probability statement or P value.
2. P value indicates the probability or likelihood of obtaining a result at least as extreme as that observed in a study by chance alone, assuming that there is truly no association.
3. In research a P-value of <=0.05 is considered statistically significant
4. P <= 0.05 indicates that there is, at most, 5 percent or 1 in 20 probability of observing an association as large or larger than that found in the study by chance alone, given that there is really no association.
5. Guidelines for P-values

P-value	Evidence against H_0
P>0.10	Weak or none
0.05<P<=0.10	Moderate
0.01<P<=0.05	Strong
P<=0.01	Very strong

B. Degree of Freedom

Degree of freedom is the number of values in the final calculation of a statistics that are free to vary.

C. One Tail Vs Two Tails P value

A one tail-tailed test will test either if the mean is significantly greater than μ or if the mean is significantly less than μ, but not both.

The one-tailed test provides more power to detect an effect in one direction by not testing the effect in the other direction.

A two-tailed test, regardless of the direction of relationship you hypothesize, you are testing for the possibility of the relationship in both directions.

A two-tailed test will test both if the mean is significantly greater than μ and if the mean significantly less than μ.

D. Testing Hypotheses and Type I and Type II Errors

Decision	H_O True/H_A False	H_O False/H_A True
Accept H_O/ Reject H_A	OK $p= 1- \alpha$	Type II error (β) $p = \beta$
Reject H_O/ Accept H_A	Type I Error (α) $p= \alpha$	OK $p= 1 – \beta$ (statistical power)

Type I Error $\alpha = 0.05$ implies there is only 5 chance in 100 that the result termed 'significant' could occur by chance alone.

The probability of making a Type I error can be decreased by altering the level of significance, the power of a test will be decreased and the risk of a Type II error will be increased.

The probability of making a Type II error can be decreased by increasing the level of significance, it will increase the chance of a Type I error.

The amount of Type I and II errors is decided by you based on willing to either risk.

Statistical Power is the ability of a study to detect a true difference between groups ($1 – \beta$)

E. Guide to Mann-Whitney Test for Two Population Means.

H_O: $\mu_1 = \mu_2$

Construct a Work Table;

Sample from Population 1	Overall rank	Sample from Population 2	Overall rank
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.

Compute the value of the Test Statistics:

M = sum of the ranks for data from Population 1

Compare with the Critical Value, M_l, M_r (for left and right tail) and M_l and M_r (for two tails).

F. Guide to Wilcoxon Signed-rank Test

a. For a Population Mean

H_O: $\mu = \mu_0$

Construct a Work Table;

Data value x	Paired difference $D = x - \mu_0$	$\lvert D \rvert$	Rank of $\lvert D \rvert$	Signed Rank Rank
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

Compute the value of the Test Statistics:

W = sum of Positive ranks

Compare with the Critical Value, W_l, W_r (for left and right tail) and W_l and W_r (for two tails).

b. For Two Population Means

H_O: $\mu_1 = \mu_2$

Construct a Work Table;

Paired difference d	$\lvert d \rvert$	Rank of $\lvert d \rvert$	Signed Rank Rank
.	.	.	.
.	.	.	.
.	.	.	.

Compute the value of the Test Statistics:

W = sum of Positive ranks

Compare with the Critical Value, W_l, W_r (for left and right tail) and W_l and W_r (for two tails).

G. Guide to Kruskal-Wallis Test for k Population Means

Assumptions: Independent samples

Populations have same shape

All sample sizes are 5 or greater

H_O: $\mu_1 = \mu_2 = \mu_3 = \mu_4 \ldots = \mu_k$

Construct a Work Table;

Sample from Population 1	Overall rank	Sample from Population 2	Overall rank	…	Sample from Population k	Overall rank
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

Compute the value of the Test Statistics:

$$H = \frac{12}{n(n+1)} \sum \frac{R_j^2}{n_j} - 3(n+1)$$

(n denotes total pieces of data, $R_1, \ldots R_k$ denotes respectively the sum of the ranks for the sample data from Population 1, 2, …,k).

Compare with the Critical Value, χ_α^2 with df = k-1

H. Test Statistics and Assumptions

a. For Single Population Mean

H_O: $\mu = \mu_0$

Type	Assumption	Test statistics
z-test	Large sample	$z = \frac{\bar{x} - \mu 0}{\sigma / \sqrt{n}}$
t-test	Normal population	$z = \frac{\bar{x} - \mu 0}{s / \sqrt{n}}$, df= n – 1
W-test	Symmetric population	W = sum of positive ranks

b. For Two Population Means

H_O: $\mu_1 = \mu_2$

Type	Assumptions	Test statistics
Two-sample z-test	1. independent samples 2. large samples	$z = \frac{\overline{x1} - \overline{x2}}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$
Pooled t-test	1. independent samples 2. normal populations 3. equal std. deviations	$t = \frac{\overline{x1} - \overline{x2}}{s_p\sqrt{(1/n_1) + (1/n_2)}}$ df= $n_1 + n_2 - 2$ ***note 1***
Nonpooled t-test	1. independent samples 2. normal populations	$t = \frac{\overline{x1} - \overline{x2}}{\sqrt{(s_1^2/n_1) + (s_2^2/n_2)}}$ ***note 2***
Mann-Whitney test	1. independent samples 2. same shape populations 3. $n_1 <= n_2$	M = sum of the ranks for sample data from Population 1
Paired t-test	1. paired samples 2. normal differences	$t = \frac{\bar{d}}{s_d/\sqrt{n}}$ df = n – 1
Paired W-test	1. paired samples 2. symmetric differences	W = sum of positive ranks
Paired z-test	1. paired samples 2. large sample	$z = \frac{\bar{d}}{\sigma_d/\sqrt{n}}$

Note 1: $s_p = \sqrt{\frac{(n1-1)s_1^2 + (n2-1)s_2^2}{n1+n2-2}}$

Note 2: df = $\frac{[(s_1^2/n_1)+(s_2^2/n_2)]\text{^}2}{\frac{(s_1^2/n_1)\text{^}2}{n_1-1} - \frac{(s_2^2/n_2)\text{^}2}{n_2-1}}$

I. Notes on Linear Regression

ANOVA calculations are displayed in an ***analysis of variance table***, which has the following format for simple linear regression:

Source	Degrees of Freedom	Sum of squares	Mean Square	F
Model	1	$\sum (\hat{y}_i - \overline{y})^2$	SSM/DFM	MSM/MSE
Error	$n - 2$	$\sum (y_i - \hat{y}_i)^2$	SSE/DFE	
Total	$n - 1$	$\sum (y_i - \overline{y})^2$	SST/DFT	

The "F" column provides a statistic for testing the hypothesis that $\beta_1 \neq 0$ against the null hypothesis that $\beta_1 = 0$. (β_1 is the gradient of the regression line)

The test statistic is the ratio MSM/MSE, the mean square model term divided by the mean square error term. When the MSM term is large relative to the MSE term, then the ratio is large and there is evidence against the null hypothesis.

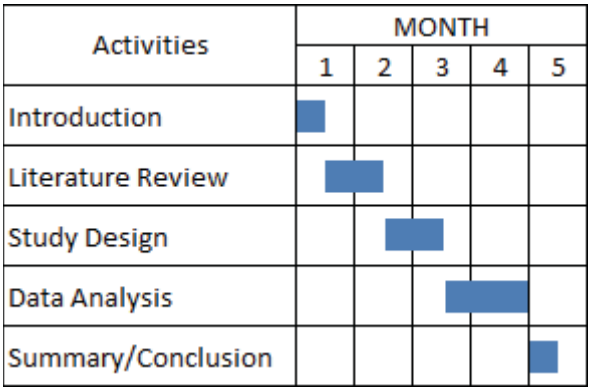
For simple linear regression, the statistic MSM/MSE has an *F* distribution with degrees of freedom (DFM, DFE) = (1, $n - 2$).

Appendix VII

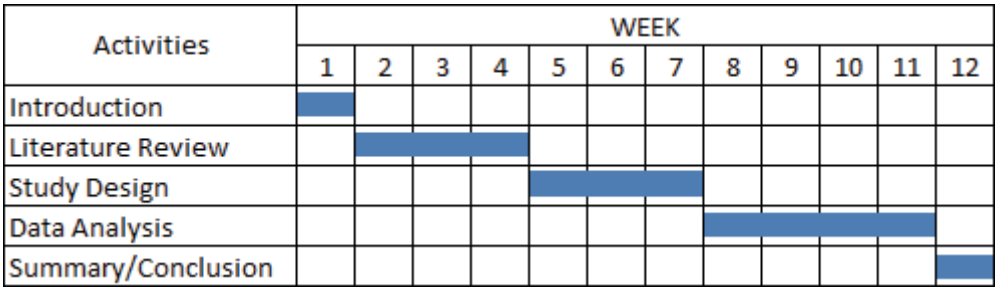
Gantt chart

A Gantt chart is a bar chart that illustrates the start and finish dates of activities of a project. The date is not limited only on the specific date of the day but can also be represented in terms of weeks, months, or quarters. Example of this is:

Choose one from these examples:



Gantt chart Example. The writing of the report is planned to be completed in approximately 4.5 months.



Gantt chart Example. The writing of the report is planned to be completed in approximately 3 months.

Appendix VIII

Searching for Key Word

The two illustrations will guide the reader through searching for key words:

Illustration 1:

- If you are searching for the description of a computer, simply key-in in any web search engines (google.com, yahoo.com, etc) the keyword '*computer*'.
- If you are interested on the examples of operating system, then simply search 'example of operating system'. Take note that it is more specific than searching 'example of computer'.

'computer' → software→ application software →word processor
or → database
'computing' → spreadsheet
→ system software→ operating system→ example
→hardware…
→Programming…
→network…
→science…

Illustration 2:

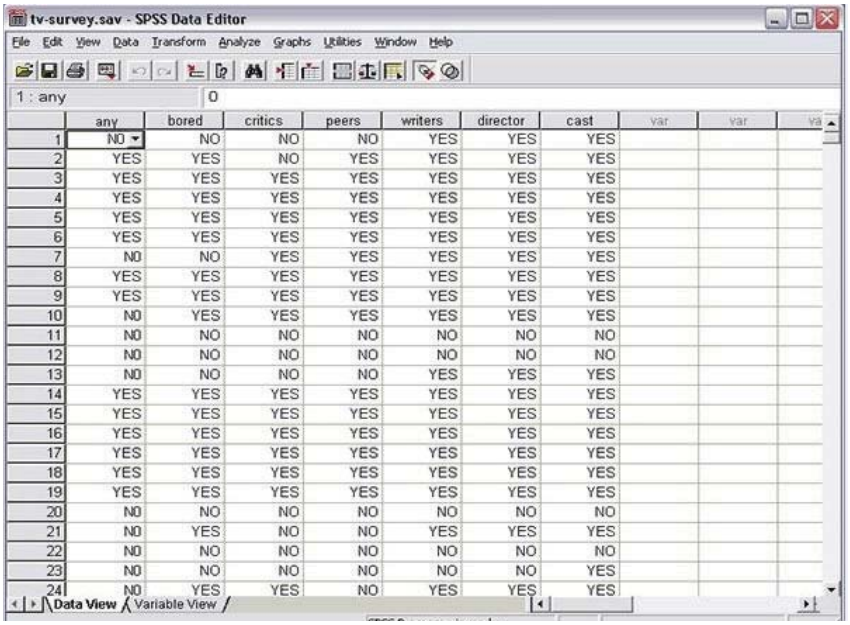
If you are searching for the research areas in management and you have simply web-searched '*areas in management*' then you will get different result if the keyword 'research' is placed before it. Or simply start from searching '*management*' which will provide you information about the description of a 'management'. By this you will get an idea (or a keyword) which area you want to further search. More examples are given below, like '*functions of management*', '*different research areas in human resource management*', etc.

'management' →functions
→roles
→skills
→levels
→research areas
→different research areas
→human resources→Labor satisfaction→Salary
→financial resources
→technological resources
→natural resources

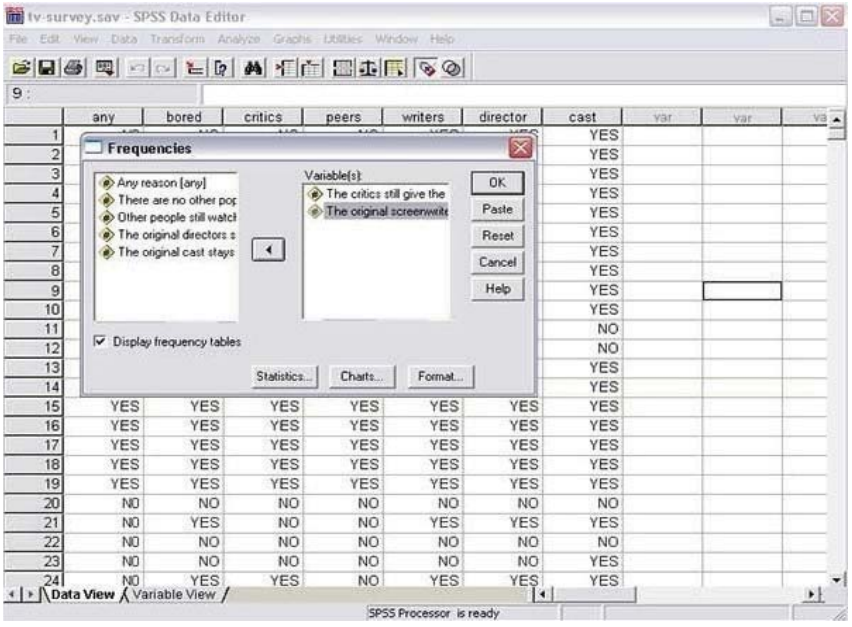
Appendix IX

SPSS Version 20

You will notice there is no difference in the way data is entered from earlier version.



Mmmm



Again you will notice the way to analysis the frequency is the same.

Index

مدينة الملك عبدالعزيز
للعلوم والتقنية KACST

King Abdulaziz City for Science and Technology has worked on providing the knowledge for Arabic readers. It has adopted a project for translating series of scientific books and journals and made them available to the readers free of charge including digital and hard copies .In addition, all kacst publications are available on its website for the browser, to be downloaded or read electronically.

This Book:

Developments in today's times are rooted in research, more specifically dynamic research. Industries allot a budget for researches to help them make sound decisions on economic planning and sustainability. Suffice it to say that without research, the business world will stagger in its direction and will not achieve its long-term targets.

Research writing though is tedious but not impossible. With the right tools to carry it out, a research writer can come out with an output whose results may benefit him personally or contribute to advancements or industrialization. A functional knowledge of research will certainly be necessary for this kind of endeavor. This book which zeroes in on business research is like a treat because it is very much a practical companion in writing a business research. From the rudiments of research writing to presentation of data, the reader will be guided systematically and directed to the most appropriate and feasible approaches and methods in the conduct of business research.

The presentation is a convergence of theory and illustration such that the user-researcher can already see an application of the concepts in the cases and samples provided and can practice on data analysis which is the core of research. This book likewise features the actual use of the SPSS (Statistical Package for the Social Sciences) and various statistical tools using MS Excel.

Overall, this book Business Research Techniques with Statistical Software Usage and Interpretation is a weapon to bridge the gap between necessity of information and development in the 21st century.

Authors:

Dr. Abdulmonem H. Alzalabani: received his MPA from the American University in Cairo, Egypt and Ph.D. in Management from the University of Kent at Canterbury, UK. He then joined Yanbu Industrial College, Saudi Arabia, in 2000, where he became the Head of Industrial Management Technology Department. At present, he is the Deputy Managing Director - Academic Affairs of the College. His research areas include human resource management, organizational behavior, and industrial relations.

Dr. G. Sundaran Sagaran: obtained his higher studies from Australia and Malaysia (PhD in Finance/Management). He has been lecturing in Yanbu Industrial College for the past eight years now.

Mr. Manuel G. Cadongonan Jr: finished his Master in Information Systems and Master in Development Administration from Philippine Universities. Aside from teaching computer applications and Information Systems at the Yanbu Industrial College, Saudi Arabia.

Tel. 011 4883444 - 011 4883555
Fax:011 4883756
P.O. Box 6086 Riyadh 11442
Kingdom of Saudi Arabia
King Abdulaziz City for Science and Technology
Doc. No. 05P042-BOK-0001-AR01

KACST Website: www.kacst.edu.sa
KACST Publications: http://publications.kacst.edu.sa
Email: awareness@kacst.edu.sa

KACST Printing Press
NO: 331203
ISBN: 978 - 603 - 8049 - 54 - 9