Topic 6 ▶ Instrumentation and Data Analysis

LEARNING OUTCOMES

By the end of this topic, you should be able to:

1. Differentiate between instrument and instrumentation;
2. Define what is a questionnaire;
3. Explain the construction of a questionnaire;
4. Design and develop an attitude scale;
5. Analyse data to show significant differences; and
6. Analyse data to show correlation or relationships.

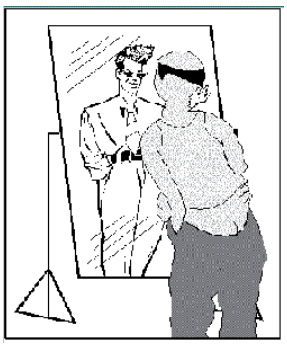
▶ INTRODUCTION

The most important and sometimes difficult aspect of educational research is determining the instruments to be used in data collection (see the *Self-Esteem Test* above which is an example of an instrument). There are numerous tools that may be used to obtain information. In this chapter we will discuss in detail the process of instrumentation and look at some examples of widely used types of instruments. We will also discuss two very important concepts related to instrumentations, namely; reliability and validity. Fraenkel and Wallen (2001) make a distinction between instrument and instrumentation. An instrument is a device or procedure for systematically collecting information while instrumentation includes both the instrument and "the conditions under which it is used, when it is used, and by whom it is used" (p. 81). Examples of instruments are tests, questionnaire, rating scales, inventories and checklists.

Do you believe in yourself?

Are you confident?

Do you like yourself?



Source:
www.cultsock.ndirect.co.uk/
../psy/self_est.html

Self-Esteem Test

This test is designed to test your level of self-esteem. Read each item and state whether you strongly agree, agree, disagree or strongly disagree with each statement by referring to yourself.

	Strongly Agree 4	Agree 3	Disagree 2	Strongly Disagree 1
1. I am an attractive person.				
2. People will dislike me if I am myself.				
3. I am happy with my present height				
4. I like my hair.				
5. I have something to say in social situations				
6. I am looked down because I am not good looking.				
7. I am a friend to myself and take care of myself.				
8. I am popular among my friends.				
9. I like the colour of my skin.				
10. I think I am useless in the company of others.				

Source: Reprinted with permission: © CoPS Project Malaysia (2005). John Arul Phillips

ACTIVITY 6.1

Examine each of the 10 items of the *Self-Esteem Test*.

1. What do you understand by self-esteem?
2. Identify the two aspects of self-esteem being measured.
3. Why do you think some items are stated in the positive while others are stated in the negative?

6.1 THE QUESTIONNAIRE

In chapter 5 we discussed in detail the Survey Method and in conducting a survey the 2nd Step is "writing the items and constructing the questionnaire". When you conduct a survey, a questionnaire is administered in which subjects respond in writing or orally in the case of an interview. The responses of the subjects to the questions are summarised, analysed, and reported. The aim is to obtain information about the characteristics of a particular group. The results from the sample are generalised to a larger group or the population, if the sample surveyed was randomly selected.

Questionnaires are widely used because they are cost effective and can be administered to large groups of people. There is greater standardisation in questionnaires as each respondent receives the same set of questions which allows interpreting from a large sample. The questionnaire if properly designed can address many issues in a relatively efficient way. Also, there is assurance of anonymity which increases the likelihood that the responses are genuine and reflective of the opinions, perception and beliefs of respondents. However, the quality of data obtained from questionnaires will depend on how well it is designed. Fowler (1984) suggests the following important criteria when designing a questionnaire:

- Is this a question that can be asked exactly the way it is written?
- Is this a question that will mean the same thing to everyone?
- Is this a question that people can answer?
- Is this a question that people will be willing to answer?

Source:
www.isothermal.cc.nc.us/admissions.htm

Designing a questionnaire is both an art and science; and takes time and precision. Before setting out to design a questionnaire from scratch, it is a good idea to find out whether such questionnaires already exist that could be used or modified to gather the information that you seek. A little bit of research could save you a lot of work in creating an entirely new questionnaire.

In designing a questionnaire you have to determine the questions that you want to ask people; select the question response format for each question, decide on question sequence and overall layout, pilot-test and implement the instrument. Deciding what questions to ask will depend on the aims of your survey. For example, if the aim of your study is to survey the reading habits and interests of teenagers, one of the questions you will ask is: 'What kind of books do you like to read?' You should constantly check to see that there is close link between the research aims and the individual questions that you ask. You do not want a situation where the data collected from your questionnaire does not answer the research questions.

6.2 QUESTION RESPONSE FORMAT

The questionnaire is designed with the purpose of getting people to respond to a set of questions. In most questionnaires, you will find two broad types of questions used: Structured (closed) and Unstructured (open).

6.2.1 Structured or Closed Questions

Structured questions are questions in which all possible answers (or responses) are identified and the respondent is asked to choose one or more among the answers.

Advantages of Structured or Closed Questions

- Structured or closed questions are useful when the answers are limited and clear-cut. They are most appropriate for asking questions about factual information; but they are also widely used in obtaining data about opinions and attitudes [We will discuss Attitude Scale later].
- Structured questions allow for greater control as they limit the choice of answer and force the respondent to answer.
- It is easier for subjects to respond to structured questions and it also saves time.
- Since it does not require writing, it may encourage unmotivated subjects to respond.
- The limited responses make the task of coding data easier.

Disadvantages of Structured or Closed Questions

- The answers are limited to what is stated which may not capture responses other than those listed.
- Questions that are poorly designed may mislead and frustrate respondents.

6.2.2 Types of Structured Questions

Types of structured questions will be discussed are:

(a) **Dichotomous questions:** These items require two possible responses such as Agree/Disagree or True/False or Yes/No.

Should the driving age be increased from 17 to 21 years?

Yes	
No	

(b) **Check Only One Answer form a List:** The respondent is required to check or tick only ONE answer.

What is your father's highest level of education?

Primary school	
Secondary school	
Form 6 / Matriculation	
Degree	

(c) **Checking More than One Answer from a List:** The respondent is required to place a check or tick next to one or more answer.

Which of the following are reasons for school bullying?

Violence on T.V	
Lack of discipline at home	
Insecure with themselves	
Peer pressure	

(d) **Ranked items:** Respondents can also be asked their preferences by ranking the items from 1, 2 and so forth. We want the respondent to put a 1, 2, 3, 4 or 5 next to the brand, where 1 is the respondent's first choice, 2 the second choice and so forth.
Rank the following brands of cars in terms of prestige:
(1 – most prestigious, 2 – second most prestigious and so forth)

Proton _____
Toyota _____
Honda _____
Nissan _____
Ford _____

(e) **Rated responses:** The Likert Scale has become a popular tool in getting people to respond to questions [*The Self-Esteem Test presented at the beginning of the Chapter consists of items using the Likert Scale*]. The Likert scale is a rating scale which measures the strength of agreement with a statement.

Common types of rating scales:

AGREEMENT				
Strong Disagree 1	Disagree 2	Undecided 3	Agree 4	Strongly Agree 5
FREQUENCY				
Rarely 1	Very rarely 2	Occasionally 3	Frequently 4	Very Frequent 5
IMPORTANCE				
Unimportant 1	Of little importance 2	Moderately important 3	Important 4	Very important 5

(f) Filter questions: This format is used when you want to obtain information from people who 'do' or 'do not' engage in something or belief in something or have an opinion about something (i.e. you are filtering).

(i) Do you smoke cigarettes Yes ☐ No ☐

If *'YES'*, answer Question 2-3. If *'NO'*, continue to Question 4.

(ii) How many sticks of cigarettes do you smoke a day?

< 5 sticks a day		11-15 sticks a day	
6-10 sticks a day		> 15 sticks a day	

(iii) When did you start smoking?

Primary school		Upper secondary school	
Lower secondary school		After secondary school	

6.2.3 Unstructured Questions

In the design of a questionnaire, one issue that will arise is whether you should use open-ended questions. Some researchers are against open-ended questions because of the problems associated with using such a format. However, open-ended questions can be useful. For example, the open-ended question:

> What do you think are the reasons for school bullying?
> What do you like most about the learning materials?

Such questions would provide many kinds of responses. Some would be long answers while others would be just a phrase. If you are interested in getting a variety of reasons and also some in depth reasons, then the unstructured or open-ended format would be useful. Unstructured or open questions should be used when you feel that the particular question cannot be categorised to include all possible responses. However, if you are concerned about the time-consuming task of processing many different responses, than closed questions should be used. But, you should be aware of the disadvantages of such questions.

SELF-CHECK 6.1

1. What are the advantages and disadvantages of using structured or closed questions in a questionnaire?
2. What is the difference between ranked response and rated responses?
3. When would you use a ranked response and a rated response? Give specific examples.

6.3 GUIDELINES ON QUESTIONNAIRE DESIGN

Most problems with a questionnaire can be traced back to the design phase. Your questionnaire would be better designed if you are clear about the purpose of your study. You should clarify your goals in relation to how you intend to use the information. Unfortunately, many researchers neglect this aspect of the questionnaire design process.

(a) **Questions related to objectives:**
Only ask questions that directly address the objectives of your study. For example, if you do not intend to compare differences between ethnic groups, you do not need to ask about ethnic origin. Avoid the temptation to ask questions because it would be "interesting to know" or "maybe I'll need the data later" or "what harm is there in collecting more than I need".

(b) **Length of questionnaire:**
As a general rule long questionnaires get less response that short questionnaire. You should keep the questionnaire short to sustain respondent interest and not exhaust your subjects [you can guess how a person will respond if he or she is tired!]. However, if the questions are interesting, respondents might overlook the length of the questionnaire.

(c) **Instructions:**

(i) Instructions on how to answer the questions should be clear and concise. They should be easy to understand and use words that are not difficult to understand.

(ii) State the purpose of the survey and who is administering or sponsoring the survey.

(iii) Indicate how confidentiality is protected. This assurance may increase the likelihood of honest responses.

(iv) Indicate who the respondent can call or write if they have questions, concerns or want a copy of the survey results.

(d) **Order of questions:**

(i) Begin with questions that you think are interesting to your respondents. For example, questions about their hobbies and interests would attract the interest of adolescents. Also, the questions should not be too threatening. Questions that are boring and threatening will cause respondents not to continue.

(ii) Group questions that are similar together in terms of format and the specific covered (e.g. group question asking respondents about their background such as occupation, income, age, etc.).

(iii) Transition from one question to another should not be too abrupt. Avoid questions that jump from one unrelated topic to another.

(e) **Scales:**

(i) When you want subjects to respond to a scale from lower to higher, it is usually better to place the lower end of the scale on the left and the higher end of the scale to the right.

1. Never 2. Seldom 3. Occasionally 4. Always

(ii) Avoid having too many divisions in your scale which strains the respondent (see example below). Usually a 4 point or 5 point scale would be adequate.

1. Never 2. Seldom 3. Occasionally 4. Fairly often
5. Often 6. Very Often 7. Always

(f) **Wordings of questions:**

(i) Avoid leading questions such as the following:

- **Is it important to treat the handicapped fairly?**
- **Do you agree with most people that the traffic system is getting worse?**

(ii) Watch out for words that are ambiguous or have more than one meaning.
Which teacher is the best?

(iii) Avoid jargon or acronyms or terms that a few people may know. Acronyms should be expanded or stated in full unless the target audience commonly knows them.

- Do you favour inclusion? [Are you sure use most of your subjects know what is 'inclusion'?].
- What are your SPM results? [Foreigners may not know what SPM stands for].

(iv) Use simple and direct language which is easily understood by the respondent. Do not use words that will probably not be understood to avoid misinterpretation. You could bold or underline words that you want to emphasise to eliminate misunderstanding.

(v) Avoid double barrelled questions such as the following:
Do you think Malaysians should eat less and do more exercise?
Break up the question into two: one asking about 'eating less' and the other asking about 'exercise'.

(vi) Avoid questions involving negatives. Do not confuse respondents with questions such as the following:

- Are you against the ban on smoking?
- Do you oppose the ban on public speaking?

(g) Layout and Design:

(i) Do not clutter the questionnaire with unnecessary headings and numbers.

(ii) Avoid using lots of lines, borders and boxes since these can make the page look too 'dense'.

(iii) Ensure that the questionnaire has a title and a brief introductory statement on the purpose of the study.

(iv) Use a good legible font (e.g. Times Roman). Make use of italics and bold types for instructions as well as headings.

SELF-CHECK 6.2

1. What are some common mistakes in the design of questionnaires?
2. What is the purpose of pilot-testing the questionnaire?

6.4 PILOT TESTING THE QUESTIONNAIRE

Test the questionnaire on a small sample of your subjects who are closely similar to your final sample. For example, if you are planning to survey adolescents, make sure that you pilot-test the questionnaire with a group of adolescents. Encourage respondents to make comments on each of the questions, on the order and format of the questions, or on the nature of the questionnaire. You could also have respondents discuss in small groups to provide insights into the questionnaires.

Note whether any questions are frequently unanswered or are frequently misunderstood and answered incorrectly. Probably the questions are poorly worded, they are located wrongly in the questionnaire or they are too difficult to answer or too sensitive to answer. Also, take note of the time taken to complete

the questionnaire. The aim of the pilot-test is to detect any mistakes in your questioning and to correct them before the main survey. The pilot-testing may also allow you to convert an unstructured or open-ended question to a structured or closed question by determining the range of possible answers or responses. Based on the pilot-test, you will be able to make changes that will help maximise response rate and minimise confusion among respondents.

ACTIVITY 6.2

Design a 15 item Teacher Workload Questionnaire focusing on the following:

(a) Planning for teaching

(b) Preparing for assessment and marking papers

(c) Clerical and record keeping

(d) Attending meetings

6.5 DESIGNING AN ATTITUDE TEST

What is an attitude? An attitude is a pattern of belief that is enduring which may influence behaviour. So if someone has a positive attitude about his or her job, you would expect the person to be committed to their job. The Likert scale was introduced by Renis Likert in 1932 as the familiar five-point response format. The scale requests people to indicate how much they agree or disagree or approve or disapprove. An attitude test or inventory is a cluster of items (or questions) that measures a unitary dimension or single attitude. The respondent is provided with a range of possible responses. An appropriate scoring scheme is associated with each of the five possible responses. For example; strongly agree = 5, agree = 4, undecided = 3, disagree – 2 and strongly disagree = 1. This could of course be reversed if desired. Sometimes, if the researcher wants to avoid an undecided category, then they may choose to use an even number of choices, that is; 4 or 6.

Steps in Developing an Attitude Test

1. **Define the Construct:** Define the attitude (or construct) that you wish to measure. You should review the literature in the field to see how the attitude has been defined. Examples of attitudes are; attitudes toward racial integration, attitudes toward women bosses, attitudes toward smoking, attitudes toward water conservation and so forth.

2. **Breakdown the Construct:** Usually an attitude (or construct) needs to broken down into a number of categories. In other words, you are providing an operational definition of the attitude. For example, racial attitudes can be broken down into: marriage, social interaction, cultural heritage, workplace and so forth.

3. **Brainstorm:** Discuss and come up with a list of about 8-10 statements for each category of the attitude. Develop an equal number of positive and negative statements about each category of the attitude object.

4. **Rating the Items:** Next is to have a group of judges or panel of experts, rate each statement on a 5 point rating scale to ensure content validity (or face validity). Do the statements cover the breadth of the attitude being assessed? Measure what it is supposed to measure. For example, is the statement: I would be or have been in a romantic relationship with a person of another race; describing a racial attitude.

   1 = strongly does not describe the attitude

   2 = somewhat describes the attitude

   3 = undecided

   4 = somewhat describes the attitude

   5 = strongly describes the attitude

   The above scale is suggested to guide judges or experts in rating the statements. Note that the judges are not telling you what they believe, but the extent to which they perceive the statements as describing the attitude (or construct) of interest.

5. **Scale:** Decide on an appropriate scale such as the following which has five possible responses: 1 = strongly disagree, 2 = somewhat agree, 3 = undecided, 4 = somewhat agree, and 5 = strongly agree. If you want to avoid an 'undecided' category, you can choose to use an even number of choices, i.e. 4 point Likert scale or a 6 point Likert scale.

6. **Pilot-Testing:** Administer the attitude scale to a group of subjects and score the instrument. Monitor to see if respondents had problems in understanding the items. The final score for the respondent on the scale is the sum of their ratings for all of the items. (this is sometimes called a "summated" scale). On some scales, you will have items that are reversed in meaning from the overall direction of the scale. These are called reversal items. You will need to reverse the response value for each of these items

before summing for the total. That is, if the respondent gave a 1, you make it a 5; if they gave a 2 you make it a 4; 3 = 3; 4 = 2; and, 5 = 1.

7. **Reliability:** You have to examine the test for reliability which means whatever the instrument measures, it measures consistently. Cronbach's alpha measures how well a set of items measures a single construct (or attitude). It measures consistency within the instrument where all items are compared with each other. Alpha coefficient ranges in value from 0 to 1 and the higher the score, the more reliable is the attitude scale. Nunnaly (1978) has indicated 0.7 to be an acceptable reliability coefficient but lower thresholds are sometimes used in the literature.

Item	Correlation with Total	Alpha when Item Deleted
Item 1	0.358869	0.772209
Item 2	0.350085	0.772623
Item 3	0.434180	0.781626
Item 4	0.176243	0.816080
Item 5	0.443533	0.769178
Item 6	0.418211	0.773390
Item 7	0.434247	0.770623
Cronbach Coefficient Alpha:		0.77102

Figure 6.1: Printout of reliability analysis

8. **Reliability Analysis:** See Figure 6.1 which is a printout of the reliability analysis done for a sample of items. The Cronbach alpha for the instrument is 0.77 (rounded from 0.77102). The second column shows the correlation between each item with the total or rest of the items. The third column shows the correlation coefficient obtained if a particular item is deleted. Note that if you were to delete Item 4, the Cronbach alpha will increase to 0.82 (rounded from 0.816080).

9. **Final Items:** Based on the reliability analysis, final items to be included in the instrument can be decided. Generally, the reliability of the scales tends to increase with the number of items. However, as the number of items in a scale increases, so the time taken to complete the attitude test will also increase, and this may demotivate respondents. There is no hard and fast rule to determine the final number of items in a scale and this will reflect the nature and complexity of the attitude being assessed. Generally fewer

than 20 items may reduce reliability acceptably, but more than 30 will begin to demotivate the respondent.

Example of an Attitude Scale:

Attitude towards Sex Education in Schools

1 = Strong disagree 2 = Disagree 3 = Agree 4 = Strongly agree

	1	2	3	4
1. Sex education should be the main responsibility of parents and not teachers.				
2. Rather than have students learn about sex from their peers or on their own, it is better if sex education is taught in schools.				
3. Since some students are likely to experiment with sex, it is a good idea to have sex education taught in the public schools.				
4. Allowing sex education to be taught in the schools would lead to an increase in teenage pregnancy.				
5. Sex education should not be taught in public schools because it will lead to students experimenting with sex earlier than they might otherwise.				
6. The reproductive system taught in science and biology is not sufficient for students to know about sex.				

ACTIVITY 6.2

Referring to the 'Attitude towards Sex Education Scale'.

1. Identify the positive and negative statements.
2. Complete the attitude scale by adding three positive statements and three negative statements.
3. Tryout the complete attitude scale with some of your colleagues and friends.

6.6 QUANTITATIVE DATA ANALYSIS

This section will the quantative data analysis such as *t*-test, ANOVA, chi-square and correlation coefficient to test the different and the relations.

6.6.1 Mean and Standard Deviation

The data collected from questionnaires, attitudes scales, achievement tests and other instruments have to be analysed and interpreted. Certain statistical procedures are used of quantitative data. We will discuss the more common statistical procedures used and we will not go into the mathematics behind the procedures. Emphasis will be on a conceptual understanding of these procedures and they might be used in analysing quantitative data.

(a) **Mean**

The mean is what is commonly called the average. The mean is the sum of all the scores divided by the number of scores. The mean is the most commonly used measures of central tendency or the location of the middle or the centre of a distribution of scores. For example, the mean for the following set of scores: 5, 3, 6, 4 & 7 = 5.

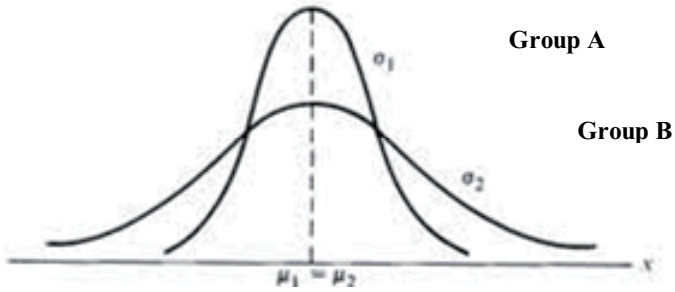


Figure 6.2: Distribution of scores on a geography test

(b) **Standard Deviation**

The standard deviation is a figure that tells you how close the scores are clustered around the mean. Look at Figure 6.1 above which shows the spread of scores obtained on a geography test for two groups of students who obtained the same mean of 20. When the scores are bunched together around the mean, the standard deviation is small and the bell-shaped curve is steep (as in the case for Group A). When the scores are spread away from the mean, the standard deviation is large and the bell curve is relatively flat (as in the case of Group B).

To further understand what does standard deviation mean refer to the graph in Figure 6.2. The mean is 20 and the standard deviation (SD) is 5.

- ONE standard deviation (SD = 5) from the mean in either direction on the horizontal axis accounts for around 68% of the students in this group. In other words, 68% of students obtained between 15 and 25 marks.
- TWO standard deviations (5 + 5 = 10) away from the mean accounts for roughly 95% of students. In other words, 95% of students obtained between 10 and 30 marks.
- THREE standard deviations (5 + 5 + 5 = 15) away from the mean accounts for roughly 99% of students. In other words, 99% of students obtained between 5 and 35 marks.

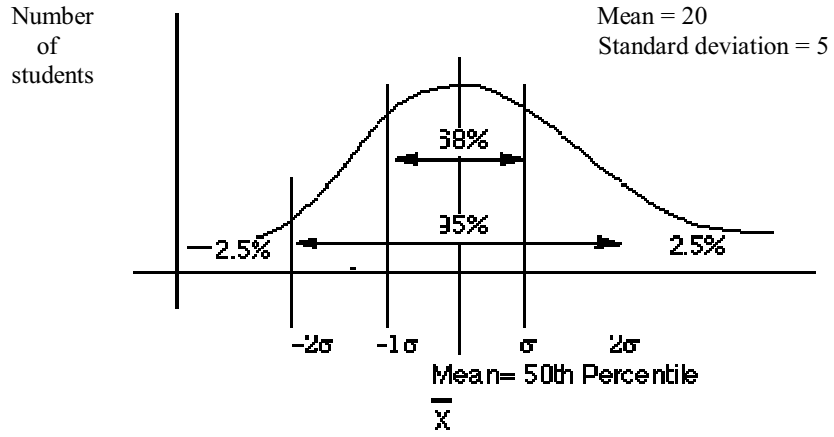


Figure 6.2: Distribution of scores on a geography test

Much of quantitative educational research is focused on answering two important questions.

- Is there a significant difference between variables?
- Is there a relationship (or correlation) between variables?

In the following section we will focus in answering these two questions and the more commonly used statistical tools. We will talk about these tools without going into the mathematics involved. You should be aware that we will discuss the use of these tools at a conceptual level and many issues (such as assumptions) related to each tool will not be mentioned [You are advised to refer to a book on statistics for the details and the mathematics involved].

6.6.2 Testing for Significant Differences between Two Means Using the Student's *t*-Test (Independent groups)

Say for example, you conducted an experiment to compare the effectiveness of the 'discovery method' (independent variable) in enhancing science achievement (dependent variable) among primary school children. The mean scores and standard deviations obtained for the science test is shown in Table 6.1. You want to test the null hypothesis:

Ho: There is no significant difference between the experimental group and the control group in terms of science achievement

Steps:

- You can use a statistical tool called the Student's t-test to obtain the t-value for independent means. ('Independent' means that the two groups consist of different subjects). The t-test gives the probability that the difference between the two means is caused by chance. To test the significance, you need to set a risk level called the alpha level. In most educational research, the "rule of thumb" is to set the alpha level at .05.

  This means that an obtained result that is significant at the .05 level could occur by chance only 5 times in 100 trials.

Table 6.1: Means and Standard Deviations Obtained for the Experimental and Control Groups

	n	Mean	Standard Deviation
Experimental group	10	13.8	2.10
Control group	10	11.4	1.96
t value = 2.65;	degrees of freedom = 18;		$p < 0.02$

- Table 6.1 shows that a t-value of 2.65 was obtained. If you are using a statistical software, the probability level is given (i.e. $p<0.02$). You could also look it up in a table of critical values to find out whether the t-value is large enough to say that the difference between the groups is not likely to have been a chance finding.
- You determine the degrees of freedom (df) for the test which is the sum of the persons in both groups minus 2 (i.e. n-2). Given the alpha level, the df, and the t-value, you can look up the t-value in the table of critical values

(available as an appendix in the back of most statistics texts) to determine whether the t-value is large enough to be significant.

- See Table 6.2. The obtained t-value (2.65) is bigger than the critical value (2.1009) for 18 degree of freedom (20-2 = 18). From this you can conclude that the difference between the means for the two groups is significantly different at the 0.05 level of significance.

Table 6.2: Extract from the Table of Critical Values

df	0.05	0.01
17	2.1098	2.8982
18	2.1009	2.8784
19	2.1009	2.8609

- Note that the difference is NOT SIGNIFICANT at the 0.01 level of significance because the t-value (2.65) is smaller than the critical value (2.8784) for 18 degrees of freedom.

6.6.3 Testing for Significant Differences between Two Means Using the t-test (Dependent groups)

Say for example, you conducted an experiment to compare the effectiveness of the 'discovery method' (independent variable) in enhancing science achievement (dependent variable) among ONE group of primary school children. You gave a pretest and after teaching the students using the discovery method, you gave a posttest. Here, the same group of subjects were tested twice. The mean score and standard deviation obtained for the science pretest and posttest is in Table 6.3. You want to test the null hypothesis:

Ho: There is no significant difference between the pretest mean and the posttest mean in terms of science achievement.

Table 6.3: Means and standard deviations obtained for the pretest and posttest scores

	Mean	**Standard Deviation**
Pretest	9.90	1.66
Posttest	10.90	0.99

n=10 t value = 1.94; degrees of freedom = 9; $p < 0.09$

Steps:

- Using the t-test for dependent groups (also called paired groups or correlated groups) and you obtain a t-value of 1.94. ('Dependent' means that the two means are obtained from the same groups of subjects).The degrees of freedom (df) for the test is the persons in the one group minus 1 (n-1). You set the alpha level at 0.05 and with the t-value you look up the table of critical values
- From Table 6.4 you find that for 9 df the critical value is 2.2622 which is larger than the t-value of 1.94. You conclude that the means is NOT significantly different at the 0.05 level of significance.

Table 6.4: Extract from the Table of Critical Values of t

df	0.05	0.01
8	2.3060	3.3554
9	2.2622	3.2498
10	2.2281	3.1693

ACTIVITY 6.3

Using the *t*-test calculator at this website:
http://www.physics.csbsju.edu/stats/t-test_NROW_form.html **OR**
http://www.physics.csbsju.edu/stats/t-test_bulk_form.html
Key in the mathematics test scores for the two groups shown below and calculate to determine if the means for the two groups are significantly different at the .05 level of significance:

	Mathematics Test Scores
Control group: (n=10)	20, 23, 25, 22, 27, 19, 28, 24, 25 & 26.
Experimental group: (n=10)	29, 30, 28, 32, 33, 37, 27, 36, 31 & 29.

6.6.4 Testing for Differences between Means Using One-Way Analysis of Variance (ANOVA)

In the analysis of variance, the approach is conceptually similar to the t-test, although the method differs. When you want to compare more than two means, the ONEWAY Analysis of Variance (ANOVA) is used. Say, for example you

conducted an experiment in which you compared the effectiveness of three teaching methods in enhancing Reading comprehension. The means and standard deviations obtained are shown in Table 6.5.

Table 6.5: Means and standard deviations obtained for science achievement among students in the three groups

	N	Mean	Standard Deviation
Method 1	10	14.6	1.83
Method 2	10	15.6	2.22
Method 3	10	18.0	2.10

Summary	Sum of square	df	Mean squares	F	p
Treatment	61.066	2	30.533	7.1811	0.003
Within	114.800	27	-		
Total	175.866	29			

Pos Hoc Comparison Using the Tukey Test

Method 1 vs Method 2 - not significant
Method 1 vs Method 3 - significant at p<.01
Method 2 vs Method 3 - significant at p<.05

You used the ONEWAY Anova and obtained an F-value of 7.1811, which is significant at 0.003. Hence, the null hypothesis of no differences between means is rejected. However, you are not sure which differences contribute to the significance. In other words, whether all three methods are different or whether only two methods are significantly different.

So, you have to perform another statistical test called Pos Hoc Tests or Pos Hoc Comparisons. Two common tests used are the Tukey Test and the Scheffe Test which is usually applied after an analysis of variance. The Tukey Test in Table 6.5 shows that there is no significant difference between the performance of subjects taught science using Method 1 and Method 2. Subjects taught with Method 3 performed significantly better than subjects taught with Method 2 at the 0.05 level of significance. Also, subjects taught with Method 3 significantly outperformed subjects taught with Method 1 at the 0.01 level of significance.

6.6.5 Testing Differences Using the Chi-Square (x^2)

The analysis of variance t-test assumes that the population from which you drew the sample is normal. What if you cannot assume that your sample is normal? In this context, the chi-square (pronounced as 'kai' square) represented by the

symbol $\chi 2$ is used. Chi-square is a non-parametric test of statistical significance for analysis of crossbreaks based on frequencies.

Table 6.6: Male and Female Undergraduates Opinions on Violence on TV

Is there too much violence on TV?

	Yes	**No**	**Total**
Males	15	38	53
Females	30	20	50
Total	**45**	**58**	**103**

df = (2-1) x (2-1) = 1 χ_2 = 10.5077 p <.0.01

The chi-square can be used to test whether or not two or more samples are different and can be generalised to the population from which the samples were drawn. The chi-square is widely used in analysing data obtained from surveys using questionnaires. Say for example, you ask a group of 103 undergraduates "Is there too much violence on TV?" The null hypothesis you want to test is:

Ho: There is no significant difference between the opinions of male and female undergraduates regarding violence on TV.

The frequencies of their responses are shown in Table 6.6. The degrees of freedom (df) are obtained by the formula df = (r-1)(c-). That is, the df equals the number of rows in the table minus one multiplied by the number of columns in the table minus one. So the df for Table 6.6 is 1.

Table 6.7: Extract from the Chi-Square Critical Values

df	0.05	0.01
1	3.841	6.634
2	5.991	9.210

The chi-square (x^2) obtained is 10.5077 and with 1 degree of freedom, it is more than the critical value of 3.84 at the 0.05 level of significance and 6.634 at the 0.01 level of significance (see Table 6.7). So, you can reject the null hypothesis and affirm that male and female undergraduate differ in their opinion about violence on TV.

ACTIVITY 6.4

Use the Chi-square calculator at this website:
http://www.georgetown.edu/faculty/ballc/webtools/web_chi.html

Enter the number of rows and columns and press *Generate table.* Enter the following data in the cells for the *Rows* and *Columns* and press *Calculate chi square.* Test the null hypothesis that there is no significant difference in the opinions of final year and first year undergraduates regarding the death sentence.

Should the death sentence be abolished?

	Yes	No
Final Year Undergraduate	54	83
First Year Undergraduates	75	55

6.6.6 Testing Relationship Using the Correlation Coefficient

When you are interested in finding out the relationship or correlation between two sets of variables, the statistic used is the correlation coefficient. For example, in a study you collected data on 'self-esteem' and also 'academic performance'. You want to find out whether there is a correlation between self-esteem and academic performance. A correlation has direction and can be either positive or negative.

- With a positive correlation, individuals who score high (or low) on one measure (or variable) tend to score similarly on the other variable.
- With a negative correlation, individuals who score high on one measure (or variable) tend to score low on the other (or vice versa).

The Pearson product-moment correlation coefficient (represented by 'r') is used to show the degree or strength of relationship between two variables. The coefficient can range from r = + 1.00 to – 1.00. See Figure 6.3 which illustrates what the coefficient means.

- Figure 6.3(a) shows a perfect **positive** correlation (r = + 1.00) which means an increase in variable y is also followed by an increase in variable x.

- Figure 6.3(b) shows a perfect **negative** correlation (r = - 1.00) which means an increase in variable y is followed by a decrease in variable x (or vice versa).

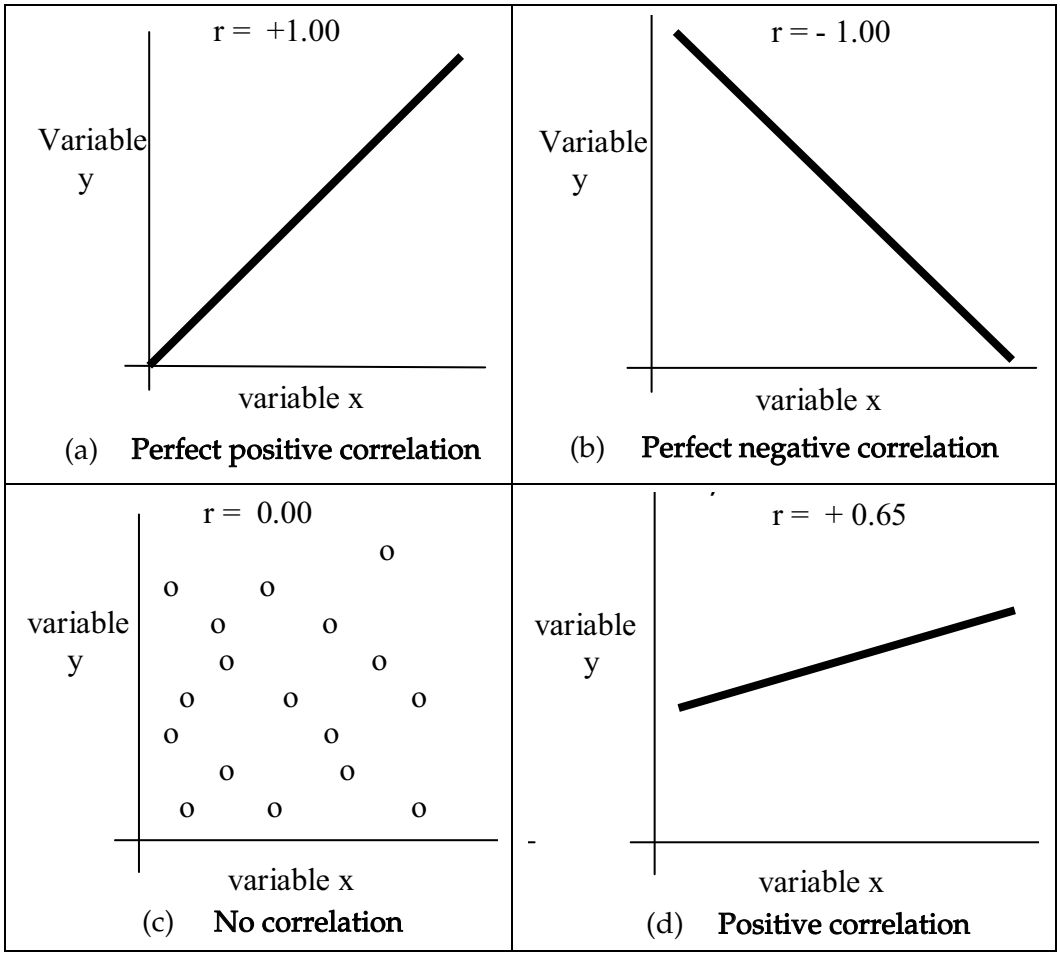


Figure 6.3: Graphs Showing Different Correlation Coefficients

- Figure 6.3(c) shows a zero correlation (r = 0.00) which means there is no relationship between variable y and variable x.
- Figure 6.1(d) shows a positive correlation (r = + 0.65) which means an increase in variable y is also followed by an increase in variable x but not perfectly. For example, if the correlation coefficient between self-esteem and academic performance is 0.65; you can conclude that there is a positive correlation between the two variables. In other words, individuals who have self-esteem also are academically high achievers. However, the relationship is not perfect, indicating that there are individuals who may have high self-esteem but are not academic high achievers and vice-versa. You should be

cautious not to assume that correlation equals causation. In other words, you cannot say that self-esteem 'caused' high academic performance. You can only say that there is a relationship between the two variables.

ACTIVITY 6.5

Use the Correlation calculator at this website:
http://calculators.stat.uda.edu/correlation.php

Enter the following data in the *Input* data column and press *Submit.* What is the correlation between attitude and performance?

Attitude towards mathematics:	23	25	26	12	20	18	21
Mathematics score:	55	50	60	41	49	52	51

SUMMARY

- Questionnaires are widely used because they are cost effective and can be administered to large groups of people. If properly designed, questionnaires can address many issues in a relatively efficient way.
- Questions in a questionnaire can be structured and unstructured.
- The questions in a questionnaire should be clear, not ambiguous and related to the goals of the study.
- Pilot-testing of the questionnaire is essential to detect weaknesses of the items.
- An attitude is a pattern of belief that is enduring which may influence behaviour.
- Attitude scales may be designed to measure attitudes using the Likert scale.
- The t-test is used to test for significant differences between means for independent and dependent groups.
- ONEWAY Anova is used when comparing the means of more than two groups.
- Chi-square is a non-parametric test of statistical significance for analysis of crossbreaks based on frequencies.
- The correlation coefficient is used to test the strength or degree of relationship between two variables.

KEY TERMS

Attitude scale

chi-square

correlation coefficient

ONEWAY Anova

Pilot-testing questionnaire

Questionnaire

Questionnaire response format

Reliability analysis

Structured questions

- dichotomous response
- ranked response
- rated response

t-test

Unstructured questions

DISCUSSION

1. What do you think are some people are *unable* and *unwilling* to respond to questions in a questionnaire?
2. What are some of the common errors you have found in questionnaires?
3. How reliable do you think is the attitude scale in measuring attitudes?

READINGS

OUM Digital Library

Brace, I. (2004). *Questionnaire design.* London: Kogan Page.

Muijs, D. (2004). *Doing quantitative research in education with SPSS.* London: Sage.
Chapter 7: Bivariate analysis: comparing two groups.
Chapter 10: Using analysis of variance to compare more than two groups.

Internet Resources

Reynaldo, J., & Santos, A. (1999). Cronbach's Alpha: A tool for assessing the reliability scales. *Journal of Extension,* 37 (2). Retrieved from: http://www.joe.org/joe/1999april/tt3.html

Siegle, D. (n. d.). *Beginning steps in developing an attitude instrument.*

StatPac. Inc. (n. d.). Questionnaire design: A free web tutorial. Retrieved from: http://www.statpac.com/surveys/