CHAPTER 8 & 9: Measurement, Scaling, Reliability & Validity

===

Any researcher has the opportunity to select a measuring system. Unfortunately, many measurement scales used in business research are not directly comparable. The first question the researcher must ask is “What is to be measured?” This question is not as simple as it first seems. A precise definition of the concept may require a description of how it will be measured, and there is frequently more than one way of measuring a concept. Further, true measurement of concepts requires a process of precisely assigning scores or numbers to the attributes of people or objects. To have precise measurement in business research requires a careful conceptual definition, an operational definition, and a system of consistent rules for assigning numbers or scales.

A. Concepts: Before the measurement process can occur, the researcher has to identify and define the concepts relevant to the problem. A **concept** (or construct) is a generalized idea about a class of objects, attributes, occurrences, or processes. Concepts such as brand loyalty, personality, and so on, present great problems in terms of definition and measurement.

B. Operational definitions: Concepts must be made operational in order to be measured. An **operational definition** gives meaning to a concept by specifying the activities or operations necessary to measure it. It specifies what the investigator must do to measure the concept under investigation. An operational definition tells the investigator “do such-and-such in so-and-so manner.”

C. RULES OF MEASUREMENT

A **rule** is a guide instructing us what to do. An example of a measurement rule might be “assign the numerals 1 through 7 to individuals according to how brand loyal they are. If the individual is an extremely brand loyal individual, assign a 1. If the individual is a total brand switcher with no brand loyalty, assign a 7.” Operational definitions help the researcher specify the rules for assigning numbers.

TYPES OF SCALES

A **scale** may be defined as any series of items that are progressively arranged according to value or magnitude into which an item can be placed according to its quantification. In other words, a scale is a continuous spectrum or series of categories. The purpose of scaling is to represent, usually quantitatively, an item’s, a person’s, or an event’s place in the spectrum.

The four types of scale in business research are as follows:

A. **Nominal scale**: The simplest type of scale. The numbers or letters assigned to objects serve as labels for identification or classification.

B. **Ordinal scale**: This scale arranges objects or alternatives according to their magnitude. For example race horse, we assign a 1 to a win position, a 2 to the place position and a 3 to a show position. A typical ordinal scale in business asks respondents to rate brands, companies, and so on as “excellent,” “good,” “fair,” or “poor.” We know that “excellent” is better than “good,” but we don’t know by how much.

C. **Interval scale**: Interval scales not only indicate order, they measure order (or distance) in units of equal intervals. The location of the zero point is arbitrary. The classic example of an interval scale is the Fahrenheit temperature scale. If the temperature is 80, it cannot be said that it is twice as hot as a 40 temperature.

D. **Ratio scale**: Ratio scales have absolute rather than relative scales. For example, both money and weight are ratio scales because they possess an absolute zero and interval properties. The absolute zero represents a point on the scale where there is an absence of the given attribute. However, for most behavioral business research, interval scales are typically the best measurements.

INDEX MEASURES

This chapter thus far focused on measuring a concept with a single question or a single observation. However, measuring more complex concepts may require more than one question because the concept has several attributes. An **attribute** is a single characteristic or fundamental feature pertaining to an object, person, situation, or issue.

Multi-itemed instruments for measuring a single concept with several attributes are called **index measures**, or **composite measures**. For example, index of social class may be based on three weighted averages: residence, occupation, and residence. Asking different questions in order to measure the same concept provides a more accurate cumulative measure than does a single-item measure.

THREE CRITERIA FOR GOOD MEASUREMENT

There are three major criteria for evaluating measurements:

A. **Reliability**: Reliability applies to a measure when similar results are obtained over time and across situations. It is the degree to which measures are free from random error and, therefore, yield consistent results.

There are two dimensions of reliability; **repeatability** and **internal consistency**. Assessing the repeatability of a measure is the first aspect of reliability. The **test-retest method** involves administering the same scale or measurement to the same respondents at two separate points in time to test for stability. If the measure is stable over time, the repeated test administered under similar conditions should obtain similar results. High stability correlation or consistency between the two measures at time one and time two, indicates a high degree of reliability. There are two problems with measures of test-retest reliability; first, the first measure may sensitize the respondents to their participation in a research project and, subsequently, influence the results of the second measure. Further, if the duration of the time period between measures is long, there may be attitude change, or some other form of maturation, of the subjects which will affect the responses.

The second underlying dimension of reliability concerns the homogeneity of the measure. An attempt to measure an attitude may require asking several questions or a battery of scale items. To measure the **internal consistency** of a multiple-item measure, scores on subsets of items within the scale must be correlated. The **split-half method**, when a researcher checks the results of one half of the scale items to the other half, is the most basic method for checking internal consistency.

The **equivalent-form method** is utilized when two alternative instruments are designed to be as equivalent as possible. If there is a high correlation between the two scales, the researcher can conclude that the scales are reliable. However, if there is a low correspondence, the researcher will be unsure as to whether the measure has intrinsically low reliability, or whether the particular equivalent-form has failed to be similar to the other form.

Both of the above methods assume that the concept is uni-dimensional; they measure homogeneity rather than over-time stability.

B. **Validity**: The purpose of measurement is to measure what we intend to measure. For example, in measuring intention to buy, there could be a systematic bias to identify brands “I wish I could afford” rather than the brand usually purchased. Validity addresses the problem of whether or not a measure does indeed measure what it purports to measure; if it does not, there will be problems.

Researchers attempt to provide some evidence of a measure’s degree of validity. There are three basic approaches to dealing with the issue of validity:

a. Face or content validity: This refers to the subjective agreement of professionals that a scale logically appears to be accurately reflecting what it purports to measure.
b. Criterion validity: Criterion validity is an attempt by researchers to answer the question “Does my measure correlate with other measures of the same construct?” Consider the physical concept of length. If a new measure of lengths were developed, finding that the new measure correlated with other measures of length would provide some assurance that the measure was valid. Criterion validity may be classified as either concurrent validity (when the measure is taken at the same time as the criterion measure) or predictive validity (when the measure predicts a future event).
c. Construct validity: Construct validity is established by the degree to which the measure confirms a network of related hypotheses generated from a theory based on the concept. In its simplest form, if the measure behaves the way it is supposed to in a pattern of inter-correlation with a variety of other variables, then there is evidence for construct validity. This is a complex method of establishing validity and of less concern to the applied researcher than to the basic researcher.
d. Convergent and discriminant validity: Convergent validity is the same as criterion validity because a new measure is expected to predict or converge with similar measures. A measure has discriminant validity when it has a low correlation with measures of dissimilar concepts.

e. Reliability versus validity: The concepts of reliability and validity should be compared. Reliability, although necessary for validity, is not in itself sufficient.

C. **Sensitivity**: The sensitivity of a scale is important; particularly when changes in attitude, or other hypothetical constructs, are under investigation. Sensitivity refers to the ability of a instrument to accurately measure variability in stimuli or responses. The sensitivity of a scale which is based on a single question or a single item can be increased by adding additional questions or items. In other words, because index measures allow for a greater range of possible scores, they are more sensitive than single item scales.

ATTITUDE DEFINED

Attitudes are usually viewed as an enduring disposition to consistently respond in a given manner to various aspects of the world, including persons, events, and objects. There are three components of attitude:

- **Affective component**: reflects an individual's general feelings towards an object.
- **Cognitive component**: represents one's awareness of and knowledge about an object.
- **Behavioral component**: reflects buying intentions and behavioral expectations.

A. Attitudes as Hypothetical Constructs

The term **hypothetical construct** is used to describe a variable that is not directly observable, but is measurable by an indirect means such as verbal expression or overt behavior. Attitudes are considered to be such variables.

B. Measuring Attitudes is Important to Managers

Most marketing managers hold the intuitive belief that changing consumers', or prospects', attitudes towards a product is a major goal, and, thus, the measurement of attitudes is a very important task.

THE ATTITUDE-MEASURING PROCESS

There are a remarkable variety of techniques that have been devised to measure attitudes. These techniques range from direct to indirect, physiological to verbal, etc.

Obtaining verbal statements from respondents generally requires that the respondent perform a task such as ranking, rating, sorting, or making a choice or a comparison.

- **Ranking** tasks require that the respondent rank order a small number of objects in overall preference on the basis of some characteristic or stimulus.
- **Rating** asks the respondent to estimate the magnitude of a characteristic, or quality, that an object possesses. The respondent indicates the position on a scale(s) where he or she would rate an object.
- **Sorting** might present the respondent with several product concepts typed on cards and require that the respondent arrange the cards into a number of piles or otherwise classify the product concepts.
- **Choice** between two or more alternatives is another type of attitude measurement—it is assumed that the chosen object is preferred over the other(s).

Physiological measures of attitudes provide a means of measuring attitudes without verbally questioning the respondent. For example, galvanic skin responses, measure blood pressure, etc., are physiological measures.

ATTITUDE RATING SCALES

Rating scales are perhaps the most common attitude measures. Some examples of ratings scales follow.

A. Simple Attitude Scaling

In its most basic form, attitude scaling requires that an individual agree with a statement or respond to a single question. This type of self-rating scale merely classifies respondents into one of two categories; thus, it has only the properties of a nominal scale. However, such scales are used if questionnaires are extremely long, or for other specific reasons. A number of simplified scales are merely checklists.

Most attitude theorists believe that attitudes vary along continua. The purpose of an attitude scale is to find out an individual's position on the continuum. These simple scales do not allow for making fine distinctions in attitudes, but several other scales have been developed that do provide more precise measurement.

B. Category Scales

A **category scale** is a more sensitive measure than a scale having only two response categories—it provides more information. Question wording is an extremely important factor in the usefulness of these scales.

C. Method of Summated Ratings: The Likert Scale

The **Likert scale** is an extremely popular means for measuring attitudes. With the Likert scale, respondents indicate their own attitudes by checking how strongly they agree or disagree with carefully constructed statements about the attitudinal object. Individuals generally choose from approximately five (although alternatives may range from three to nine) response alternatives: “strongly agree,” “agree,” “uncertain,” “disagree,” and “strongly disagree.”

To measure the attitude, researchers assign scores or weights, which are not printed on the questionnaire, to the answers. Strong agreement indicates the most favorable attitudes on the statement, and the weight of five is assigned to this response. If a negative statement toward the object were given, the weights would be reversed and “strongly disagree” would be assigned the weight of 5. The total score is the summation of the weights assigned to an individual’s total responses. A single scale item on a summated rating scale is an ordinal scale.

In the Likert procedure, a large number of statements are generated and then an **item analysis** is performed. The purpose of the item analysis is to ensure that final items evoke a wide response and discriminate among those with positive and negative attitudes. Items that are poor because they lack clarity or elicit mixed response patterns are eliminated from the final statement list. This step in the questionnaire design is too often neglected by business researchers.

D. Semantic Differential

The **semantic differential** is actually a series of attitude scales. This popular attitude measurement technique consists of the identification of a product, brand, store, or other concept, followed by a series of seven-point bipolar rating scales. Bipolar adjectives, such as “good” and “bad,” anchor either ends (or poles) of the scale. The subject makes repeated judgments of the concept under investigation on each of the scales.

Business researchers have found the semantic differential versatile and have modified the use of the scale for business applications. Replacing the bipolar adjectives with descriptive phrases is an adaptation in image studies.

A weight is assigned to each position on the rating scale. Traditionally, scores are 7, 6, 5, 4, 3, 2, 1 or +3, +2, +1, 0, -1, -2, -3. Many business researchers assume that the semantic differential provides interval data, but some critics argue that the data has only ordinal properties since the weights are arbitrary.

E. Numerical Scales

Numerical scales have numbers as response options, rather than “semantic space” or verbal descriptions, to identify categories (response positions). If the scale items have five response positions, the scale is called a five-point numerical scale. The numerical scale utilizes bipolar adjectives in the same manner as the semantic differential.

F. Stapel Scale

Modern versions of the **Stapel scale** place a single adjective as a substitute for the semantic differential when it is difficult to create pairs of bipolar adjectives.

The advantages and disadvantages of a Stapel scale, as well as the results, are very similar to those for a semantic differential. However, the Stapel scale tends to be easier to conduct and administer.

G. Constant Sum Scale

An example of a **constant-sum scale** is as follows:

Divide 100 points among each of the following brands according to your preference for the brand:

Brand A ___________
Brand B ___________
Brand C ___________

This constant-sum scale works best with respondents having a higher education level. The results will approximate interval measures. However, as the number of stimuli increase, this technique becomes more complex.

H. Graphic Rating Scales

A graphic rating scale presents respondents with a graphic continuum. The respondents are allowed to choose any point on the continuum to indicate their attitude. Typically, the respondent’s score is determined by measuring the length (in millimeters) from one end of the continuum to the point marked by the respondent.

The graphic scale has the advantage of allowing the researcher to choose any interval he wishes for purposes of scoring. The disadvantage of the graphic scale is that there are no standard answers.

A frequently used variation on the graphic scale design is the scale ladder; this and other picture response options enhance communication with respondents.

I. **Thurstone Interval Scale**

The construction of the **Thurstone scale** is a rather complex process that requires two stages. The first stage is a ranking operation performed by judges who assign scale values to attitudinal statements. The second stage consists of asking subjects to respond to the attitudinal statements.

This method is time consuming and costly and is rarely used in applied business research.

MEASURING BEHAVIORAL INTENTION

The behavioral component of an attitude involves the behavioral expectation of an individual toward an attitudinal object. Category scales to measure the behavioral component of an attitude ask a respondent's "likelihood" of purchase or intention to perform some future action. The wording of statements used in these cases often includes phrases such as "I would recommend," "I would write," or "I would buy," to indicate action tendencies.

- **Behavioral differential**: The **behavioral differential** instrument has been developed for measuring the behavioral intentions of subjects towards any object or category of objects. A description of the object to be judged is placed on the top of a sheet, and the subjects indicate their behavioral intentions toward this object on a series of scales. For example:

  A 25-year old woman sales representative
  Would _ : _ : _ : _ : _ : _ : _ : _ : _ : Would not
  Ask this person for advice.

RANKING

Consumers often **rank order** their preferences. An ordinal scale may be developed by asking respondents to rank order (from most preferred to least preferred) a set of objects or attributes. This technique is easily understood by the respondents.

- Paired comparisons: In **paired comparisons** the respondents are presented with two objects at a time and asked to pick the one they prefer. Ranking objects with respect to one attribute is not difficult if only a few products are compared, but as the number of items increases, the number of comparisons increases geometrically (n*(n - 1)/2). If the number of comparisons is too great, respondents may fatigue and no longer carefully discriminate among them.

SORTING

Sorting tasks require that respondents indicate their attitudes or beliefs by arranging items. R. H. Bruskin's A.I.M. (Association-Identification-Measurement) technique, which measures how well customers associate and identify elements of advertising with a product, is a commonly used sorting technique. The BBDO photo sort process is another sorting technique used to measure the affective component of attitudes.

A variation of the constant-sum technique uses physical counters (e.g., coins or cards) to be divided among items being tested.

RANDOMIZED RESPONSE QUESTIONS

In special cases, when respondents are asked to provide sensitive or embarrassing information in a survey, the researcher may utilize **randomized response questions.** In this type of questionnaire, each question has two possible questions associated with it—one sensitive and one non-sensitive. The respondent will answer "yes" or "no" to the question asked. The question asked is randomly selected by the respondent, who confidentially determines which of the two questions will be answered (e.g., by the toss of a coin). A formula is used to estimate the proportion of "yes" answers to the sensitive question. While estimates are subject to error, the respondent remains anonymous, and response bias is therefore reduced.

OTHER METHODS OF ATTITUDE MEASUREMENT

Attitudes, as hypothetical constructs, are not measured directly. Therefore, measurement of attitude is, to an extent, subject to the imagination of the researcher. There are several other techniques, other than the ones presented, that can be utilized when a situation dictates. With the growth of computer technology, techniques such as **multidimensional scaling** and **conjoint analysis** are used more frequently.

SELECTING A MEASUREMENT SCALE: SOME PRACTICAL QUESTIONS

There is no best scale that applies to all research projects. The choice of scale will be a function of the nature of the attitudinal object to be measured, the manager's problem definition, and the backward and forward linkages to other choices that have already been made (e.g., telephone survey versus mail survey). There are several issues that will be helpful to consider:

- ✓ Is a ranking, sorting, rating, or choice technique best? The answer to this question is largely determined by the problem definition and especially by the type of statistical analysis that is desired.
- ✓ Should a monadic or comparative scale be used? If a scale is other than a ratio scale, the researcher must make a decision whether to use a standard of comparison. A monadic rating scale uses no such comparison; it asks a respondent to rate a single concept in isolation. A comparative rating scale asks a respondent to rate a concept in comparison with a benchmark—in many cases, "the ideal situation" presents a reference for comparison with the actual situation.
- ✓ What type of category labels, if any, will be used for the rating scale? We have discussed verbal labels, numerical labels, and unlisted choices. The maturity and educational levels of the respondents and the required statistical analysis will influence this decision.
- ✓ How many scale categories or response positions are required to accurately measure an attitude? The researcher must determine the number of meaningful positions that is best for each specific project.
- ✓ Should a balanced or unbalanced rating scale be chosen? The fixed-alternative format may be balanced—with a neutral or indifferent point at the center of the scale—or unbalanced. Unbalanced scales may be used when the responses are expected to be distributed at one end of the scale; an unbalanced scale may eliminate this type of "end piling."
- ✓ Should respondents be given a forced-choice scale or a non-forced-choice scale? In many situations, a respondent has not formed an attitude towards a concept, and simply cannot provide an answer. If many respondents in the sample are expected to be unaware of the attitudinal object under investigation, this problem may be eliminated by using a non-forced-choice scale that provides a "no opinion" category. The argument for forced choice is that people really do have attitudes, even if they are unfamiliar with the attitudinal object.
- ✓ Should a single measure or an index measure be used? The researcher's conceptual definition will be helpful in making this choice. The researcher has many scaling options. The choice is generally influenced by what is planned for the later stages of the research project.

CHAPTER 10: Questionnaire Design

==

The importance of question wording is easily overlooked, but questionnaire design is one of the most critical stages in the survey research process. Unfortunately, newcomers, who naively believe that common sense and good grammar are all one needs to construct a questionnaire, generally learn that their hasty efforts are inadequate.

QUESTIONNAIRE DESIGN: AN OVERVIEW OF THE MAJOR DECISIONS

Relevance and **accuracy** are the two most basic criteria to be met if the questionnaire is to achieve the researcher's purpose. In order to achieve this, several decisions must be made.

WHAT SHOULD BE ASKED?

The specific questions to be asked will be a function of the previous decisions. The later stages of the research process will have an important impact on the questionnaire wording. When designing the questionnaire, the researcher must also be thinking about the types of statistical analysis that will be conducted.

A. Questionnaire relevancy: A questionnaire is **relevant** if no unnecessary information is collected and if only the information needed to solve the problem is obtained. To ensure information relevance, the researcher must be specific about data needs; there should be a rationale for each item, and all possible omissions should be considered.

B. Questionnaire accuracy. **Accuracy** means that the information is reliable and valid. While it is generally believed that one should use simple, understandable, unbiased, unambiguous, and nonirritating words, no step-by-step procedure can be generalized. Respondents tend to be most cooperative when the subject of the research is interesting - if questions are lengthy, difficult to answer, or ego threatening, there is a high probability of biased answers. Question wording and sequence substantially influence accuracy.

PHRASING QUESTIONS

There are many ways to phrase questions and many standard question formats have been developed in previous research.

A. Open-ended response versus fixed-alternative questions: We may categorize two basic types of questions asked on the amount of freedom respondents are given in answering.

Open-ended response questions pose some problem or topic and ask the respondents to answer in their own words. Open-ended response questions are free response questions. In contrast, **fixed-alternative questions**, or closed questions, give the respondents specific, limited, alternative responses and ask the respondent to choose the response closest to his or her viewpoint. Open-ended response questions are most beneficial when the researcher is conducting exploratory research. By gaining free and uninhibited responses, the researcher may find some unanticipated reaction toward the project. They may also be useful at the beginning of an interview as they allow the respondent to warm up to the questioning process.

The cost of open-ended response questions is substantially higher than that of fixed-alternative questions, since the job of coding, editing, and analyzing the data is quite extensive. Also, open-ended response questions allow potential interviewer bias to influence the answer - even the best interviewer can take shortcuts in recording answers.

B. Fixed-alternative questions require less interviewer skill, take less time, and are easier for the respondent to answer. There are various types of fixed-alternative questions.

1) **Single-dichotomy** or **dichotomous-alternative** questions: require that the respondents choose one of two alternatives. The answer can be a simple “yes” or “no” or a choice between “this” and “that.”
2) Multi-choice alternatives**:** there are several types.

a) **Determinant choice** questions ask a respondent to choose one and only one - response from among several possible alternatives.
b) The **frequency determination** question is a determinant choice question that asks for an answer about the general frequency of occurrence.
c) **Attitude rating scales**, such as the Likert scale.
d) The **check list question** allows respondents to provide multiple answers to a single question. In many cases, the choices are adjectives that describe a particular object. There should be no overlap among categories in the check list - each alternative should be **mutually exclusive**, that is, only one dimension of an issue should be related to that alternative. The researcher should strive to ensure that there are sufficient response choices to include almost all possible answers. However, including a category lower than the answers you expect often helps to negate the potential bias caused by respondents avoiding an extreme category. Respondents, rather than stating why they chose a given product, may select an alternative among those presented. Or, as a matter of convenience, they may select a given alternative rather than thinking of the most correct alternative. Most questionnaires include a mixture of open-ended and closed questions. Each form has unique benefits; in addition, respondent boredom and fatigue are eliminated with a change of pace offered by a mixture of question types.

C. Phrasing questions for mail, telephone, and personal interview surveys: In general, mail and telephone questions must be less complex than those utilized in personal interviews. Questionnaires for telephone and personal interviews should be written in a “conversational” manner.

THE ART OF ASKING QUESTIONS

In developing a questionnaire, there are no hard and fast rules. Some guidelines have been developed to avoid the most common mistakes.

a) **Avoid complexity**: use simple, conversational language: Words used in a questionnaire should be readily understandable to all respondents. The technical jargon of top executives should be avoided.

b) **Avoid leading and loaded questions**: Asking leading and loaded questions is a major source of bias in question wording. **Leading questions** suggest or imply certain answers. Such questions may result in a "bandwagon effect" which threatens the study's validity. Partial mention of alternatives is a variation of this phenomenon.

   Loaded questions suggest social desirability or are emotionally charged. Some questions invite only positive answers. In other cases, respondents are able to interpret which answers are most socially acceptable, and the resulting answers may not portray the respondent's true feelings.

   Invoking the status quo is a form of loading that result in bias because the majority of people tend to be resistant to change.

   Asking respondents "how often" leads them to generalize about their behavior and one is more likely to portray one's ideal behavior rather than one's average behavior. An introductory **counter biasing statement** to a question, that reassures respondents that their "embarrassing" behavior is not abnormal, may help yield truthful responses. Also, an assurance of anonymity may help elicit honest responses to embarrassing questions.

   A question statement may be leading because it is phrased to reflect either the negative or positive aspects of an issue. To control for this bias, **split-ballot technique**, which reverses the wording of attitudinal questions for 50 percent of the sample, can be used.

C. **Avoid ambiguity** - be as specific as possible: Items on questionnaires are often ambiguous because they are too general. Indefinite words such as frequently, often, ready, etc., have many different meanings. Use of such words should be avoided - the questions should be as specific as possible.

D. **Avoid double-barreled items**: A question covering several items at once is referred to as a **double-barreled question** and should always be avoided. There is no need for the confusion that results in a double-barreled question.

E. **Avoid making assumptions**: The researcher should not place the respondent in a bind by including an implicit assumption in the question. Another mistake that question writers sometimes make is assuming that the respondent has previously thought about an issue -research that induces people to express attitudes on subjects that they do not ordinarily think about is meaningless.

WHAT IS THE BEST QUESTION SEQUENCE?

The order of questions may serve several functions for the researcher. For example, if the respondents' curiosity is not aroused at the outset, they can become disinterested and terminate the interview.

Order bias: Order bias results from an alternative answer's position in a set of answers or from the sequencing of questions. Order bias tends to distort survey results. Specific questions tend to influence more general ones. Thus, it is advisable to ask the general questions before the specific questions to obtain the freest of open-ended responses. This technique is known as the **funnel technique**, and it allows researchers to understand the respondent's frame of reference before asking more specific questions about the respondent's particular level of information and intensity of opinion.

When using attitude scales, there also may be an **anchoring effect**. That is, the first concept measured tends to become a comparison point from which subsequent evaluations are made. Randomization of these items on a questionnaire helps to minimize this order bias.

Filter questions **minimize the asking of questions that may be inapplicable**, and **pivot questions** may be used to obtain information that the respondent may be reluctant to provide. For example, a respondent is asked "Is your family income over $30,000?" If under, ask "Is it over or under $10,000?" If over, ask "Is it over or under $50,000?" Notice the logical order of questions which can help ensure the respondent's cooperation, and can help eliminate any confusion or indecision.

WHAT IS THE BEST LAYOUT?

The layout and attractiveness of the questionnaire are of crucial importance. In mail questionnaires, often the rate of return can be improved by adding the money that might have been used as an incentive, to improve the attractiveness and quality of the questionnaire. Questionnaires should be designed to appear as short as possible and experienced researchers have found that it pays to carefully phrase the title to be printed on the questionnaire.

The researcher can design the questionnaire to make the interviewee's job of following interconnected questions much easier by utilizing several forms, special instructions, and other tricks of the trade.

HOW MUCH PRETESTING AND REVISING IS NECESSARY

Usually, the questionnaire is "tried out" on a group selected on a convenience basis and similar in make-up to the one that ultimately will be sampled; it is not necessary to get a statistical sample for pretesting. Pretesting allows the researcher to determine if the respondents have any difficulty understanding the questionnaire or if there are any ambiguous or biased questions. This process can save the potential disaster of administering an invalid questionnaire to several hundred individuals.

A preliminary tabulation of the pretest results often illustrates that while a question is easily comprehended and answered by the respondent, it is inappropriate because it is does not solve the problem.

Pretests are typically conducted to answer questions about the questionnaire such as:

- Can the questionnaire format be followed by the interviewer?
- Does the questionnaire flow naturally and conversationally?
- Are the questions clear and easy to understand?
- Can respondents answer the questions easily?
- Which alternative forms of questions work best?

Pretests provide the means for testing the sampling procedure and also provide estimates for the response rates for the mail surveys and completion rates for telephone surveys.

DESIGNING QUESTIONNAIRES FOR GLOBAL MARKETS

International business researchers must take cultural factors into account when designing questionnaires. The most widespread problem involves translation into another language. International questionnaires are often back translated. **Back Translation** is the process of translating the questionnaire from one language to another and then having it translated back again by a second, independent translator. The back translator is often a person whose native tongue is the language that will be used on the questionnaire.

CHAPTER 11: SAMPLING

==

The process of **sampling** involves using a small number of items or parts of the population to make conclusions regarding the whole population. The purpose of sampling is to estimate some unknown characteristic of the population. A **sample** is a subset or some part of a larger population.

A **population** (finite group) or **universe** (infinite group) is any complete group sharing some common set of characteristics. The term **population element** refers to and individual member of the population. A **census** is an investigation of all the individual elements making up the population - a total enumeration rather than a sample.

WHY SAMPLE?

There are three groups of reasons why a sample rather than a complete census should be taken.

A. Pragmatic reasons: Applied business research projects usually have budget and time constraints. Often, it would not be possible to contact the whole population within a short period of time. Sampling cuts costs, reduces manpower requirements, and gathers vital information quickly.

B. Accurate and reliable results: Another major reason for sampling is that properly selected samples are sufficiently accurate in most cases. Even when the populations have considerable heterogeneity, large samples provide data of sufficient precision to make most decisions. Of course, samples are accurate only when researchers have taken care to properly draw representative samples. A sample may be more accurate than a census. In a census of a large population, there is greater likelihood of non-sampling errors. In a field survey a small, well-trained, closely supervised group may do a more careful and accurate job of collecting information than a large group of nonprofessional interviewers trying to contact everyone.

C. Destruction of test units: Many research projects, especially those in quality control testing, require the destruction of the items being tested. For example, if the manufacturer of firecrackers wished to find out whether each unit met a specific production standard, there would be no product left after testing.

PRACTICAL SAMPLING CONCEPTS

A. Defining the target population: Once the decision to sample has been made, the first question concerns identifying the target population. What is the relevant population? In many cases this is not a difficult question, but in other cases, the decision may be a difficult one. At the outset of the sampling process it is vitally important to carefully define the target population so that the proper source from which the data are to be collected can be identified. Answering questions about the crucial characteristics of the population is the usual technique for defining the target population. The question “Whom do we want to talk to?” must be answered.

B. The sampling frame. In actual practice, the sample will be drawn from a list of population elements that is often somewhat different from the defined target population. A **sampling frame** is a list of elements from which the sample may be drawn. The sampling frame is also called the working population, because it provides the list that can be operationally worked with. The discrepancy between the definition of the population and a sampling frame is the first potential source of error associated with sample selection.

1. Mailing lists: Some firms, called sampling services or list brokers, specialize in providing lists/databases that give the names, addresses, phone numbers, and e-mail addresses of specific populations.
   International Sampling frames: The availability of sampling frames around the world varies dramatically.
2. Sampling frame error: **Sampling frame error** occurs when certain elements are excluded or when the entire population is not accurately represented in the sample frame. It is possible for elements to be either over- or underrepresented in a sampling frame.

C. Sampling units: During the actual sampling process, the elements of the population must be selected according to a certain procedure. The **sampling unit** is a singles element or group of elements subject to selection in the sample. For example, if an airline wishes to sample passengers, every 25th name on a complete list of passengers may be taken. In this case the sampling unit would be the same as the element.

If the target population has first been divided into units, such as airline flights, additional terminology must be used. The term **primary sampling unit (PSU)** is used to designate units selected in the first stage of sampling. If successive stages of sampling are conducted, sampling units are called **secondary sampling units**, or **tertiary sampling units**. When there is no list of population elements, the sampling unit is generally something other than the population element. For example, in a random digit dialing study the sampling unit will be telephone numbers.

RANDOM SAMPLING ERROR AND NONSAMPLING ERROR

Investigators expect a sample to be representative of the population. However, if there is a difference between the value of a sample statistic of interest (for example, average-willingness-to-buy-the-service score) and that of the corresponding value of the population parameter (again, willingness-to-buy score), there is a **statistical error**.

Random sampling error is the difference between the sample result and the result of a census conducted using identical procedures. Random sampling error occurs because of chance variation in the scientific selection of sampling units. It will be seen when we discuss the process of randomization that because random sampling errors follow chance variations they tend to cancel each other out when averaged. What this means is that properly selected samples are generally good approximations of the population. Every once in a while a very unusual sample will be selected because “too many” of the unusual people were included in the sample and there is a large sampling error.

Random sampling error is a function of sample size. As sample size increases, random sampling error decreases. Of course, the resources available will influence how large a sample may be taken.

It is possible to estimate the random sampling error that may be expected with various sample sizes. For example, based on the laws of probability, 95 percent of the time a survey of slightly fewer than 900 people will produce results with an error of approximately plus or minus 3 percent. If the survey had been collected with only 325 people, the margin of error would increase to approximately plus or minus 5 percent. This example illustrates random sampling errors.

Systematic (non-sampling) errors result from non-sampling factors, primarily the nature of a study's design and the correctness of execution. These errors are not due to chance fluctuations. Sample biases account for a large portion of errors in business research.

A. Less than perfectly representative samples: Random sampling errors and systematic errors associated with the sampling process may combine to yield a sample that is less than perfectly representative of the population.

PROBABILITY VERSUS NONPROBABILITY SAMPLING

The major alternative sampling plans may be grouped into probability techniques and nonprobability techniques. In **probability sampling** every element in the population has a known nonzero probability of selection; each member of the population has an equal probability of being selected.

In **nonprobability sampling**, the probability of any particular member of the population being chosen is unknown. It should be noted that there are no appropriate statistical techniques for measuring random sampling error from a nonprobability sample. Thus, projecting the data beyond the sample is inappropriate. Nevertheless, there are occasions when the nonprobability samples are best suited for the researcher's purpose.

NONPROBABILITY SAMPLING

A. **Convenience sampling**: Convenience sampling (also called haphazard or accidental sampling) refers to the sampling procedure of obtaining the people who are most conveniently available. For example, a college professor who uses his or her students has a captive sample—convenient but perhaps unwilling and unrepresentative.

Researchers generally use convenience samples to obtain a large number of completed questionnaires quickly and economically. However, the user of research that is based on a convenience sample should remember that projecting the results beyond the specific sample is inappropriate. Convenience samples are best utilized for exploratory research when additional research will subsequently be conducted with a probability sample.

B. **Judgment sampling**: Judgment or **purposive sampling** is a nonprobability technique in which an experienced individual selects the sample upon his or her judgment about some appropriate characteristic required of the sample members. For example, a fashion manufacturer regularly selects a sample of key accounts that it believes are capable of providing the information to predict what will sell in the fall; the sample has been selected to satisfy a specific objective.

C. **Quota sampling**: The purpose of quota sampling is to ensure that the various subgroups in a population are represented on pertinent sample characteristics to the exact extent that the investigators desire. In quota sampling, the interviewer has a quota to achieve. For example, an interviewer in a particular city may be assigned 100 interviews, 30 of which are with Panasonic owners, 28 with Sony owners, 10 with Magnavox owners, 7 with Toshiba owners, and the rest with owners of other brands. Aggregating the various interview quotas yields a sample representing the desired proportion of the subgroups.

1) Possible sources of bias: the logic of classifying the population by pertinent subgroups is essentially sound. However, because respondents are selected according to a convenience sampling procedure rather than on a probability basis, as in stratified sampling, the haphazard selection of subjects may introduce bias. Quota samples have the tendency to include people who are easily found, willing to be interviewed, and middle class.

2) Advantages of quota sampling: Speed of data collection, lower costs, and convenience are the major advantages of quota sampling over probability sampling. Although there are many problems with this method, careful supervision of the data collection may provide a representative for analyzing the various subgroups within a population.

D. **Snowball sampling**: Snowball sampling refers to a variety of procedures in which initial respondents are selected by probability methods, but additional respondents are then obtained from information provided by the initial respondents. This technique is used to locate members of rare populations by referrals.

Reduced costs and sample sizes are clear-cut advantages of snowball sampling. Bias is likely to enter into the study, however, because a person who is known to someone also in the sample has a higher probability of being similar to the first person. If there are major differences between those who are widely known by others and those who are not, there may be some serious problems with this technique. Since the focus group is not expected to be a generalized sample, snowball sampling may be very appropriate.

PROBABILITY SAMPLING

All probability samples are based on chance selection procedures. This eliminates the bias inherent in nonprobability sampling procedures because the probability sampling process is random. Randomness refers to a procedure, the outcome of which cannot be predicted because it is dependent on chance. It should not be thought of as unplanned or unscientific - it is the basis of all probability sampling techniques. There is several probability sampling techniques.

A. **Simple random sampling**: A simple random sample is a sampling procedure that assures that each element in the population will have an equal chance of being included in the sample. Drawing names from a hat is a typical example of simple random sampling; each person has an equal chance of being selected. This process is simple because is simple because it requires only one stage of sample selection, in contrast to other, more complex probability samples. When populations consist of large numbers of elements, tables of random numbers or computer-generated random numbers are utilized for sample selection.

   Selecting a random sample; to use a table of random numbers, a serial number is assigned to each element of the population. Then, assuming a population of 99,999 or less, five-digit numbers are selected from the table of random numbers merely by reading the numbers in any column or row, by moving upward, downward, left, or right. A random starting point should be selected at the outset.

   The random digit dialing technique of sample selection requires that the researcher identify the exchange or exchanges of interest (the first three numbers) and then use a table of numbers to select the next four numbers.

B. **Systematic sampling**: Systematic sampling is extremely simple: An initial starting point is selected by a random process; then every nth number on the list is selected. To illustrate this procedure, suppose one wish to take a sample of 1,000 from a list consisting of 200,000 names. Using systematic selection every 200th name from the list will be drawn. In this example, the **sampling interval** is 200.

   While this procedure is not actually a random selection procedure, it does yield random results if the arrangements of the items in the list are random in character. The problem of **periodicity** occurs if a list has a systematic pattern, that is, if it is not random in character. If the sampling interval is every 200th name, this could cause a problem, however, periodicity is rarely a problem for most sampling in business research, but researchers should be aware of its possibility.

C. **Stratified sampling**: The first step of choosing strata on the basis of existing information, such as classification of retail outlets, size based on annual sales volume, is the same for both stratified and quota sampling. However, the process of selecting sampling units within the strata differs substantially. In stratified sampling, a subsample is drawn utilizing a simple random sample within each stratum. This is not true with quota sampling.

   The reason for taking a stratified sample is to have a more efficient sample than could be taken on the basis of simple random sampling. A smaller standard error may be the result of a stratified sample because the groups will be adequately represented when strata are combined.

   Another reason for conducting a stratified sample is the assurance that the sample will accurately reflect the population on the basis of the criterion or the criteria used for stratification. Occasionally a simple random sample yields a disproportionate number of one group or another and the representativeness of the sample needs improving.

   A researcher selecting a stratified sample will proceed as follows. First, a variable (sometimes several variables) is identified as an efficient basis for stratification. The variable chosen should increase the homogeneity within each stratum and increase the heterogeneity between strata. The stratification variable is usually a categorical variable or one easily converted into categories, that is, subgroups.

   Next, for each separate subgroup or strata, a list of population elements must be obtained. If a complete listing is not available, a true stratified probability cannot be selected. A table of random numbers or some other device is then used to take a separate random sample within each stratum. Of course, the research must determine how large a sample must be drawn for each stratum.

D. **Proportional versus disproportional strata**: If the number of sampling units from each stratum is in proportion to the relative population size of the stratum, the sample is a **proportional stratified sample**. However, if the primary purpose of the research is to estimate some characteristic of the total among strata, disproportional stratified sampling should be used. To avoid under representing the medium-sized and smaller stores in the sample, a disproportionate sample is taken. In a **disproportional stratified sample**, sample size for each stratum is not allocated on a proportional basis with the population but dictated by analytical considerations. Thus, the strata exhibiting the greatest variability are sampled more heavily to increase sample efficiency, that is, smaller random sampling error. A simplified rule of thumb for understanding the concept of **optimal allocation** is that the stratum sample size increases for strata of larger sizes with the greatest relative variability.

E. **Cluster sampling**: The purpose of **cluster sampling** is to sample economically while retaining the characteristics of a probability sample. In a cluster sample, the primary sampling unit is no, longer the individual element in the population (for example, grocery stores) but a larger cluster of elements located in proximity to one another (for example, cities). The area sample is the most popular type of cluster sample.

Cluster samples are frequently utilized when there are no lists of the sample population available. Ideally a cluster should be as heterogeneous as the population itself - indeed, a mirror image of the population. A problem may arise with cluster sampling if the characteristics and the attitudes of the elements within the cluster are too similar. To an extent this problem may be mitigated by having constructed clusters that are composed of diverse elements and by selecting a large number of sampled clusters.

Multistage area sampling: multistage area sampling involves two or more steps that combine some of the probability techniques already described. It is possible to take as many steps as are necessary to achieve a representative sample. Progressively smaller geographic areas are chosen until a single housing unit is selected for interviewing. The Bureau of the Census provides maps, population information, and demographic characteristics of the population, and so on broken down into several small geographical areas that may be useful in sampling.

INTERNET SAMPLING IS UNIQUE

Internet surveys allow researchers to rapidly reach a large sample. This is both an advantage and a disadvantage. Sample size requirements can be met overnight or in some cases almost instantaneously. A major disadvantage of Internet surveys is the lack of computer ownership and Internet access among certain segments of the population. A sample of Internet users is only representative of Internet users, who tend to be younger, better educated, and more affluent. This is not to say that all Internet samples are unrepresentative of all target populations. Nevertheless, when using Internet surveys, researchers should be keenly aware of potential sampling problems because some members of target populations do not have Internet access.

A. **Web Site Visitors**: At the present time, many Internet surveys are conducted with volunteer respondents who by intention or happenstance visit an organization's web site. These unrestricted samples are clearly convenience samples. They may not be representative because of the haphazard manner by which many respondents arrive at a particular web site or because of self-selection bias.

A better technique for sampling Web site visitors is randomly select sampling units. SurveySite, a company that specializes in Internet surveys collects data by using its "pop-up survey software". The software selects Web visitors at random and "pops up" a small javascript window asking the person if they want to participate in an evaluation survey. If the person clicks "Yes", then a new window opens up containing the online survey. The person can then browse the site at their normal pace and switch to the survey at anytime to express their opinions.

Randomly selecting web site visitors can cause a potential problem. It is possible to over represent the more frequent visitors to the site, and thus represent site visits rather than visitors. There are several programming techniques and technologies that can help accomplish more representative sampling based on site traffic ("cookies", registration data, or pre-screening).

This type of sampling is most valuable if the target population is defined as visitors to a particular Web site. Evaluation and analysis of the visitor's perceptions and experiences of the Web site would be a typical survey objective with this type of sample.

Researchers who have broader interest may obtain Internet samples a variety of other ways.

B. **Panel Samples**: Drawing a probability sample from an established consumer panel or other pre-recruited, membership panel is a popular, scientific, and effective method for drawing Internet samples. Typically samples from a panel yield a high response rate because panel members have already agreed to cooperate with the research organization via e-mail and the Internet. Often panel members are compensated for their time with a sweepstake or a small, cash incentive. Further because the panel has already supplied demographic characteristics and other information from previous questionnaires, researchers have the ability to select panelists based on product ownership, lifestyle, or other characteristics. A variety of sampling methods and data transformation techniques can be applied to assure that sample results are representative of the general public or a targeted population.

Consider Harris Interactive Inc. Harris Interactive is an Internet survey research organization that maintains a United States panel of more than 6.5 million individuals

Because Harris Interactive knows that all demographic groups are not fully accessible via the Internet, it uses a propensity-weighting scheme to ensure that survey results are representative. The research company does parallel studies - phone as well as Internet - to test the accuracy of its Internet data gathering capabilities. Researchers look at the results of the telephone survey and match those against the Internet-only survey results. Next, propensity weighting adjusts the results taking into account the motivational and behavioral differences between the online and offline populations.

C. **Recruited Ad Hoc Samples**: Another means for obtaining an Internet sample is to obtain or create an e-mail addresses sampling frame on an ad hoc basis. Researchers may create the sampling frame offline or online. Databases containing e-mail addresses can be compiled from many sources including customer/client lists, advertising banner recruiting survey participants, online sweepstakes, pop-up windows and registration forms that must be filled out in order to gain access to a particular Web site. Researchers may contact respondents by snail mail or by telephone, ask for their email addresses, and obtain permission for an Internet survey. Using off-line techniques, such as random digit dialing and short telephone screening interview, to recruit respondents can be a very practical way to get a representative sample for an Internet survey. For companies anticipating future Internet research, adding an optional e-mail registration into customer relationship databases (product registration cards, telephone interactions, on-site registration, etc.) can prove to be a valuable database for sample recruitment.

By whatever means the sampling frame is compiled, it is important not to send unauthorized email to respondents. If individuals do not opt-in to receive e-mail from a particular organization - they may consider unsolicited survey requests to be spam. A researcher cannot expect high response rates from individuals who have not given permission to be surveyed. Spamming is not tolerated by advanced Internet uses and can easily backfire to create a host of problems. The most extreme being complaints to the Internet service provider (ISP) and the ISP shutting down the survey site.

Opt-in Lists: Another means for obtaining an Internet sample is to obtain list of e-mail addresses from individuals who have given permission to receive e-mail messages related to a particular topic of interest. Survey Sampling Incorporated is a company that specializes in providing sampling frames and scientifically drawn samples.

WHAT IS THE APPROPRIATE SAMPLE DESIGN?

We have discussed the various advantages and disadvantages of each of the sampling techniques. This section outlines and briefly discusses the most common criteria involved in selecting a sampling design.

i. Degree of accuracy: The degree of accuracy required or the researcher’s tolerance for sampling and non-sampling error may vary from project to project, especially when cost savings or other considerations may be a trade-off for a reduction in accuracy.
ii. Resources: If the researcher’s financial and human resources are restricted, this limitation of resources will eliminate certain methods. Managers concerned with the cost of the research versus the value of the information often will opt for a cost savings from a certain nonprobability sample design rather than make the decision to conduct no research at all.
iii. Time: Researchers who need to make a deadline or complete a project quickly will be more likely to select simple, less time-consuming sample designs.
iv. Advance knowledge of the population: In many cases, a list of population elements will not be available to the researcher. A lack of adequate lists may automatically rule out systematic sampling, stratified sampling, or another sampling design, or it may dictate that a preliminary study, such as a short telephone survey utilizing

random digit dialing, be conducted to generate information to build a sampling frame for the study of primary interest.

v. National versus local project: Geographic proximity of population elements will influence sample design. When population elements are unequally distributed geographically, a cluster sampling design may become much more attractive.

vi. Need for statistical analysis: The need for statistical projections based on the sample is often a criterion. Nonprobability sampling techniques do not allow the researcher to utilize statistical analysis to project the data beyond the sample.

CHAPTER 12: EXPERIMENTS

==

The purpose of experimental research is to allow the investigator to control the research situation so that causal relationships among variables may be evaluated. Events may be controlled in an experiment in a way not possible in a survey. In an **experiment** one variable (the independent variable) is manipulated and its effect upon another variable (the dependent variable) is measured, while other variables that may confound the relationship are eliminated or controlled.

BASIC ISSUES IN EXPERIMENTAL DESIGN

Decisions must be made about four basic elements of an experiment. These elements are described below.

a) **Manipulation of the independent variable:** The experimenter has some degree of control over the **independent variable**. Its value may be changed or altered independently of any other variable. The independent variable is hypothesized to be the causal influence. **Experimental treatments** are the alternative manipulations of the independent variable being investigated.

   In business research, the independent variable is often a **categorical** or **classificatory variable** representing some classifiable or qualitative aspect of business strategy. However, in some situations, the independent variable is a **continuous variable**. The researcher must select the appropriate levels of the variable as experimental treatments. For example, the number of dollars that can be spent on advertising may be any number of different values.

1) **Experimental and control groups:** In the simplest form of experiment only two values of the independent variable are manipulated. By holding conditions in a **control group** at zero, or unchanged, the researcher can control for potential sources of error in the experiment and get a better measure of the effect that the independent variable had on the **experimental group**.
2) **Several experimental treatment levels:** If the use of only two groups does not tell the researchers everything that they wish to know about the relationship between the dependent and independent variable, then additional experimental groups can be added (each with the independent variable at different levels) to get a better idea of the relationship.
3) **More than one independent variable:** It is possible to assess the effects of more than one independent variable. For example, a restaurant chain might investigate the combined effects on sales of an increase in advertising and a change in prices.

b) **Selection and measurement of the dependent variable:** The **dependent variable** is the criterion or standard by which the results are judged. It is presumed that changes in the dependent variable are a consequence of changes in the independent variable.

   Selection of the dependent variable is a crucial decision in the design of an experiment. Often the dependent variable selection process is not as carefully considered by the researcher as it should be. The experimenter's choice of a dependent variable determines what type of answer is given to the research question. For example, in a test market, the time period for the effects to become evident should be carefully considered. Consumers may try a "loser" once, but they may not re-buy.

c) **Selection and assignment of test units: Test units** are the subjects or entities whose responses to the experimental treatment are measured or observed. People are the most common test units in organizational behavior and consumer behavior experiments.

1) **Sample selection error**: Sample selection error may occur because of the procedure utilized to assign subjects or test units to either the experimental group or the control group.

2) **Random sampling error**: This type of error may occur if repetitions of the basic experiment sometimes favor one experimental condition and sometimes the other on a chance basis. However, it is generally accepted that random assignment of participants to groups and experimental treatments is still the best procedure.

3) **Randomization**: This is the random assignment of subjects and treatments to groups. Randomization is one device for equally distributing or scattering the effects of extraneous variables. Thus, the chance of unknown nuisance effects piling up in particular experimental groups can be identified. Random assignment of subjects allows the researcher to assume that the groups are identical with respect to all variables except the experimental treatment.

4) **Matching**: Matching the respondents on the basis of pertinent background information is another technique for controlling assignment errors. For example, if age is expected to influence savings behavior, a savings and loan conducting an experiment may have greater assurance that there are no differences among subjects if subjects in all experimental conditions are matched according to age.

5) **Repeated measure**: Exposing the same subjects to all experimental treatments eliminates any problems due to subject differences, but it causes some other problems which will be discussed later. Such an experimental technique is said to have repeated measure**.**

d) **Control over extraneous variables:** To understand this issue, the various types of experimental error should be understood. The errors in research design can be classified into two basic categories—random error, which is comparable to sampling error in total survey error, and constant error, which is comparable to systematic error in total survey error.

1) **Constant experimental error**: Random error has already been discussed. **Constant error** (bias) occurs when the extraneous variables or the conditions of administering the experiment are allowed to have an influence on dependent variables every time the experiment is repeated. For example, if the subjects of the experimental group are always administered the treatment in the morning and the subjects of the control group are always administered the treatment in the afternoon, this will result in a constant, systematic error which must be taken into account.

2) **Extraneous variables**: A number of extraneous variables may affect the dependent variable, thereby distorting the experiment. For example, when testing the efficiency of a car fuel, if cars with different engine sizes were used in the test then extraneous variables would affect the results.

3) **Demand characteristics**: The term demand characteristic refers to experimental design procedures that unintentionally hint to subjects about the experimenter's hypothesis. If participants recognize the experimenter's expectation or demand, they are likely to act in a manner consistent with the experimental treatment; even slight nonverbal cues may influence their reactions. The **guinea pig effect**, which occurs when subjects exhibit the behavior that they feel, is expected of them rather than their actual behavior may occur.

   A common effect caused by demand characteristic is the **Hawthorne effect.** This causes people to perform differently just because they know they are experimental subjects.

   Another common effect occurs when people in an experiment interact; this may result in "joint" decisions rather than a desired individual decision. To reduce demand characteristics, steps are typically taken to make it difficult for subjects to know what the researcher is trying to find out. If the purpose of the experiment is disguised, the participant does not know how to be a "good" subject to "help" confirm the hypothesis. Of course, the use of deception presents an ethical question that must be resolved by the researcher.

4) **Establishing control**: The major difference between experimental research and other research is the experimenter's ability to hold conditions constant and manipulate the treatment. When extraneous variables cannot be eliminated, experimenters strive for **conditions of consistency**. This procedure strives to have all the

subjects in each experimental group exposed to situations that are exactly alike except for the differing conditions on the independent variable.

If the experimental method requires that the same subjects be exposed to two or more experimental treatments, there may be an error due to **order of presentation**. For example, consumers in a taste test may have a tendency to prefer the first product they taste if they cannot taste the difference between the two products. **Counterbalancing** attempts to eliminate the confounding effects of order of presentation by requiring that half of the subjects be exposed to treatment A first and then to treatment B while the other half receive treatment B first then treatment A.

Blinding is utilized to control subjects' knowledge of whether or not they have been given a particular experimental treatment. This technique is frequently used in medical research when chemically inert pills (placebos) are given to subjects who do not receive any medication. It may also involve experimenters; the researchers do not know which of the pills are real and which placebos are. This kind of experiment is known as a **double-blind design** because neither the subject nor the experimenter knows which are the experimental and which are the controlled conditions.

The random assignment of subjects and experimental treatments to groups is an attempt to control extraneous variations resulting from chance. If certain extraneous variations cannot be controlled, it must be assumed that the confounding effects will be present in all experimental conditions with approximately the same influence.

It is not always possible to control everything that would be desirable in the perfect experiment.

ETHICAL ISSUES IN EXPERIMENTATION

The researchers must concern themselves with issues relating to privacy, confidentiality, deception, and other ethical concerns.

Debriefing is the process of providing subjects with all the pertinent facts about the experiment after the experiment has been completed. When deception is used, is recommended that researchers debrief subjects of laboratory experiments to explain the purpose of the research.

FUNDAMENTAL QUESTIONS IN EXPERIMENTATION

A. **Basic versus factorial experimental designs:** In **basic experimental designs**, a single independent variable is manipulated to observe its effect on another, single dependent variable. **Factorial experimental designs** are more sophisticated than basic experimental designs. They allow for an investigation of the interaction of two or more independent variables.

B. **Field and laboratory experiments**: A experiment can be conducted in a natural setting (**field experiment**) or in an artificial setting, one contrived for a specific purpose (**laboratory experiment**). In a laboratory experiment, the researcher has almost complete control over the research setting. Some laboratory experiments are more controlled than others. An example of a very controlled experiment is a **tachioscope** which can be used to experiment with the visual impact of advertising, packaging and so on by controlling the amount of time a visual image is exposed to a subject.

Field experiments are generally utilized to fine-tune business strategies and to determine sales volume.

As experiments increase in naturalism they begin to approach the pure field experiment, and as they become more artificial they approach the laboratory type. The degree of artificiality in experiments refers to the amount of manipulation and control of the situation that the experiment creates to ensure that the subjects will be exposed to the exact conditions desired.

Generally, subjects will be aware of their participation in laboratory experiments and it is common to debrief subjects of laboratory experiments to explain the purpose of the research.

ISSUES OF EXPERIMENTAL VALIDITY

A. ***Internal validity:*** Internal validity refers to whether or not the experimental treatment was the sole cause of observed changes in the dependent variable. It is helpful to classify several different types of extraneous variables that may jeopardize internal validity.

1) **History**: History refers to specific events in the environment between the first and second measurement that are beyond the control of the experimenter. A common history effect occurs when competitors change their business strategies during a test marketing experiment. A special case of the history effect occurs when a change in the

dependent variable occurs because members of one experimental condition experienced historical situations different from those of the other experimental conditions. This is referred to as the **cohort effect**.

2) **Maturation**: The effects due to maturation are changes within the respondents that operate as a function of time rather than of a specific event. For example, during a day-long experiment subjects may grow tired, bored, or hungry.
3) **Testing effects**: These are also called pretesting effects because the initial measurement or test alerts respondents to the nature of the experiment and respondents may act differently than they would have if no pretest measures were taken. In a before-and-after study, taking a pretest before the independent variable is manipulated may sensitize respondents when they are taking the test for a second time. If the identical measuring instrument is used more than once, there may be a testing effect. To avoid the effects of testing, an alternate form of measuring instrument may be given during the posttest or after measurement. Although this may reduce the effect of testing because of a change in the measuring instrument, this may result in an instrumentation effect. There are numerous sources of instrument decay or variation.
4) **Selection**: The selection effect is a sample bias resulting in differential selection of respondents for the comparison groups.
5) **Mortality**: If the experiment is conducted over a period of a few weeks or more, there may be some sample error due to mortality, or **sample attrition**. This occurs when subjects withdraw from the experiment before it is completed. Mortality effects may occur if many subjects drop from one experimental group and not from other treatment or control groups.

B. ***External validity*:** External validity is the quality of being able to generalize beyond the data of the experiment. In essence, it is a sampling question: To what extent can the results of a simulated shopping experiment be transferred to real-world supermarket shopping? Of course, if an experiment lacks internal validity, projecting the result is not possible. Thus, the threats to internal validity may also jeopardize external validity.

1) **Student surrogates**: One issue of external validity concerns the common use of students as experimental subjects. The student population is atypical and, although students are easily accessible, they often are not representative of the total population.
2) **Trade-offs between internal and external validity**: Natural experiments tend to have greater external validity than artificial laboratory experiments, but laboratory experiments with many controlled factors are usually higher in internal validity. A trade-off must be made when designing the experiment.

CLASSIFICATION OF EXPERIMENTAL DESIGNS

There are various types of experimental designs. If only one variable is manipulated, the experiment is a **basic design**. If the experimenter wishes to investigate several levels of the independent variable (for example, four price levels), or if he or she wants to investigate effects of two or more independent variables, then the experiment requires a **complex**, or **statistical**, experimental design.

A. Symbolism for Diagramming Experimental Designs

The following symbolism facilities the description of the various experimental designs:

X = exposure of a group to an experimental treatment.

O = observation or measurement of the dependent variable. If more than one observation is taken, subscripts will be given to indicate temporal order.

R = random assignment of test units.

As we diagram designs utilizing these symbols, the reader should assume a time flow from left to right.

B. Three examples of quasi-experimental design: Quasi-experimental designs do not qualify as true experimental designs because they do not adequately control for the problems associated with loss of external or internal validity.

1) **One-shot design, or the after-only design:**

$$X \quad O_1$$

This one-shot design is a case study fraught with problems. In this experiment, we do not have any kind of comparison. We have no means of controlling extraneous variables. We need a measure of what would happen when the test units have not been exposed to X to compare with the measure when the subjects have been exposed to X. Nevertheless, under certain circumstances this design is the only viable choice.

2) One-group pretest-posttest design:

$$O_1 \quad X \quad O_2$$

This design offers a comparison on the same individuals before and after training. Although this is an improvement over the one-shot design, this design still has certain weaknesses such as maturation, testing effect, and mortality that may jeopardize internal validity. However, despite its weaknesses, this design is used frequently in business research.

3) Static group design:

Experimental group:	**X**	O_1
Control group:		O_2

The results of a static control group are computed by subtracting the observed results in the experimental group from those in the control group ($O_1 - O_2$). A major weakness of this design is that we have no assurance that the groups were equal on variables of interest before the experimental group received the treatment. If the groups were selected arbitrarily by the investigator or if entry into either group was voluntary then there may be systematic differences between the groups that could invalidate the conclusions about the effect of the treatment.

Random assignment of subjects may minimize problems with group differences. If the groups can be determined by the experimenter rather than existing as a function of some other causation, the static group design is referred to as an "after-only" design with control group.

On many occasions "after-only" designs are the only possible ones. This is particularly true when conducting "use tests" for new products or brands.

C. Three good experimental designs: The three basic experimental designs discussed next will have the symbol R to the left of the diagram to indicate that the first step in a ***true experimental design*** is the randomization of subject assignment.

1) Pretest-posttest control group design (before-after with control):

Experimental group:	**R**	O_1	**X**	O_2
Control group:	**R**	O_3		O_4

This design has the advantage of the before-after design with the additional advantages gained from having a control group. The effect of the experimental treatment equals $(O_2 - O_1) - (O_4 - O_3)$.

It is assumed that the effect of extraneous variables will be the same on both the experimental and the control groups. This assumption is also made for history, maturation, testing effects, instrument decay, and regression effects. However, **interactive testing effect** is possible when the subjects are sensitized to the subject of the research. This weakness in the before-after with control group design can be corrected (see the next two designs).

2) Posttest-only control group design (after-only with control):

Experimental group:	**R**	**X**	O_1
Control group:	**R**		O_2

The effect of the experimental treatment is equal to O_2 - O_1. In some situations no pretest measure is possible about the effect of the treatment (for example, an experiment testing the effectiveness of athlete's foot remedy). With only posttest measurement, the effects of testing and instrument decay are eliminated. Further, all the same assumptions about extraneous variables are made, that is, they operate equally on both groups.

3) Solomon four-group design:

Experimental group 1:	**R**	O_1	**X**	O_2
Control group 1:	**R**	O_3		O_4
Experimental group 2:	**R**		**X**	O_5
Control group 2:	**R**			O_6

Although we will not go through the calculations, it is possible to isolate the effects of the experimental treatment and interactive testing in this design. Although this design allows for isolation of the various effects, it is rarely used in business research because of the time, effort, and cost of implementing it.

D. Compromise experimental designs: In many instances of business research true experimentation is not possible—the best the researcher can do is approximate an experimental design. These **compromise designs** may fall short of the requirements of assigning subjects or treatments randomly to groups. The alternative to the compromise design is to conduct the experiment without a control group. Generally this is considered a greater weakness than utilizing groups that have already been established.

When the experiment involves a longitudinal study, circumstances usually dictate a compromise with true experimentation.

E. Time series experimental designs: Business experiments that are investigating long-term structural change may require a **time series design**. When experiments are conducted over long periods of time, they are most vulnerable to historical changes. In such cases the following quasi-design is utilized:

$$O_1 \quad O_2 \quad O_3 \quad X \quad O_4 \quad O_5 \quad O_6$$

Several observations are taken to identify trends before the treatment is administered. After the treatment, several observations are made to determine if the patterns after the treatment are similar to those before. Of course, this time series design cannot give the researcher complete assurance that the treatment caused the change in the trend, but it does enable the researcher to distinguish temporary changes from permanent changes. An improvement on the basic time series design is to utilize a time series with control group.

COMPLEX EXPERIMENTAL DESIGNS

Complex experimental designs are statistical designs that allow for the manipulation of more than one independent variable in the experiment.

A. Completely randomized designs: This design uses a random process to assign experimental units to treatments. Randomization of experimental units is the researcher's attempt to control all extraneous variables while manipulating a single factor, the researcher variable. Several of the experiments discussed in the previous are completely randomized.

B. Randomized block design: The **randomized block design** is an extension of the completely randomized design. A form of randomization is utilized to control for most extraneous variables. However, if the researcher has identified a single "extraneous variable" that might affect test units' response to the treatment, then the researcher will attempt to isolate the single variable by blocking out its effects.

By grouping test units into homogeneous blocks of some relevant characteristic, one known source of extraneous variation may be separately accounted for. The logic behind the randomized block design is similar to the logic underlying the selection of a stratified sample rather than a simple random sample. By isolating the block effects, one type of extraneous variation are partitioned out and a more efficient experimental design results.

C. Factorial design: Even though the single-factor experiments already considered may have one specific variable blocked and others controlled for, they are still limited. **Factorial designs** allow for the testing of two or more treatments (factors) at various levels.

A factorial design allows for the simultaneous manipulation of two or more independent variables at various levels. The **main effect** is the influence on the dependent variable by each independent variable, and the **interaction effect** is the effect of the combination of the independent variables on the dependent variable. A major advantage of the factorial design is its ability to measure the interaction effect which may be more or less than the total of the main effects.

D. Latin square design: This design attempts to block out the effect of two or more confounding extraneous factors. The Latin square is a balanced, two-way classification square. In the following 3 x 3 matrix, each letter occurs only once in each row and in each column.

Subject	Order of Usage		
	1	2	3
1	A	B	C
2	B	C	A
3	C	A	B

The letters A, B, and C identify the three treatments; the rows and columns of the table identify the confounding factors. The end result of this design is that each treatment will be administered under conditions involving all levels of both confounding factors.
A major assumption of the Latin square design is that interaction effects are expected to be minimal or nonexistent.
A Latin square design may have any number of treatments. For example, a 5 x 5 matrix may be used with 25 cells. The number of treatment levels for the confounding factors must be equal which can present certain problems. For example, suppose a retail group wishes to control for shelf space and city where the product is sold. The retail group may be limited in its experiment because it markets in only three cities, but it wishes to experiment with four shelf heights. Even with some drawbacks, the Latin square design can be very advantageous to use in certain situations.

CHAPTER 13: QUANTITATIVE DATA ANALYSIS

===

REVIEW OF BASIC TERMINOLOGY

Descriptive and Inferential Statistics; There are two applications of statistics: (1) to describe characteristics of the population or sample (descriptive statistics) and (2) to generalize from the sample to the population (inferential statistics).
Sample Statistics and Population Parameters; The primary purpose of inferential statistics is to make a judgment about the population, or the collection of all elements about which one seeks information. The sample is a subset or relatively small fraction of the total number of elements in the population. **Sample statistics** are variables in the sample or measures computed from the sample data. **Population parameters** are variables or measured characteristics of the population. We will generally use Greek lowercase letters to denote population parameters (e.g., μ or σ) and English letters to denote sample statistics (e.g., X or S).

MAKING THE DATA USABLE

To make data usable, this information must be organized and summarized.
Methods for doing this include:

- frequency distributions
- proportions
- measures of central tendency and dispersion

Frequency Distributions; Constructing a frequency table or **frequency distribution** is one of the most common means of summarizing a set of data. The frequency of a value is the number of times a particular value of a variable occurs. It is also quite simple to construct a distribution of relative frequency, or a **percentage distribution**, which is developed by dividing the frequency of each value by the total number of observations, and multiply the result by 100.
Probability is the long-run relative frequency with which an event will occur. Inferential statistics uses the concept of a probability distribution, which is conceptually the same as a percentage distribution except that the data are converted into probabilities.

A **proportion** indicates the percentage of population elements that successfully meet some standard on the particular characteristic. May be expressed as a percentage, a fraction, or a decimal number.

Measures of Central Tendency; There are three ways to measure the central tendency, and each has a different meaning.

- **Mean;** The **mean** is simply the arithmetic average, and it is a common measure of central tendency. It is the sum of all the observations divided by the number of observations. Often we will not have enough data to calculate the population mean, μ, so we will calculate a sample mean, $\overline{X}$ (read as "X bar").
- **Median;** The **median** is the midpoint of the distribution, or the 50th percentile. In other words, the median is the value below which half the values in the sample fall.
- **Mode;** The **mode** is the measure of central tendency that merely identifies the value that occurs most often. Determined by listing each possible value and noting the number of times each value occurs.

Measures of Dispersion; Accurate analysis of data also requires knowing the tendency of observations to depart from the central tendency. Thus, another way to summarize the data is to calculate the dispersion of the data, or how the observations vary from the mean. There are several measures of dispersion discussed below.

- **Range;** The simplest measure of dispersion. It is the distance between the smallest and largest values of a frequency distribution. Does not take into account all the observations; it merely tells us about the extreme values of the distribution. While we do not expect all observations to be exactly like the mean, in a skinny distribution they will be a short distance from the mean, while in a fat distribution they will be spread out. The **interquartile range** is the range encompassing the middle 50 percent of the observations (i.e., the range between the bottom quartile and the top quartile).
- **Deviation Scores;** A method of calculating how far any observation is from the mean is to calculate individual deviation scores. A deviation of any observation from the mean can be calculated by subtracting the mean from that observation.

▪ **Why Use the Standard Deviation?** It is perhaps the most valuable index of spread, or dispersion. Learning about the standard deviation will be easier if we present several other measures of dispersion that may be used. Each of these has certain limitations that the standard deviation does not. **Average deviation** – determined by calculating the deviation score of each observation value (i.e., its difference from the mean) and summing up each score; then we divide by the sample size (*n*). While this means of calculating a measure of spread seems of interest, it is never used because the positive deviation scores are always canceled out by the negative deviation scores, thus leaving an average deviation value of zero. One might correct the disadvantage of the average deviation by computing the absolute values of the deviations. We could ignore all the positive and negative signs and utilize only the absolute values of each deviation to give us the mean absolute deviation, but there are some technical mathematical problems that make the mean absolute deviation less valuable than some other measures.

▪ **Variance;** This is another means of eliminating the sign problem caused by the negative deviations canceling out the positive deviations. Useful for describing the sample variability. This procedure is to square the deviation scores and divide by the number of observations. (In actual fact *n* - 1 rather than *n* is used in most research problems.). A very good index of the degree of dispersion. The variance, S^2, will be equal to zero if—and only if—each and every observation in the distribution is the same as the mean.

▪ **Standard Deviation; t**he variance does have one major drawback—it reflects a unit of measurement that has been squared. Because of this, statisticians have taken the square root of the variance. The square root of the variance for distribution is called the **standard deviation**. *S* is the symbol for the sample standard deviation, while σ is the symbol for the population standard deviation.

THE NORMAL DISTRIBUTION

- One of the most common probability distributions in statistics is the **normal distribution** (a.k.a., the *normal curve*). Bell shaped and almost all (99 percent) of its values are within ± 3 standard deviations from its mean. The **standardized normal distribution** is a specific normal curve that has several characteristics. It is symmetrical about its mean. The mean identifies its highest point (the mode) and vertical line about which this curve is symmetrical. The normal curve has an infinite number of cases (it is a continuous distribution), and the area under the curve has a probability density equal to 1.0. The standardized normal distribution has a mean of 0 and a standard deviation of 1. The standardized normal distribution is a purely theoretical probability distribution, but it is the most useful distribution in inferential statistics. The standardized normal distribution is extremely valuable because we can translate or transform any normal variable, *X*, into the standardized value, *Z*. Computing the standardized value, *Z*, of any measurement expressed in original units is simple: Subtract the mean from the value to be transformed, and divide by the standard deviation (all expressed in original units). In the formula note that σ, the population standard deviation, is used for calculation:

$$Z = \frac{X - \mu}{\sigma}$$ where μ = hypothesized or expected value of the mean

POPULATION DISTRIBUTION, SAMPLE DISTRIBUTION, AND SAMPLING DISTRIBUTION

Three additional types of distribution must be defined:

1. population distribution
2. sample distribution
3. sampling distribution

A frequency distribution of the population elements is called a **population distribution**. The population distribution has its mean and standard deviation represented by the Greek letters μ and σ.

A frequency distribution of a sample is called the **sample distribution**. The sample mean is designated with $\overline{X}$, and the sample standard deviation is designated *S*.

A **sampling distribution** is a theoretical probability that shows the functional relation between the possible values of some summary characteristic of *n* cases drawn at random and the probability (density) associated with each value over all possible samples of size *n* from a particular population. The sampling distribution's mean is called the *expected value* of the statistic. The expected value of the mean of the sampling distribution is equal to μ. The standard deviation of the sampling distribution is called the **standard error of the mean** ($S_{\overline{X}}$) and is approximately equal to $\sigma/\sqrt{n}$. As sample size increases, the spread of the sample mean around μ decreases (i.e., larger samples will have a skinnier sampling distribution).

CENTRAL-LIMIT THEOREM

The **central-limit theorem** states: As the sample size, *n*, increases, the distribution of the mean, $\overline{X}$, of a random sample taken from practically any population approaches a normal distribution (with a mean μ and a standard deviation, $\sigma/\sqrt{n}$). The central-limit theorem works regardless of the shape of the original population distribution. This theoretical knowledge about distributions can be used to solve two very practical research problems: (1) estimating parameters & (2) determining sample size

ESTIMATION OF PARAMETERS

- **Point Estimates;** Our goal in utilizing statistics is to make an estimate about the population parameters. The population mean, μ, and standard deviation, σ, are constants, but in most instances of research they are unknown. To estimate the population values, we are required to sample. Example: To estimate the average number of people participating in racquetball in one week we may take a sample of 300 racquetball players throughout the area where our researcher is thinking of building club facilities. If the sample mean, $\overline{X}$, equals 2.6 days per week, we may use this figure as a **point estimate**. This single value, 2.6, is the best estimate of the population mean. However, we would be extremely lucky if the sample estimate were exactly the same as the population value. A less risky alternative would be to calculate a confidence interval.

Confidence Intervals; A **confidence interval** estimate is based on the knowledge that $\mu = \overline{X} \pm$ a small sampling error. After calculating an interval estimate, we can determine how probable it is that the population mean will fall within this range of statistical values. In the racquetball example the researcher, after setting up a confidence interval, would be able to make a statement such as "with 95 percent confidence, I think that the average number of days played per week is between 2.3 and 2.9." The researcher has a certain confidence that the interval contains the true value of the population mean. The crux of the problem for the researcher is to determine how much random sampling error to tolerate. The **confidence level** is a percentage or decimal that indicates the long-run probability that the results will be correct. Traditionally, researchers have utilized the 95 percent confidence level. The confidence interval gives the estimated value of the population parameter, plus or minus an estimate of the error:

$$\mu = \overline{X} \pm \text{ a small sampling error (E)}$$

$$\text{where } E = Z_{c.1.} S_{\overline{X}}$$

$Z_{c.1.}$ =the value of Z, our standardized normal variable at a specified confidence level.

$S_{\overline{X}}$ =the standard error of the mean.

The following is a step-by-step procedure for calculating confidence intervals:

1. Calculate $\overline{X}$ from the sample.
2. Assuming σ is unknown, estimate the population standard deviation by finding *S,* the sample deviation.
3. Estimate the standard error of the mean, using the following formula: $S_{\overline{X}} = S/\sqrt{n}$.
4. Determine the *Z*-value associated with the confidence level desired. The confidence level should be divided by 2 to determine what percentage of the area under the curve must be included on each side of the mean.
5. Calculate the confidence interval.

Sample statistics, such as the sample means, $\overline{X}$ s, can provide good estimates of population parameters such as μ .

There will be a random sampling error, which is the difference between the survey results and the results of surveying the entire population.

SAMPLE SIZE

Random Error and Sample Size; Random sampling error varies with samples of different sizes.

Increasing the sample size decreases the width of the confidence interval at a given confidence level.

When the standard deviation of the population is unknown, a confidence interval is calculated by using the following formula:

$$\overline{X} \pm Z \frac{S}{\sqrt{n}}$$

Observe that the equation for the plus or minus error factor in the confidence interval includes *n*, the sample size. If *n* increases, E is reduced. Increases in sample size reduce sampling error at a *decreasing rate*. More technically, random sampling error is inversely proportional to the square root of *n*. Thus, the main issue becomes ones of determining the optimal sample size.

1. **Factors in Determining Sample Size for Questions Involving Means;** Three factors are required to specify sample size: (1) the variance, or heterogeneity, of the population, (2) the magnitude of acceptable error & (3) the confidence level

- The *variance*, or *heterogeneity*, of the population characteristic in statistical terms refers to the *standard deviation* of the population parameter. Only a small sample is required if the population is homogeneous. As *heterogeneity* increases, so must sample size.
  The *magnitude of error*, or the confidence interval, is defined in statistical terms as *E*, and indicates how precise the estimate must be. It indicates a certain precision level. From a managerial perspective, the importance of the decision in terms of profitability will influence the researcher's specifications of the range of error.
- The third factor of concern is the *confidence level*. We will typically use the 95 percent confidence level. This, however, is an arbitrary decision based on convention.

Estimating Sample Size for Questions Involving Means; The researcher must follow three steps: (1) Estimate the standard deviation of the population, (2) Make a judgment about the desired magnitude of error & (3) Determine a confidence level. The only problem is estimating the standard deviation of the population. Ideally, similar studies conducted in the past will be used as a basis for judging the standard deviation. In practice, researchers without prior information conduct a pilot study to estimate the population parameters so that another, larger sample, of the appropriate sample size, may be drawn. This procedure is called ***sequential sampling***, because researchers take an initial look at the pilot study results before deciding on a larger sample to provide more precise information. *A rule of thumb for estimating the value of the standard deviation is to expect it to be one-sixth of the range.* In a general sense, doubling sample size will reduce error by only approximately one-quarter.

The Influence of Population Size on Sample Size; In most cases the size of the population does not have a major effect on the sample size. The variance of the population has the largest effect on sample size. However, a finite correction factor may be needed to adjust the sample size if that size is more than 5 percent of a finite population. If the sample is large relative to the population, the above procedures may overestimate sample size, and there may be a need to adjust sample size.

Determining Sample Size on the Basis of Judgment; Sample size may also be determined on the basis of managerial judgments. Using a sample size similar to those used in previous studies provides the inexperienced researcher with a comparison of other researchers' judgments. However, the cost of data collection becomes a major consideration, and judgment must be exercised regarding the importance of such information. Another consideration stems from most researchers' need to analyze the various subgroups within the sample.

- There is a judgmental rule of thumb for selecting minimum subgroup sample size: Each subgroup to be separately analyzed should have a minimum of 100 or more units in each category of the major breakdowns. With this procedure, the total sample size is computed by totaling the sample size necessary for these subgroups.

Determining Sample Size for Stratified and Other Probability Samples; Stratified sampling involves drawing separate samples within the subgroups to make the sample more efficient. With a stratified sample, the sample variances are expected to differ by strata. This makes the determination of sample size more complex. Increased complexity may also characterize the determination of sample size for cluster sampling and other probability sampling methods.

CHAPTER 14: RESEARCH PROJECTS: PRESENTATION & DISCUSSION

==

The theory of communications helps clarify the importance of the research report. Several elements enter into a successful communication.

A. The communicator, the writer of the report.
B. The message, the findings of the report.
C. The medium, the oral or written report itself.
D. The audience, the manager who will make a decision based on the report findings.
E. Feedback, the manager's response to the report.

However, this model of communication simplifies the case. It seems that the message flows smoothly along from the writer to the reader, who in turn promptly provides the writer with feedback. Actually things are more complex.

The communicator and the audience each have a field of experience. These overlap to some extent, otherwise no communication would be possible. The message is successfully communicated only if there is enough common experience for it to be encoded, transmitted, and decoded with roughly the same meaning.

In the research setting we have a communicator (the researcher) who has spent a great deal of time studying a problem. The researcher may assume the reader has a lot of background information on the project and may produce pages of unexplained tables, assuming the reader will dig out from them the same patterns the researcher has observed. At the other extreme, the report may go overboard, explaining everything in sixth-grade terms to make sure the reader is not lost, thereby insulting the reader.

In reality, when the reader receives the report, he or she usually has not thought much about the project. The reader may or may not know anything about statistics and if the report cannot be understood quickly, it may be put on the stack of things to do "someday."

Under these circumstances, delivering the report to an audience is not sufficient. It needs to be written so as to hit the area of common experience between the researcher and the manager. The effort to hit that zone is the responsibility of the writer, not the reader.

THE REPORT IN CONTEXT

The **research report** is an oral presentation and/or a written statement whose purpose is to communicate research results, strategic recommendations, and/or other conclusions to management and/or other specific audiences.

REPORT FORMAT

Although every research report is custom-made for the project it represents, some conventions of **report format** are universal. They represent a consensus about what parts are necessary to a good research report and how they should be ordered.

a) **Tailoring the format to the project:** The format may need adjustment for two reasons: (1) to obtain the proper level of formality and (2) to alter the complexity of the report. The format given here is for the most formal type of report. How does the researcher decide on the appropriate level of formality? The general rule is to include all the parts needed for effective communication in the particular circumstances and no more. This factor relates to how far up in management the report is expected to go and on how routine the matter is.

b) Parts of the report

1) **The title page**: The title page should include the title of the report, for whom the report was prepared, who prepared it, and the date of release or presentation. The title should give a brief but complete indication of the purpose of the research project.
2) **Letter of transmittal**: This element is included in relatively formal to very formal reports. Its purpose is to release or deliver the report to the recipient. It also serves to establish some rapport between the reader and the writer.
3) **Letter of authorization**: This is a letter to the researcher approving the project, detailing who has responsibility for the project, and what resources are available to support it. The researcher would not write this personally.
4) **Table of contents**: The table of contents is essential to any report more than a few pages long. It should list the divisions and subdivisions of the report with page references. If the report includes many figures or tables, a list of these should also be included, immediately following the table of contents.
5) **Summary**: The summary is a vital part of the report. Studies have indicated that nearly all managers read a report's summary while only a minority read the rest of the report. The summary should be written only after the rest of the report has been completed. Its length should be about one page, so the writer must carefully sort out what is important enough to include in it. The summary contains four elements. First, the objectives of the report are stated, including the most important background information and the specific purposes of the project. Second, the major results regarding each purpose are presented. Third conclusions. Finally recommendations or suggestions for action, based on the conclusions. In many cases, managers prefer not to have recommendations included in the report or summary.
6) **The body**: The body constitutes the bulk of the report. It begins with an **introduction** which explains why the project was done and what it aimed to discover. Enough background should be included to explain why the project was worth doing, but essential historical factors should be omitted. The last part of the introduction explains what this particular project tried to discover. The second division of the body explains the **research methodology**. This involves an explanation of the research design, data collection methods, sampling procedures, and other technical procedures dealing with collection of data such as field work and analysis. The presentation of **results** should occupy the bulk of the report. No report is perfect so its limitations should be indicated. However, the discussion of limitations should avoid overemphasizing the weaknesses. The last division of the body presents the **conclusions and recommendations**. The conclusions and recommendations should be presented here in more detail than in the summary, with whatever justification is needed.
7) **The appendix**: Any material that is too technical or too detailed to go in the body of the paper should appear in the appendix.

EFFECTIVE USE OF GRAPHIC AIDS

Used properly, graphic aids can clarify complex points or emphasize a message, but used improperly or sloppily they can be distracting or misleading. The key to effective use of graphic aids is to make them an integral part of the text; the key points should be pointed out and related to the discussion in progress. Several types of graphic aids may be useful in research reports.

a) **Tables**: Tables are most useful for presenting numerical information, especially when several pieces of information have been gathered about each item discussed. Each table should include a table number, which allows simple reference from the text, a title, a boxhead and stubhead, footnotes, and a source note.

b) **Charts**: Charts translate numerical information into visual form so that relationships can be easily grasped. The accuracy of the numbers is reduced to gain this advantage. Each chart should include a figure number, allowing easy reference from the text, a title, an explanation of the chart, and a source and footnotes. Charts are subject to distortion, whether unintentional or deliberate. A particularly severe kind of distortion comes from treating unequal intervals as if they were equal. Another common distortion is to begin the vertical scale at some value larger than zero; graphs should always start at zero on the vertical axis.

1) **Pie charts**: One of the most useful types of charts is the pie chart. A pie chart shows the composition of some total quantity at a particular time. Each angle or "slice" is proportional to its percentage of the whole and should be labeled with its description and percentage. The writer should not try to include too many slices—about six slices is a usual maximum.

2) **Line graphs**: Line graphs are useful to show the relationship of one variable to another. The dependent variable is generally shown on the vertical axis and the independent variable on the horizontal axis. A simple line graph shows the relationship of one dependent variable to the independent variable, whereas a multiple line graph shows the relationship of more than one dependent variable to the independent variable. The lines for each dependent variable should be distinguishable and clearly labeled. A second variation is the stratum chart which shows the composition of a total quantity and its changes as the independent variable changes.

3) **Bar charts**: Bar charts show changes in the dependent variable at discrete intervals of the independent variable. Common variants are the subdivided bar chart or the multiple bar charts. In each of these cases, each variable needs to be clearly identified.

Too much detail obscures the essential advantage of charts, which is to make relationships easy to grasp.

THE ORAL PRESENTATION

The conclusions and recommendations of most research reports will be presented orally as well as in writing. The **oral presentation** is a verbal summary of the major findings and conclusions, and the recommendations given to clients or line managers to provide them with the opportunity to clarify any ambiguous issues by asking questions.

The key to effective presentation is preparation. The researcher should select the three or four most important findings for emphasis and rely on the written report for full summary. The researcher needs to be ready to defend the results.

Another key to effective oral presentation is adaptation to the audience. Also, the presenter should plan to rely on brief notes, along with memory, and as much rehearsal as the occasion calls for.

Graphic aids and other visual aids can be useful in an oral presentation as in a written one.

Whatever medium is chosen, the visual aids should be designed to convey a simple, attention-getting message that supports a point on which the audience should focus its thinking.

Reports on the Internet

Many clients want many employees to have access to research findings. One easy way to share data is to have executive summaries and reports available on a company Intranet. Information technology exists so that a company can use the Internet to design questionnaires, administer surveys, analyze data, and share the results in a presentation-ready format. Real-time data capture allows for beginning-to-end reporting. A number of companies offer fully Web-based research management systems.

THE RESEARCH FOLLOW-UP

The **research follow-up** is a recontacting of decision makers and/or clients after they have had a chance to read over the research report. Its purpose is to determine if the researchers need to provide additional information or can clarify some issues that may concern management. The research follow-up can help researchers ensure the satisfaction of their customers.

THE WRITING PROCESS

Organizing the report; To a large degree, the organization flows from the earlier stages of the project. This is especially true of a good, well-organized project. Nevertheless, even after the best-run project has been completed, writing a report calls for consideration of how to organize the presentation effectively. The reader should be able to read the objectives, turn to the conclusion section, and find specific conclusions relative to each objective. The reader should also be able to turn to any section of the report and see the whole tied together by concern with one problem stated in terms of a small number of interrelated objectives. A unified theme helps the reader understand what has been accomplished.

If the research did not produce a conclusion on a point, the writer should say so. Since the report is written in fact, it can take all these factors into consideration and show how they are interrelated.

Good organization is achieved by outlining. If the writer plans ahead, looking at the relationships among the ideas and the order in which they should be presented, the report will usually be better organized.

1) **The outline:** The outline has two main functions: (1) to show the order of presentation and (2) to show the way in which the parts relate to one another, particularly how small parts go together to form larger ones. The first is shown by order on the page, the second by indentation of subordinate parts. There are also two major types of outline notation. The traditional method uses alternating letters and numbers to show levels of subordination, while the newer method uses a decimal system with successive places to the right of the decimal showing smaller and smaller levels of subordination. Attention to good practices in classification is useful in developing good outlines. First, at each level, use only one basis for organization and this basis should include all the topics to be discussed. The writer must decide which classification principle to use at which level.
2) **Writing the first draft:** A first draft can almost always be improved. Often it needs major work, but doing that work is much easier than writing the draft in the first place. Every research writer should memorize two maxims: “Easy writing is hard reading” and “Good writing comes from good rewriting.” Trying to produce a perfect job on the first draft means combining creation and criticism. Trying to create and criticize simultaneously makes it almost impossible to write anything. Rewriting a first draft is always easier than getting the first draft down. The key is to get the first draft down before that last-minute desperation - that is, suppressing the critical “internal editor” and releasing the creative side. Here are some suggestions:

a) Consolidate your time. Once started don’t stop.
b) If you have trouble getting started, set your pen on the paper and keep it going writing anything that hits your mind.
c) The introduction is one of the hardest parts to write, so start with a later section that you are very sure of.
d) Don’t keep shifting gears from writing to revising.
e) Many writers find it easier to “talk” the first draft into a tape recorder.

3) **Criticizing and rewriting:** The professional writer who does not make extensive revisions is a rarity. The key to effective revision is objectivity. In revision the task is to look at the work as a self-editor. One of the best ways to achieve objectivity is to put the first draft aside for a while before revising it. Even an hour or two will help you to gain a critical distance from the work. Another suggestion is to let someone look at it and encourage that person not to withhold suggestions for fear of offending you. How do you go about doing the revision? The revision inventory is an attempt to overcome the problem of vagueness. It presents details to check in a framework of four general criteria: ***readability, correctness, appropriateness, and thought***. These criteria may be considered one at a time in successive reviews of the report or all together, depending on what the individual writer finds most effective.