

HKBM7000

Approaches to Research and Academic Communications

Day 3 (28 June 2020)

Data Processing

Ethical Issues

[1]

Topics

- Secondary Data
- Measurement and Scaling
- Questionnaire Design
- Sampling and Data Collection
- Sample Size Determination
- Fieldwork
- Data Preparation
- Data Analysis Methods and Hypothesis Testing
- Results Presentation
- International Research
- Research Ethics

[2]

SECONDARY DATA

Primary Vs. Secondary Data

- **Primary data** are originated by a researcher for the specific purpose of addressing the problem at hand. The collection of primary data involves all six steps of the marketing research process (Chapter 1).
- **Secondary data** are data that have already been collected for purposes other than the problem at hand. These data can be located quickly and inexpensively.

Comparison of Primary & Secondary Data

	Primary Data	Secondary Data
Collection purpose	For the problem at hand	For other
Collection process	Very involved	Rapid & easy
Collection cost	High	Relatively low
Collection time	Long	Short



Uses of Secondary Data

- Identify the problem
- Better define the problem
- Develop an approach to the problem
- Formulate an appropriate research design (for example, by identifying the key variables)
- Answer certain research questions and test some hypotheses
- Interpret primary data more insightfully

Criteria for Evaluating Secondary Data

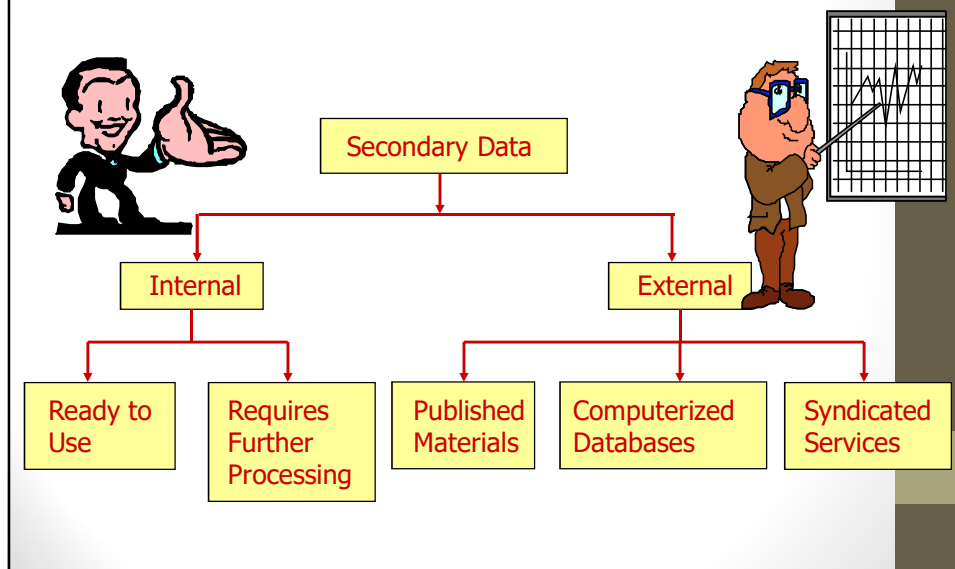
- **Specifications:** Methodology Used to Collect the Data
- **Error:** Accuracy of the Data
- **Currency:** When the Data Were Collected
- **Objective(s):** The Purpose for Which the Data Were Collected
- **Nature:** The Content of the Data
- **Dependability:** Overall, How Dependable Are the Data

Criteria for Evaluating Secondary Data

Table 4.2

Criteria	Issues	Remarks
Specifications & Methodology	Data collection method, response rate, quality & analysis of data, sampling technique & size, questionnaire design, fieldwork.	Data should be reliable, valid, & generalizable to the problem.
Error & Accuracy	Examine errors in approach, research design, sampling, data collection & analysis, & reporting.	Assess accuracy by comparing data from different sources.
Currency	Time lag between collection & publication, frequency of updates.	Census data are updated by syndicated firms.
Objective	Why were the data collected?	The objective determines the relevance of data.
Nature	Definition of key variables, units of measurement, categories used, relationships examined.	Reconfigure the data to increase their usefulness.
Dependability	Expertise, credibility, reputation, and trustworthiness of the source.	Data should be obtained from an original source.

A Classification of Secondary Data



Internal Secondary Data

Department Store Project

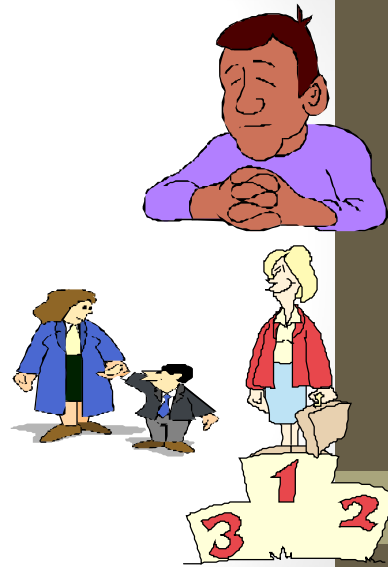
Sales were analyzed to obtain:

- Sales by product line
- Sales by major department (e.g., men's wear, house wares)
- Sales by specific stores
- Sales by geographical region
- Sales by cash versus credit purchases
- Sales in specific time periods
- Sales by size of purchase
- Sales trends in many of these classifications were also examined

Type of Individual/Household Level Data Available from Syndicated Firms

I. Demographic Data

- Identification (name, address, email, telephone)
- Sex
- Marital status
- Names of family members
- Age (including ages of family members)
- Income
- Occupation
- Number of children present
- Home ownership
- Length of residence
- Number and make of cars owned



Type of Individual/Household Level Data Available from Syndicated Firms

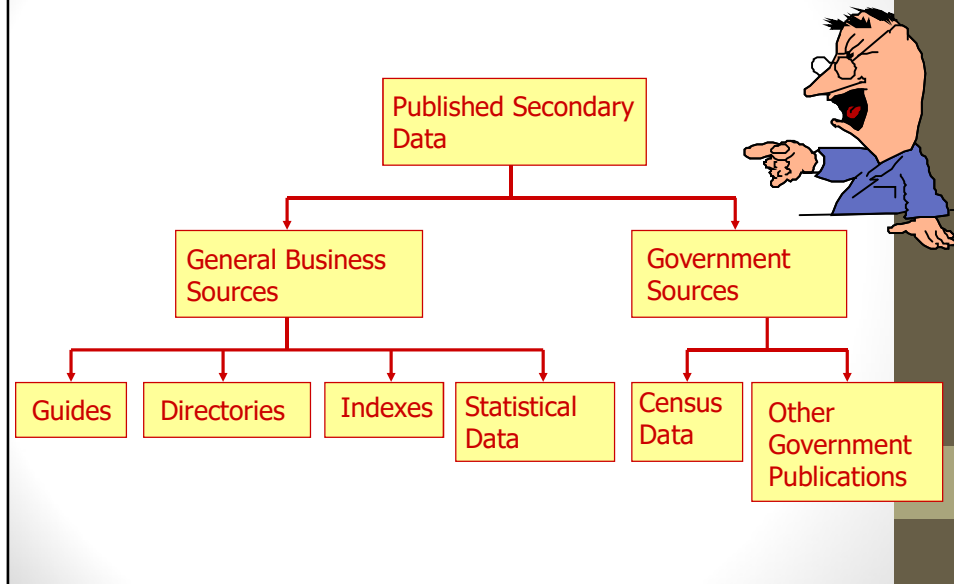
II. Psychographic Lifestyle Data

- Interest in golf
- Interest in snow skiing
- Interest in book reading
- Interest in running
- Interest in bicycling
- Interest in pets
- Interest in fishing
- Interest in electronics
- Interest in cable television



There are also firms such as Dun & Bradstreet and American Business Information which collect demographic data on businesses.

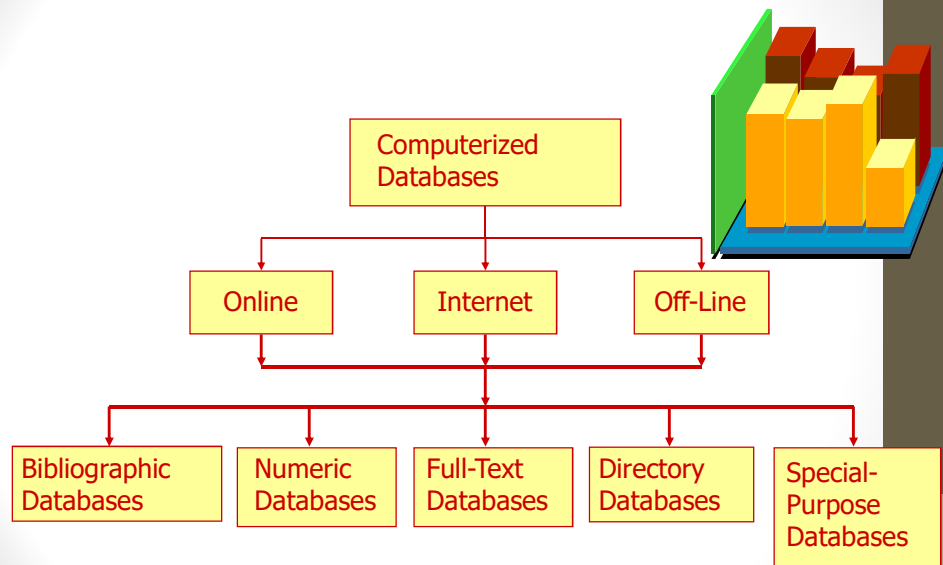
A Classification of Published Secondary Sources



InfoUSA: Here, There, Everywhere

InfoUSA (www.infousa.com) markets subsets of its data in a number of forms, including the professional online services (LEXIS-NEXIS and DIALOG), the general online services (CompuServe and Microsoft Network), the Internet (look-ups), and on CD-ROM. The underlying database on which all these products are based contains information on 115 million residential listings and 14 million business listings, as of 2005. These are verified with over 17 million phone calls annually. The products derived from these databases include sales leads, mailing lists, business directories, mapping products, and also delivery of data on the Internet.

A Classification of Computerized Databases



Big Data

Published External Secondary Sources

Guides

- An excellent source of standard or recurring information
- Helpful in identifying other important sources of directories, trade associations, and trade publications
- One of the first sources a researcher should consult

Directories

- Helpful for identifying individuals or organizations that collect specific data
- Examples: Consultants and Consulting Organizations Directory, Encyclopedia of Associations, FINDEX: The Directory of Market Research Reports, Studies and Surveys, and Research Services Directory

Indices

- Helpful in locating information on a particular topic in several different publications

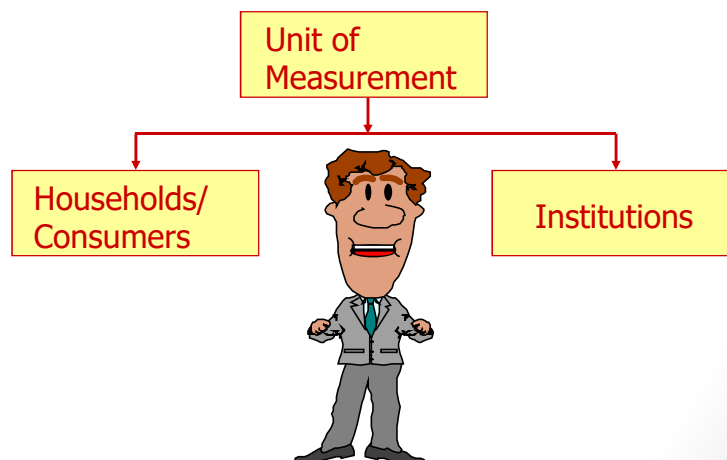
Classification of Computerized Databases

- **Bibliographic databases** are composed of citations to articles
- **Numeric databases** contain numerical and statistical information
- **Full-text databases** contain the complete text of the source documents comprising the database
- **Directory databases** provide information on individuals, organizations, and services
- **Special-purpose databases** provide specialized information

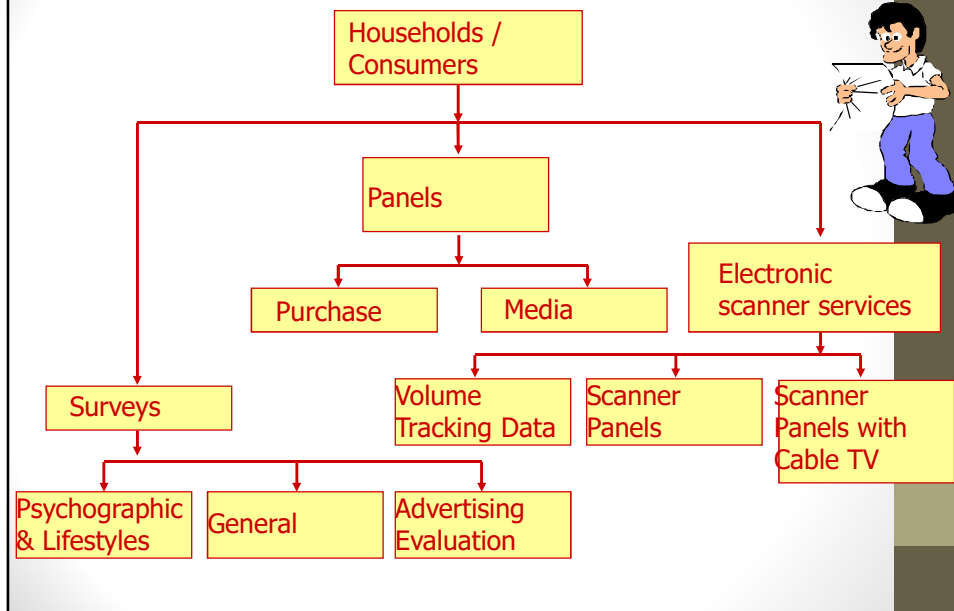
Syndicated Services

- **Syndicated services** are companies that collect and sell common pools of data of known commercial value designed to serve a number of clients
- Syndicated sources can be classified based on the unit of measurement (households/consumers or institutions)
- Household/consumer data may be obtained from surveys, diary panels, or electronic scanner services
- Institutional data may be obtained from retailers, wholesalers, or industrial firms

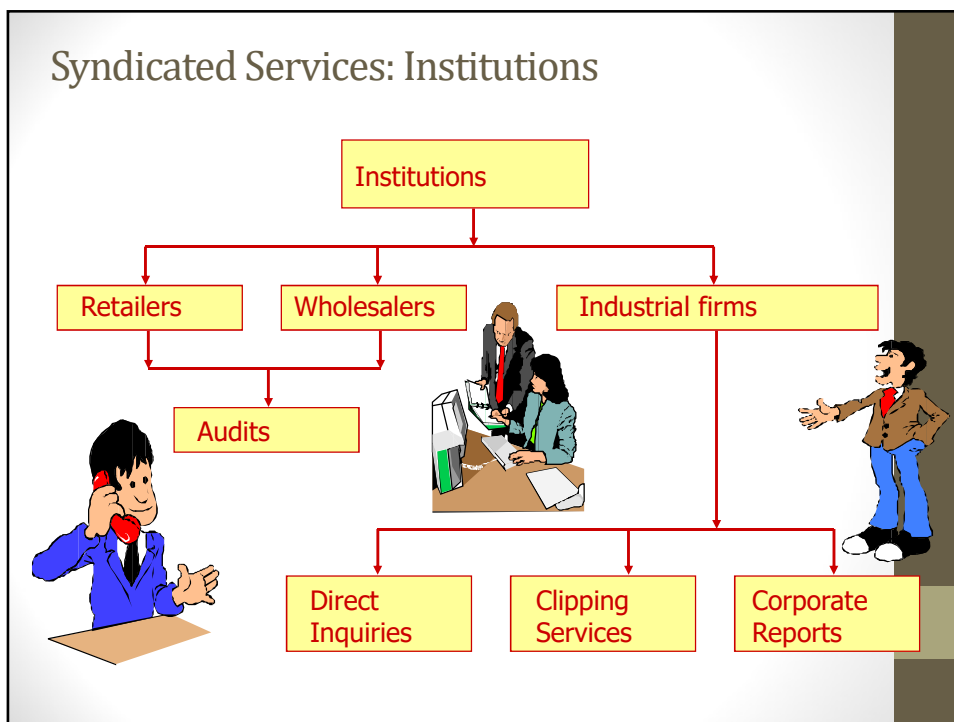
A Classification of Syndicated Services



Syndicated Services: Consumers



Syndicated Services: Institutions



Overview of Syndicated Services



Type	Characteristics	Advantages	Disadvantages	Uses
Surveys	Surveys conducted at regular intervals	Most flexible way of obtaining data; information on underlying motives	Interviewer errors; respondent errors	Market segmentation, advertising theme selection and advertising effectiveness
Purchase Panels	Households provide specific information regularly over an extended period of time; respondent asked to record specific behaviors as they occur	Recorded purchase behavior can be linked to the demographic/psychographic characteristics	Lack of representativeness; response bias; maturation	Forecasting sales, market share and trends; establishing consumer profiles, brand loyalty and switching; evaluating test markets, advertising, and distribution
Media Panels	Electronic devices automatically recording behavior, supplemented by a diary	Same as purchase panel	Same as purchase panel	Establishing advertising rates; selecting media program or air time; establishing viewer profiles

Overview of Syndicated Services

Scanner Diary Panels with Cable TV	Scanner panels of households that subscribe to cable TV	Data reflect actual purchases; sample control; ability to link panel data to household characteristics	Data may not be representative; quality of data limited	Promotional mix analyses; copy testing; new product testing; positioning
Audit services	Verification of product movement by examining physical records or performing inventory analysis	Relatively precise information at the retail and wholesale levels	Coverage may be incomplete; matching of data on competitive activity may be difficult	Measurement of consumer sales and market share; competitive activity; analyzing distribution patterns; tracking of new products
Industrial Product Syndicated Services	Data banks on industrial establishments created through direct inquiries of companies, clipping services, and corporate reports	Important source of information on industrial firms, particularly useful in initial phases of the projects	Data are lacking in terms of content, quantity, and quality	Determining market potential by geographic area; defining sales territories; allocating advertising budget

Single-Source Data

Single-source data provide integrated information on household variables, including media consumption and purchases, and marketing variables, such as product sales, price, advertising, promotion, and in-store marketing effort

- Recruit a test panel of households and meter each home's TV sets
- Survey households periodically on what they read
- Grocery purchases are tracked by UPC scanners
- Track retail data, such as sales, advertising, and promotion

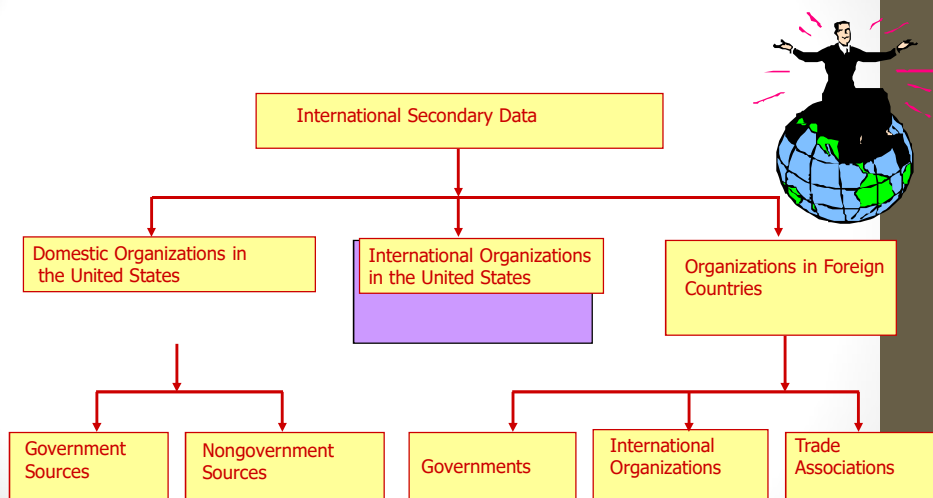
The NYT on the Web: A New Way To Target Customers

To handle alternate forms of interaction and updates, The New York Times created a separate unit, The New York Times Electronic Media Co. *The New York Times* on the Web (www.nytimes.com) has drawn over 11.4 million national unique users as of October 2005. The database contains demographic information, such as age, gender, income, and zip code, that ties to an e-mail address for each of the members. This new database marketing system can identify and customize user groups, target Web messages to specific segments of the population, and adjust the message based on audience reaction. It can also increase targeting opportunities through third-party data or additional information supplied by the user.

The NYT on the Web: A New Way To Target Customers

For example, the database enables an automobile firm to emphasize safety to older customers, luxury to affluent ones, and roominess to families. The system is set up so that near real-time data can be received from the Web that indicates how well ads are performing relative to age, gender, and income characteristics. Thus, this system allows a firm to maintain up-to-date information on audiences in order to position its products effectively.

A Classification of International Sources



MEASUREMENT AND SCALING

Measurement and Scaling

Measurement means assigning numbers or other symbols to characteristics of objects according to certain pre-specified rules.

- One-to-one correspondence between the numbers and the characteristics being measured.
- The rules for assigning numbers should be standardized and applied uniformly.
- Rules must not change over objects or time.

Scale Characteristics

Description

By description, we mean the unique labels or descriptors that are used to designate each value of the scale. All scales possess description.

Order

By order, we mean the relative sizes or positions of the descriptors. Order is denoted by descriptors such as greater than, less than, and equal to.

Scale Characteristics

Distance

The characteristic of distance means that absolute differences between the scale descriptors are known and may be expressed in units.

Origin

The origin characteristic means that the scale has a unique or fixed beginning or true zero point.

Measurement and Scaling

Scaling involves creating a continuum upon which measured objects are located.

Consider an attitude scale from 1 to 100. Each respondent is assigned a number from 1 to 100, with 1 = Extremely Unfavorable, and 100 = Extremely Favorable. Measurement is the actual assignment of a number from 1 to 100 to each respondent. Scaling is the process of placing the respondents on a continuum with respect to their attitude toward department stores.

Primary Scales of Measurement

Scale				
Nominal	Numbers Assigned to Runners			 Finish
Ordinal	Rank Order of Winners	 Third place	 Second place	 First place Finish
Interval	Performance Rating on a 0 to 10 Scale	8.2	9.1	9.6
Ratio	Time to Finish in Seconds	15.2	14.1	13.4

Primary Scales of Measurement

Nominal Scale

- The numbers serve only as labels or tags for identifying and classifying objects.
- When used for identification, there is a strict one-to-one correspondence between the numbers and the objects.
- The numbers do not reflect the amount of the characteristic possessed by the objects.
- The only permissible operation on the numbers in a nominal scale is counting.
- Only a limited number of statistics, all of which are based on frequency counts, are permissible, e.g., percentages, and mode.

Primary Scales of Measurement

Ordinal Scale

- A ranking scale in which numbers are assigned to objects to indicate the relative extent to which the objects possess some characteristic.
- Can determine whether an object has more or less of a characteristic than some other object, but not how much more or less.
- Any series of numbers can be assigned that preserves the ordered relationships between the objects.
- In addition to the counting operation allowable for nominal scale data, ordinal scales permit the use of statistics based on centiles, e.g., percentile, quartile, median.

Primary Scales of Measurement

Interval Scale

- Numerically equal distances on the scale represent equal values in the characteristic being measured.
- It permits comparison of the differences between objects.
- The location of the zero point is not fixed. Both the zero point and the units of measurement are arbitrary.
- Any positive linear transformation of the form $y = a + bx$ will preserve the properties of the scale.
- It is not meaningful to take ratios of scale values.
- Statistical techniques that may be used include all of those that can be applied to nominal and ordinal data, and in addition the arithmetic mean, standard deviation, and other statistics commonly used in marketing research.

Primary Scales of Measurement

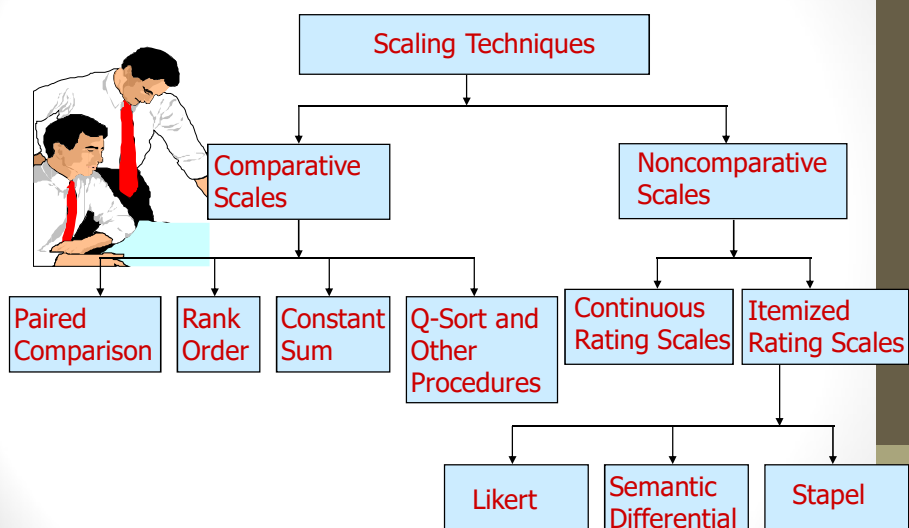
Ratio Scale

- Possesses all the properties of the nominal, ordinal, and interval scales.
- It has an absolute zero point.
- It is meaningful to compute ratios of scale values.
- Only proportionate transformations of the form $y = bx$, where b is a positive constant, are allowed.
- All statistical techniques can be applied to ratio data.

Primary Scales of Measurement

Scale	Basic Characteristics	Common Examples	Marketing Examples	Permissible Statistics	
				Descriptive	Inferential
Nominal	Numbers identify & classify objects	Social Security nos., numbering of football players	Brand nos., store types	Percentages, mode	Chi-square, binomial test
Ordinal	Nos. indicate the relative positions of objects but not the magnitude of differences between them	Quality rankings, rankings of teams in a tournament	Preference rankings, market position, social class	Percentile, median	Rank-order correlation, Friedman ANOVA
Interval	Differences between objects	Temperature (Fahrenheit)	Attitudes, opinions, index	Range, mean, standard	Product-moment
Ratio	Zero point is fixed, ratios of scale values can be compared	Length, weight	Age, sales, income, costs	Geometric mean, harmonic mean	Coefficient of variation

A Classification of Scaling Techniques



A Comparison of Scaling Techniques

- **Comparative scales** involve the direct comparison of stimulus objects. Comparative scale data must be interpreted in relative terms and have only ordinal or rank order properties.
- In **noncomparative scales**, each object is scaled independently of the others in the stimulus set. The resulting data are generally assumed to be interval or ratio scaled.

Relative Advantages of Comparative Scales

- Small differences between stimulus objects can be detected.
- Same known reference points for all respondents.
- Easily understood and can be applied.
- Involve fewer theoretical assumptions.
- Tend to reduce halo or carryover effects from one judgment to another.

Relative Disadvantages of Comparative Scales

- Ordinal nature of the data
- Inability to generalize beyond the stimulus objects scaled.

Comparative Scaling Techniques

Paired Comparison Scaling

- A respondent is presented with two objects and asked to select one according to some criterion.
- The data obtained are ordinal in nature.
- Paired comparison scaling is the most widely used comparative scaling technique.
- With n brands, $[n(n - 1) / 2]$ paired comparisons are required.
- Under the assumption of transitivity, it is possible to convert paired comparison data to a rank order.

Obtaining Shampoo Preferences Using Paired Comparisons

Instructions: We are going to present you with ten pairs of shampoo brands. For each pair, please indicate which one of the two brands of shampoo you would prefer for personal use.

Recording Form:



	Jhirmack	Finesse	Vidal Sassoon	Head & Shoulders	Pert
Jhirmack		0	0	1	0
Finesse	1 ^a		0	1	0
Vidal Sassoon	1	1		1	1
Head & Shoulders	0	0	0		0
Pert	1	1	0	1	
Number of Times Preferred ^b	3	2	0	4	1

^aA 1 in a particular box means that the brand in that column was preferred over the brand in the corresponding row. A 0 means that the row brand was preferred over the column brand. ^bThe number of times a brand was preferred is obtained by summing the 1s in each column.

Paired Comparison Selling

The most common method of taste testing is paired comparison. The consumer is asked to sample two different products and select the one with the most appealing taste. The test is done in private and a minimum of 1,000 responses is considered an adequate sample. A blind taste test for a soft drink, where imagery, self-perception and brand reputation are very important factors in the consumer's purchasing decision, may not be a good indicator of performance in the marketplace. The introduction of New Coke illustrates this point. New Coke was heavily favored in blind paired comparison taste tests, but its introduction was less than successful, because image plays a major role in the purchase of Coke.

A paired comparison
Blind taste test



Comparative Scaling Techniques

Rank Order Scaling

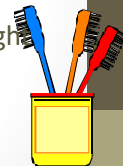
- Respondents are presented with several objects simultaneously and asked to order or rank them according to some criterion.
- It is possible that the respondent may dislike the brand ranked 1 in an absolute sense.
- Furthermore, rank order scaling also results in ordinal data.
- Only $(n - 1)$ scaling decisions need be made in rank order scaling.

Preference for Toothpaste Brands Using Rank Order Scaling

Instructions: Rank the various brands of toothpaste in order of preference. Begin by picking out the one brand that you like most and assign it a number 1. Then find the second most preferred brand and assign it a number 2. Continue this procedure until you have ranked all the brands of toothpaste in order of preference. The least preferred brand should be assigned a rank of 10.

No two brands should receive the same rank number.

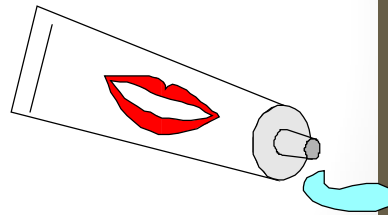
The criterion of preference is entirely up to you. There is no right or wrong answer. Just try to be consistent.



Preference for Toothpaste Brands Using Rank Order Scaling (cont.)

Form

<u>Brand</u>	<u>Rank Order</u>
1. Crest	_____
2. Colgate	_____
3. Aim	_____
4. Gleem	_____
5. Sensodyne	_____
6. Ultra Brite	_____
7. Close Up	_____
8. Pepsodent	_____
9. Plus White	_____
10. Stripe	_____



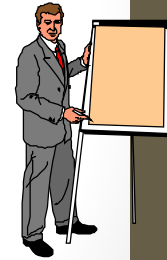
Comparative Scaling Techniques Constant Sum Scaling

- Respondents allocate a constant sum of units, such as 100 points to attributes of a product to reflect their importance.
- If an attribute is unimportant, the respondent assigns it zero points.
- If an attribute is twice as important as some other attribute, it receives twice as many points.
- The sum of all the points is 100. Hence, the name of the scale.

Importance of Bathing Soap Attributes Using a Constant Sum Scale

Instructions

On the next slide, there are eight attributes of bathing soaps. Please allocate 100 points among the attributes so that your allocation reflects the relative importance you attach to each attribute. The more points an attribute receives, the more important the attribute is. If an attribute is not at all important, assign it zero points. If an attribute is twice as important as some other attribute, it should receive twice as many points.



Importance of Bathing Soap Attributes Using a Constant Sum Scale (cont.)

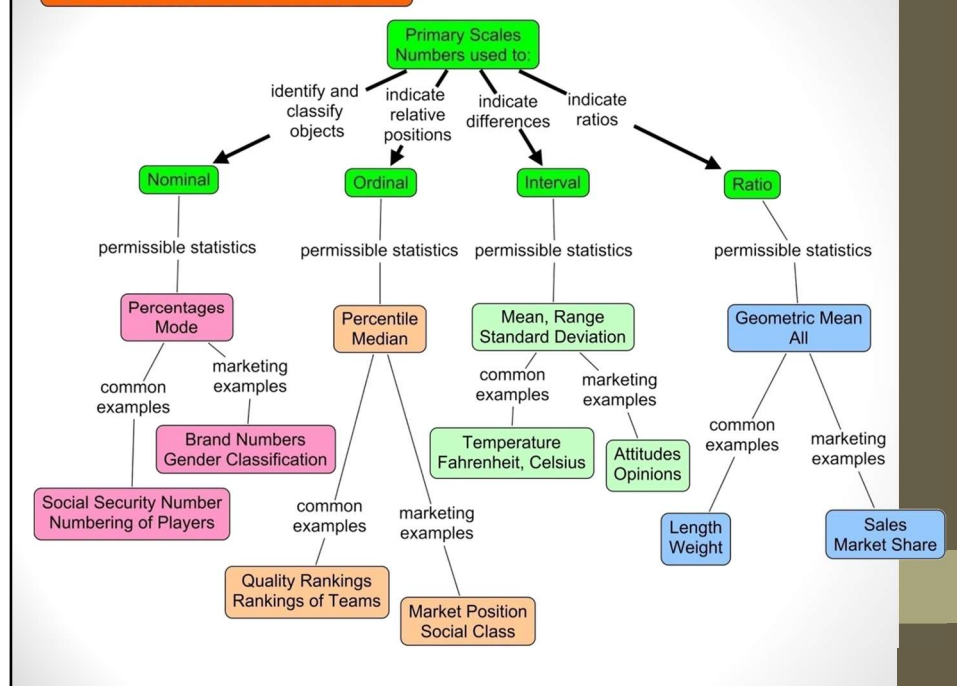
Form

Average Responses of Three Segments

Attribute	Segment I	Segment II	Segment III
1. Mildness	8	2	4
2. Lather	2	4	17
3. Shrinkage	3	9	7
4. Price	53	17	9
5. Fragrance	9	0	19
6. Packaging	7	5	9
7. Moisturizing	5	3	20
8. Cleaning Power	13	60	15
Sum	100	100	100



Figure 8.7 A Concept Map for Primary Scales



Noncomparative Scaling Techniques

- Respondents evaluate only one object at a time, and for this reason **non-comparative scales** are often referred to as monadic scales.
- Non-comparative techniques consist of **continuous** and **itemized rating scales**.

Continuous Rating Scale

Respondents rate the objects by placing a mark at the appropriate position on a line that runs from one extreme of the criterion variable to the other.

The form of the continuous scale may vary considerably.

How would you rate Sears as a department store?

Version 1

Probably the worst -----|----- Probably the best

Version 2

Probably the worst -----|----- Probably the best

0 10 20 30 40 50 60 70 80 90 100

Version 3

Very bad

Neither good
nor bad

Very good

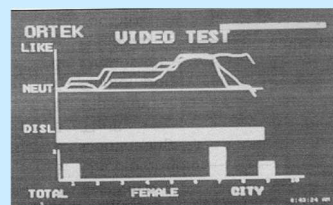
Probably the worst -----|----- Probably the best

0 10 20 30 40 50 60 70 80 90 100

Perception Analyzer

A relatively new research tool, the perception analyzer, provides continuous measurement of "gut reaction." A group of up to 400 respondents is presented with TV or radio spots or advertising copy. The measuring device consists of a dial that contains a 100-point range. Each participant is given a dial and instructed to continuously record his or her reaction to the material being tested.

As the respondents turn the dials, the information is fed to a computer, which tabulates second-by-second response profiles. As the results are recorded by the computer, they are superimposed on a video screen, enabling the researcher to view the respondents' scores immediately. The responses are also stored in a permanent data file for use in further analysis. The response scores can be broken down by categories, such as age, income, sex, or product usage.



Itemized Rating Scales

- The respondents are provided with a scale that has a number or brief description associated with each category.
- The categories are ordered in terms of scale position, and the respondents are required to select the specified category that best describes the object being rated.
- The commonly used itemized rating scales are the Likert, semantic differential, and Stapel scales.

Likert Scale

The **Likert scale** requires the respondents to indicate a degree of agreement or disagreement with each of a series of statements about the stimulus objects.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1. Sears sells high-quality merchandise.	1	2X	3	4	5
2. Sears has poor in-store service.	1	2X	3	4	5
3. I like to shop at Sears.	1	2	3X	4	5

- The analysis can be conducted on an item-by-item basis (profile analysis), or a total (summated) score can be calculated.
- When arriving at a total score, the categories assigned to the negative statements by the respondents should be scored by reversing the scale.

Semantic Differential Scale

The **semantic differential** is a seven-point rating scale with end points associated with bipolar labels that have semantic meaning.

SEARS IS:

Powerful --:--:--:--X--:--: Weak

Unreliable --:--:--:--X--:--: Reliable

Modern --:--:--:--:--X--: Old-fashioned

- The negative adjective or phrase sometimes appears at the left side of the scale and sometimes at the right.
- This controls the tendency of some respondents, particularly those with very positive or very negative attitudes, to mark the right- or left-hand sides without reading the labels.
- Individual items on a semantic differential scale may be scored on either a -3 to +3 or a 1 to 7 scale.

A Semantic Differential Scale for Measuring Self- Concepts, Person Concepts, and Product Concepts

1) Rugged	:--:--:--:--:--X--:--: Delicate
2) Excitable	:--:--:--:--:--X--:--: Calm
3) Uncomfortable	:--:--:--:--:--X--:--: Comfortable
4) Dominating	:--:--:--:--:--X--:--: Submissive
5) Thrifty	:--:--:--:--:--X--:--: Indulgent
6) Pleasant	:--:--:--:--:--X--:--: Unpleasant
7) Contemporary	:--:--:--:--:--X--:--: Obsolete
8) Organized	:--:--:--:--:--X--:--: Unorganized
9) Rational	:--:--:--:--:--X--:--: Emotional
10) Youthful	:--:--:--:--:--X--:--: Mature
11) Formal	:--:--:--:--:--X--:--: Informal
12) Orthodox	:--:--:--:--:--X--:--: Liberal
13) Complex	:--:--:--:--:--X--:--: Simple
14) Colorless	:--:--:--:--:--X--:--: Colorful
15) Modest	:--:--:--:--:--X--:--: Vain



Stapel Scale

The **Stapel scale** is a unipolar rating scale with ten categories numbered from -5 to +5, without a neutral point (zero). This scale is usually presented vertically.

SEARS	
+5	+5
+4	+4
+3	+3
+2	+2X
+1	+1
HIGH QUALITY	POOR SERVICE
-1	-1
-2	-2
-3	-3
-4X	-4
-5	-5

The data obtained by using a Stapel scale can be analyzed in the same way as semantic differential data.

Basic Noncomparative Scales

Scale	Basic Characteristics	Examples	Advantages	Disadvantages
Continuous Rating Scale	Place a mark on a continuous line	Reaction to TV commercials	Easy to construct	Scoring can be cumbersome unless computerized
Itemized Rating Scales				
Likert Scale	Degrees of agreement on a 1 (strongly disagree) to 5 (strongly agree) scale	Measurement of attitudes	Easy to construct, administer, and understand	More time-consuming
Semantic Differential	Seven - point scale with bipolar labels	Brand, product, and company images	Versatile	Controversy as to whether the data are interval
Stapel Scale	Unipolar ten - point scale, - 5 to +5, without a neutral point (zero)	Measurement of attitudes and images	Easy to construct, administer over telephone	Confusing and difficult to apply

Summary of Itemized Scale Decisions

- | | |
|--------------------------------------|--|
| 1) Number of categories | Although there is no single, optimal number, traditional guidelines suggest that there should be between five and nine categories |
| 2) Balanced vs. unbalanced | In general, the scale should be balanced to obtain objective data |
| 3) Odd/even no. of categories | If a neutral or indifferent scale response is possible for at least some respondents, an odd number of categories should be used |
| 4) Forced vs. non-forced | In situations where the respondents are expected to have no opinion, the accuracy of the data may be improved by a non-forced scale |
| 5) Verbal description | An argument can be made for labeling all or many scale categories. The category descriptions should be located as close to the response categories as possible |
| 6) Physical form | A number of options should be tried and the best selected |

Balanced and Unbalanced Scales

Balanced Scale

Jovan Musk for Men is:

Extremely good _____

Very good _____

Good _____

Bad _____

Very bad _____

Extremely bad _____

Unbalanced Scale

Jovan Musk for Men is:

Extremely good _____

Very good _____

Good _____

Somewhat good _____

Bad _____

Very bad _____



Rating Scale Configurations

Cheer detergent is:

1) Very harsh --- --- --- --- --- --- --- Very gentle

2) Very harsh 1 2 3 4 5 6 7 Very gentle

3) . Very harsh
.
.
Neither harsh nor gentle
.
.
Very gentle

4)

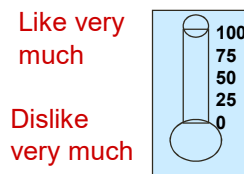
Very harsh	Harsh	Somewhat harsh	Neither harsh nor gentle	Somewhat gentle	Gentle	Very gentle
-3	-2	-1	0	+1	+2	+3
Very harsh			Neither harsh nor gentle			Very gentle



Some Unique Rating Scale Configurations

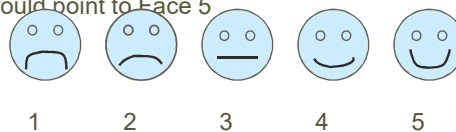
Thermometer Scale

Instructions: Please indicate how much you like McDonald's hamburgers by coloring in the thermometer. Start at the bottom and color up to the temperature level that best indicates how strong your preference is.



Smiling Face Scale

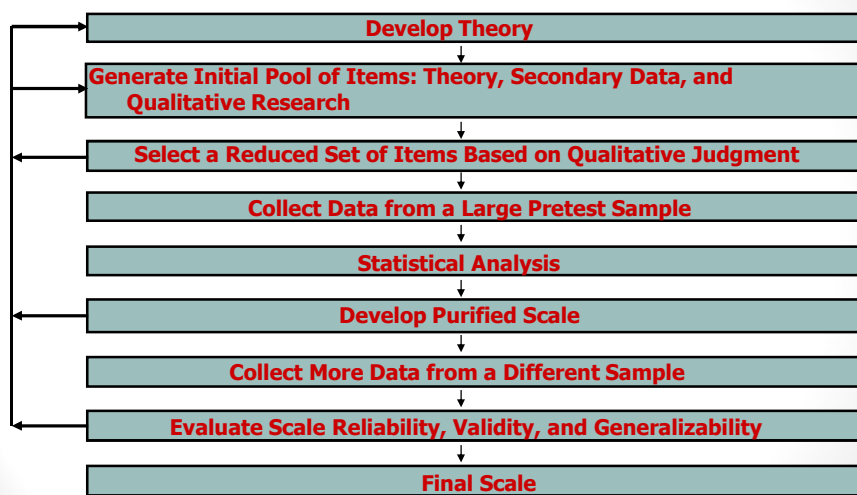
Instructions: Please point to the face that shows how much you like the Barbie Doll. If you do not like the Barbie Doll at all, you would point to Face 1. If you liked it very much, you would point to Face 5.



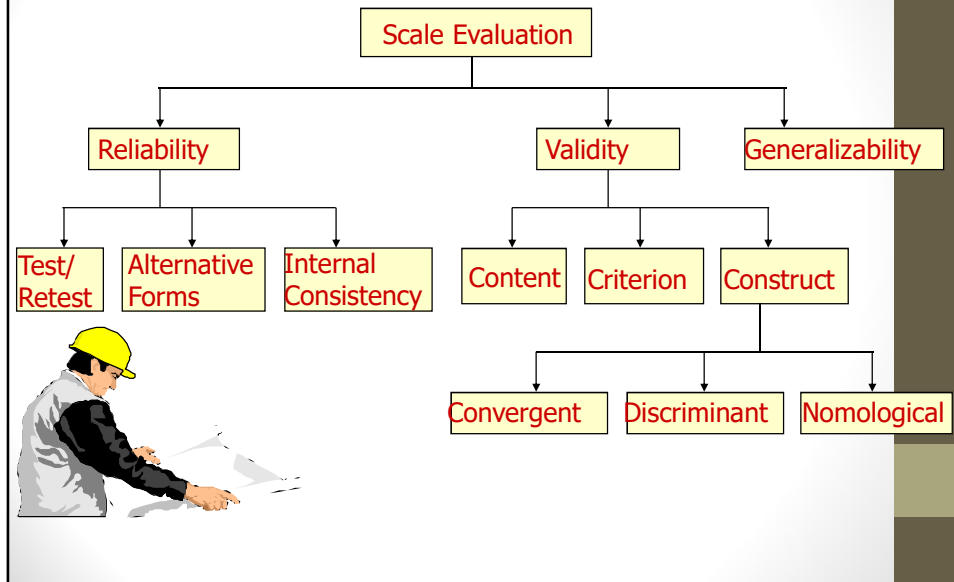
Some Commonly Used Scales in Marketing

CONSTRUCT	SCALE DESCRIPTORS				
Attitude	Very Bad	Bad	Neither Bad Nor Good	Good	Very Good
Importance	Not at All Important	Not Important	Neutral	Important	Very Important
Satisfaction	Very Dissatisfied	Dissatisfied	Neither Dissat. Nor Satisfied	Satisfied	Very Satisfied
Purchase Intent	Definitely will Not Buy	Probably Will Not Buy	Might or Might Not Buy	Probably Will Buy	Definitely Will Buy
Purchase Freq	Never	Rarely	Sometimes	Often	Very Often

Development of a Multi-item Scale



Scale Evaluation



Measurement Accuracy

The **true score model** provides a framework for understanding the accuracy of measurement.

$$X_O = X_T + X_S + X_R$$

where

X_O = the observed score or measurement

X_T = the true score of the characteristic

X_S = systematic error

X_R = random error

Potential Sources of Error on Measurement

- 1) Other relatively stable characteristics of the individual that influence the test score, such as intelligence, social desirability, and education.
- 2) Short-term or transient personal factors, such as health, emotions, and fatigue.
- 3) Situational factors, such as the presence of other people, noise, and distractions.
- 4) Sampling of items included in the scale: addition, deletion, or changes in the scale items.
- 5) Lack of clarity of the scale, including the instructions or the items themselves.
- 6) Mechanical factors, such as poor printing, overcrowding items in the questionnaire, and poor design.
- 7) Administration of the scale, such as differences among interviewers.
- 8) Analysis factors, such as differences in scoring and statistical analysis.

Reliability

- **Reliability** can be defined as the extent to which measures are free from random error, X_R . If $X_R = 0$, the measure is perfectly reliable.
- In **test-retest reliability**, respondents are administered identical sets of scale items at two different times and the degree of similarity between the two measurements is determined.
- In **alternative-forms reliability**, two equivalent forms of the scale are constructed and the same respondents are measured at two different times, with a different form being used each time.

Reliability

- **Internal consistency reliability** determines the extent to which different parts of a summated scale are consistent in what they indicate about the characteristic being measured.
- In **split-half reliability**, the items on the scale are divided into two halves and the resulting half scores are correlated.
- The **coefficient alpha**, or Cronbach's alpha, is the average of all possible split-half coefficients resulting from different ways of splitting the scale items. This coefficient varies from 0 to 1, and a value of 0.6 or less generally indicates unsatisfactory internal consistency reliability.

Types of Indicators

- Formative indicators
- Reflective indicators
- Implication for reliability of the indicator

Scale Attenuation

- Collectively to ceiling and floor effects, where performance is at an extreme of the measurement scale
- Both types of effects pose problems for determining whether or not the dependent measure reflects performance limits
- Ceiling effects
 - When performance is at the upper limit of the scale for the dependent measure
 - What is the typical shape of the distribution? Normal? Skewed?
- Floor effects
 - When performance is at the lower limit of the scale for the dependent measure

Regression Artifacts (regression to the mean)

- Can anyone explain what is *regression to the mean*?
- An unwanted effect in which initial, extreme scores along a dependent measure tend to approach average scores upon further testing
- It is particularly a problem when assignment to conditions is based upon a single measure of performance
- Example
 - http://www2.warwick.ac.uk/newsandevents/pressreleases/extreme_statistics/figure-2.jpg

Validity

- The **validity** of a scale may be defined as the extent to which differences in observed scale scores reflect true differences among objects on the characteristic being measured, rather than systematic or random error. Perfect validity requires that there be no measurement error ($X_O = X_T$, $X_R = 0$, $X_S = 0$).
- **Content validity** is a subjective but systematic evaluation of how well the content of a scale represents the measurement task at hand.
- **Criterion validity** reflects whether a scale performs as expected in relation to other variables selected (criterion variables) as meaningful criteria.

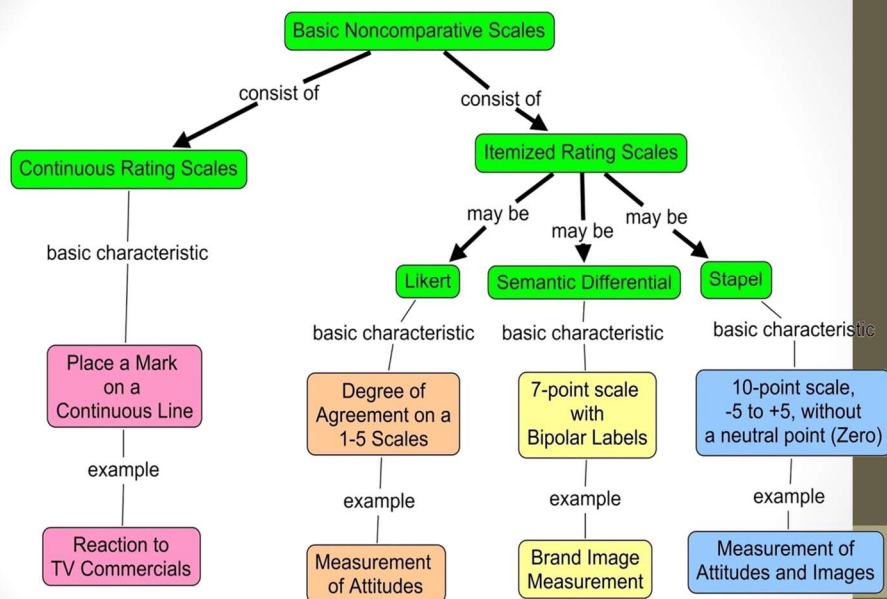
Validity

- **Construct validity** addresses the question of what construct or characteristic the scale is, in fact, measuring. Construct validity includes convergent, discriminant, and nomological validity.
- **Convergent validity** is the extent to which the scale correlates positively with other measures of the same construct.
- **Discriminant validity** is the extent to which a measure does not correlate with other constructs from which it is supposed to differ.
- **Nomological validity** is the extent to which the scale correlates in theoretically predicted ways with measures of different but related constructs.

Relationship Between Reliability and Validity

- If a measure is perfectly valid, it is also perfectly reliable. In this case $X_O = X_T$, $X_R = 0$, and $X_S = 0$.
- If a measure is unreliable, it cannot be perfectly valid, since at a minimum $X_O = X_T + X_R$. Furthermore, systematic error may also be present, i.e., $X_S \neq 0$. Thus, unreliability implies invalidity.
- If a measure is perfectly reliable, it may or may not be perfectly valid, because systematic error may still be present ($X_O = X_T + X_S$).
- Reliability is a necessary, but not sufficient, condition for validity.

Figure 9.8 A Concept Map for Noncomparative Scales



QUESTIONNAIRE DESIGN

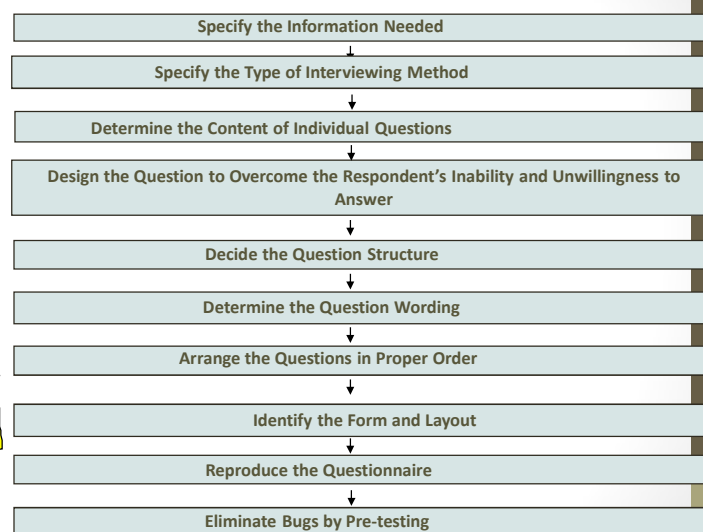
Questionnaire Definition

- A **questionnaire** is a formalized set of questions for obtaining information from respondents.

Questionnaire Objectives

- It must translate the information needed into a set of specific questions that the respondents can and will answer.
- A questionnaire must uplift, motivate, and encourage the respondent to become involved in the interview, to cooperate, and to complete the interview.
- A questionnaire should minimize response error.

Questionnaire Design Process



Biases in Scales

- Central tendency bias (no extreme scores)
 - Use 5-7 item ratings
- Acquiescence bias (agreeing with item)
 - Have reverse scored items
- Social desirability bias
 - Have social desirability items
 - “I never lie”

[85]

Choosing a response format

- Open ended
 - Imagine you were buying a new car. Describe your decision making process
- Categorical
 - When buying a new car, I would ask my friends about their cars.
 - Yes – No
 - Correct – Incorrect – Neither – Don't Know

[86]

Choosing a response format

- Rating Scales

- Likert (measuring attitude)
- Balance positive and negative
- Equal distances between items
- Often treated as interval scales

1. It is the duty of doctors to keep people alive for as long as possible.

- ☐ Strongly Agree
- ☐ Agree
- ☐ Agree somewhat
- ☐ Undecided
- ☐ Disagree somewhat
- ☐ Disagree
- ☐ Strongly disagree

Some 5-Item Scales Balanced Positive and Negative

Equal
Distances??

Very Interested 5	Somewhat Interested 4	Neutral 3	Not Very Interested 2	Not at All Interested 1
Very Much 5	Somewhat 4	Undecided 3	Not Really 2	Not at All 1
Very Much Like Me 5	Somewhat Like Me 4	Neutral 3	Not Much Like Me 2	Not at All Like Me 1
Very Happy 5	Somewhat Happy 4	Neutral 3	Not Very Happy 2	Not at All Happy 1
Almost Always 5	Sometimes 4	Every Once In a While 3	Rarely 2	Never 1

Marking only endpoints: more equal distances

Early Clinical and Professional Development A11PD1

	Low				High	
1.	☐	☐	☐	☐	☐	The usefulness of the module overall
2.	☐	☐	☐	☐	☐	The objectives of the module were made clear
3.	☐	☐	☐	☐	☐	Rate the usefulness of Skills Centre practical classes in helping you attain the module objectives
4.	☐	☐	☐	☐	☐	Rate the usefulness of lectures in helping you attain the module objectives
5.	☐	☐	☐	☐	☐	Rate the usefulness of GP visits in helping you attain the module objectives
6.	☐	☐	☐	☐	☐	Rate the usefulness of the Workbook & Guide in helping you attain the module objectives

[89]

Even number of rating options: Forced choice scale

FIGURE 6.1 The Rosenberg Self-Esteem Scale

Circle one response for each of the following ten items.

	Strongly Agree	Agree	Disagree	Strongly Disagree
1. I feel that I am a person of worth, at least on an equal basis with others.	1	2	3	4
2. I feel that I have a number of good qualities.	1	2	3	4
*3. All in all, I am inclined to feel that I am a failure.	1	2	3	4
4. I am able to do things as well as most other people.	1	2	3	4
*5. I feel I do not have much to be proud of.	1	2	3	4
6. I take a positive attitude toward myself.	1	2	3	4
7. On the whole, I am satisfied with myself.	1	2	3	4
*8. I wish I could have more respect for myself.	1	2	3	4
*9. I certainly feel useless at times.	1	2	3	4
*10. At times I think I am no good at all.	1	2	3	4

*Items marked with an asterisk have reversed wording. The numbers on items with reversed wording should be reversed before summing the responses for the ten items. For example, on item 3, "strongly agree" becomes 4, "agree" becomes 3, "disagree" becomes 2, and "strongly disagree" becomes 1.

Source: Morris Rosenberg's "Self-Esteem Scale" from pp. 325–327 of *Society and Adolescent Self-Image*. © 1989 by Morris Rosenberg. Wesleyan University Press.

[90]

Writing good items

- Clear and simple
- Informal language
- Questions with variance
- Avoid negatives
- Multiple questions for a construct
- Avoid leading questions/demand characteristics

[91]

How many items?

- Content validity
 - Cover all relevant categories
 - e.g., verbal and quantitative
 - Multiple items per category
 - But don't be boring



[92]

Clear Presentation

- Interesting start
- Organized
- Balance order (if necessary)
- End with demographics
 - Easy
 - Boring
 - Stereotype Threat

[93]

Effect of Interviewing Method on Questionnaire Design

Department Store Project

Mail Questionnaire

- Please rank order the following department stores in order of your preference to shop at these stores. Begin by picking out the one store that you like most and assign it a number 1. Then find the second most preferred department store and assign it a number 2. Continue this procedure until you have ranked all the stores in order of preference. The least preferred store should be assigned a rank of 10. No two stores should receive the same rank number.

Store	Rank Order
1. Nordstrom	_____
2. Macy's	_____
.	
.	
10. Wal-Mart	_____

Effect of Interviewing Method on Questionnaire Design

Telephone Questionnaire

- I will read to you the names of some department stores. Please rate them in terms of your preference to shop at these stores. Use a ten-point scale, where 1 denotes not so preferred and 10 denotes greatly preferred. Numbers between 1 and 10 reflect intermediate degrees of preference. Again, please remember that the higher the number, the greater the degree of preference. Now, please tell me your preference to shop at(READ ONE STORE AT A TIME)

Store	Not So Preferred									Greatly Preferred
1. Nordstrom	1	2	3	4	5	6	7	8	9	10
2. Macy's	1	2	3	4	5	6	7	8	9	10
.										
.										
10. Wal-Mart	1	2	3	4	5	6	7	8	9	10

Effect of Interviewing Method on Questionnaire Design

Personal Questionnaire

- (HAND DEPARTMENT STORE CARDS TO THE RESPONDENT.) Here is a set of department store names, each written on a separate card. Please examine these cards carefully. (GIVE RESPONDENT TIME.) Now, please examine these cards again and pull out that card which has the name of the store you like the most, i.e., your most preferred store for shopping. (RECORD THE STORE NAME AND KEEP THIS CARD WITH YOU.) Now, please examine the remaining nine cards. Of these remaining nine stores, what is your most preferred store for shopping? (REPEAT THIS PROCEDURE SEQUENTIALLY UNTIL THE RESPONDENT HAS ONLY ONE CARD LEFT.)

Store Rank	Name of the Store
1. <u>1</u>	_____
2. <u>2</u>	_____
.	
10. <u>10</u>	_____

Effect of Interviewing Method on Questionnaire Design

Electronic Questionnaire

- This question for e-mail and Internet questionnaires will be very similar to that for the mail questionnaire.
- In all these methods, the questionnaire is self-administered by the respondent.

Individual Question Content — Is the Question Necessary?

- If there is no satisfactory use for the data resulting from a question, that question should be eliminated.

Individual Question Content — Are Several Questions Needed Instead of One?

- Sometimes, several questions are needed to obtain the required information in an unambiguous manner. Consider the question:

“Do you think Coca-Cola is a tasty and refreshing soft drink?”
(Incorrect)

- Such a question is called a **double-barreled question**, because two or more questions are combined into one. To obtain the required information, two distinct questions should be asked:

“Do you think Coca-Cola is a tasty soft drink?” and
“Do you think Coca-Cola is a refreshing soft drink?”
(Correct)

Overcoming Inability To Answer – Is the Respondent Informed?

- In situations where not all respondents are likely to be informed about the topic of interest, **filter questions** that measure familiarity and past experience should be asked before questions about the topics themselves.
- A “don't know” option appears to reduce uninformed responses without reducing the response rate.

Overcoming Inability To Answer – Can the Respondent Remember?

How many gallons of soft drinks did you
consume during the last four weeks? (Incorrect)

How often do you consume soft drinks in a
typical week? (Correct)

1. ☐ Less than once a week
2. ☐ 1 to 3 times per week
3. ☐ 4 to 6 times per week
4. ☐ 7 or more times per week

Overcoming Inability To Answer – Can the Respondent Articulate?

- Respondents may be unable to articulate certain types of responses, e.g., describe the atmosphere of a department store.
- Respondents should be given aids, such as pictures, maps, and descriptions to help them articulate their responses.

Overcoming Unwillingness To Answer – Effort Required of the Respondents

- Most respondents are unwilling to devote a lot of effort to provide information.

Overcoming Unwillingness To Answer

Please list all the departments from which you purchased merchandise on your most recent shopping trip to a department store. (Incorrect)

In the list that follows, please check all the departments from which you purchased merchandise on your most recent shopping trip to a department store.

- | | | | |
|-----|------------------------|-------|-----------|
| 1. | Women's dresses | _____ | |
| 2. | Men's apparel | _____ | _____ |
| 3. | Children's apparel | _____ | |
| 4. | Cosmetics | _____ | _____ |
| . | | | |
| . | | | |
| . | | | |
| 16. | Jewelry | _____ | |
| 17. | Other (please specify) | _____ | (Correct) |

Overcoming Unwillingness To Answer

Context

- Respondents are unwilling to respond to questions which they consider to be inappropriate for the given context.
- The researcher should manipulate the context so that the request for information seems appropriate.

Legitimate Purpose

- Explaining why the data are needed can make the request for the information seem legitimate and increase the respondents' willingness to answer.

Sensitive Information

- Respondents are unwilling to disclose, at least accurately, sensitive information because this may cause embarrassment or threaten the respondent's prestige or self-image.

Overcoming Unwillingness To Answer – Increasing the Willingness of Respondents

- Place sensitive topics at the end of the questionnaire.
- Preface the question with a statement that the behavior of interest is common.
- Ask the question using the third-person technique: phrase the question as if it referred to other people.
- Hide the question in a group of other questions which respondents are willing to answer. The entire list of questions can then be asked quickly.
- Provide response categories rather than asking for specific figures.
- Use randomized techniques.

Choosing Question Structure – Unstructured Questions

- **Unstructured questions** are open-ended questions that respondents answer in their own words.

What is your occupation?

Who is your favorite actor?

What do you think about people who shop at high-end department stores?

Choosing Question Structure – Structured Questions

- **Structured questions** specify the set of response alternatives and the response format. A structured question may be multiple-choice, dichotomous, or a scale.

Choosing Question Structure – Multiple-Choice Questions

- In **multiple-choice questions**, the researcher provides a choice of answers and respondents are asked to select one or more of the alternatives given.

Do you intend to buy a new car within the next six months?

- ☐ Definitely will not buy
- ☐ Probably will not buy
- ☐ Undecided
- ☐ Probably will buy
- ☐ Definitely will buy
- ☐ Other (please specify)

Choosing Question Structure – Dichotomous Questions

- A **dichotomous question** has only two response alternatives: yes or no, agree or disagree, and so on.
- Often, the two alternatives of interest are supplemented by a neutral alternative, such as “no opinion,” “don’t know,” “both,” or “none.”

Do you intend to buy a new car within the next six months?

- ☐ Yes
- ☐ No
- ☐ Don't know

Choosing Question Structure – Scales

Do you intend to buy a new car within the next six months?

Definitely will not buy	Probably will not buy	Undecided	Probably will buy	Definitely will buy
1	2	3	4	5

Choosing Question Wording – Define the Issue

- Define the issue in terms of who, what, when, where, why, and way (the six Ws). Who, what, when, and where are particularly important.

Which brand of shampoo do you use?

(Incorrect)

Which brand or brands of shampoo have you personally used at home during the last month?

In case of more than one brand, please

list all the brands that apply.

(Correct)

Choosing Question Wording

The W's	Defining the Question
Who	The Respondent It is not clear whether this question relates to the individual respondent or the respondent's total household.
What	The Brand of Shampoo It is unclear how the respondent is to answer this question if more than one brand is used.
When	Unclear The time frame is not specified in this question. The respondent could interpret it as meaning the shampoo used this morning, this week, or over the past year.
Where	Unclear At home, at the gym, on the road?

Choosing Question Wording – Use Ordinary Words

“Do you think the distribution of soft drinks is adequate?”
(Incorrect)

“Do you think soft drinks are readily available when you want to buy them?”
(Correct)

Choosing Question Wording – Use Unambiguous Words

In a typical month, how often do you shop in department stores?

- ☐ Never
- ☐ Occasionally
- ☐ Sometimes
- ☐ Often
- ☐ Regularly

(Incorrect)

In a typical month, how often do you shop in department stores?

- ☐ Less than once
- ☐ 1 or 2 times
- ☐ 3 or 4 times
- ☐ More than 4 times

(Correct)

Choosing Question Wording – Avoid Leading or Biasing Questions

- A **leading question** is one that clues the respondent to what the answer should be, as in the following:

Do you think that patriotic Americans should buy imported automobiles when that would put American labor out of work?

- ☐ Yes
- ☐ No
- ☐ Don't know

(Incorrect)

Do you think that Americans should buy imported automobiles?

- ☐ Yes
- ☐ No
- ☐ Don't know

(Correct)

Choosing Question Wording – Avoid Implicit Alternatives

- An alternative that is not explicitly expressed in the options is an **implicit alternative**.

1. Do you like to fly when traveling short distances?

(Incorrect)

2. Do you like to fly when traveling short distances, or would you rather drive?

(Correct)

Choosing Question Wording – Avoid Implicit Assumptions

- Questions should not be worded so that the answer is dependent upon **implicit assumptions** about what will happen as a consequence.

1. Are you in favor of a balanced budget?

(Incorrect)

2. Are you in favor of a balanced budget if it would result in an increase in the personal income tax?

(Correct)

Choosing Question Wording – Avoid Generalizations and Estimates

“What is the annual per capita expenditure on groceries in your household?”

(Incorrect)

“What is the monthly (or weekly) expenditure on groceries in your household?”

and

“How many members are there in your household?”

(Correct)

Choosing Question Wording Dual Statements: Positive and Negative

- Questions that are in the form of statements should be worded both positively and negatively.

Determining the Order of Questions

Opening Questions

- The opening questions should be interesting, simple, and non-threatening.

Type of Information

- As a general guideline, basic information should be obtained first, followed by classification, and, finally, identification information.

Difficult Questions

- Difficult questions or questions which are sensitive, embarrassing, complex, or dull, should be placed late in the sequence.

Determining the Order of Questions

Effect on Subsequent Questions

- General questions should precede the specific questions (funnel approach).

Q1: "What considerations are important to you in selecting a department store?"

Q2: "In selecting a department store, how important is convenience of location?"

(Correct)

Determining the Order of Questions

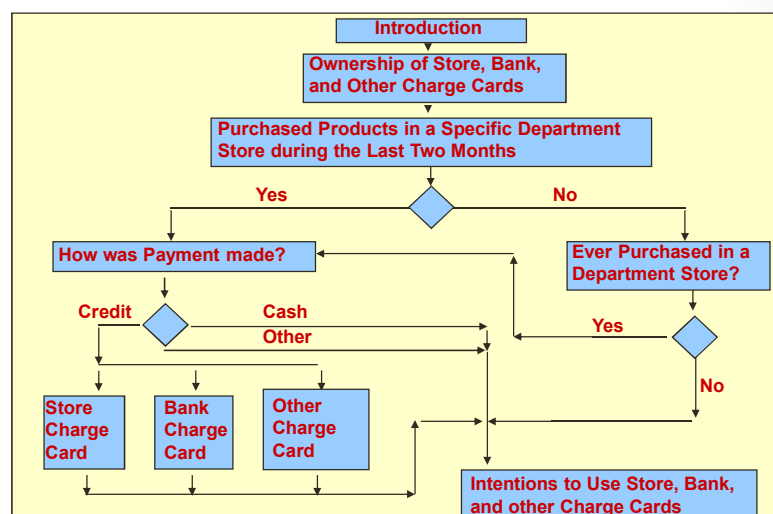
Logical Order

The following guidelines should be followed for

branching questions:

- The question being branched (the one to which the respondent is being directed) should be placed as close as possible to the question causing the branching.
- The branching questions should be ordered so that the respondents cannot anticipate what additional information will be required.

Flow Chart for Questionnaire Design



Form and Layout

- Divide a questionnaire into several parts.
- The questions in each part should be numbered, particularly when branching questions are used.
- The questionnaires should preferably be precoded.
- The questionnaires themselves should be numbered serially.

Example of a Precoded Questionnaire

The American Lawyer

A Confidential Survey of Our Subscribers



(Please ignore the numbers alongside the answers. They are only to help us in data processing.)

1. Considering all the times you pick it up, about how much time, in total, do you spend reading or looking through a typical issue of THE AMERICAN LAWYER?

Less than 30 minutes..... ☐-1	1 1/2 hours to 1 hour 59 minutes..... ☐-4
30 to 59 minutes..... ☐-2	2 hours to 2 hours 59 minutes..... ☐-5
1 hour to 1 hour 29 minutes..... ☐-3	3 hours or more..... ☐-6

Reproduction of the Questionnaire

- The questionnaire should be reproduced on good-quality paper and have a professional appearance.
- Questionnaires should take the form of a booklet rather than a number of sheets of paper clipped or stapled together.
- Each question should be reproduced on a single page (or double-page spread).
- Vertical response columns should be used for individual questions.
- Grids are useful when there are a number of related questions that use the same set of response categories.
- The tendency to crowd questions together to make the questionnaire look shorter should be avoided.
- Directions or instructions for individual questions should be placed as close to the questions as possible.

Pretesting

Pretesting refers to the testing of the questionnaire on a small sample of respondents to identify and eliminate potential problems.

- A questionnaire should not be used in the field survey without adequate pretesting.
- All aspects of the questionnaire should be tested, including question content, wording, sequence, form and layout, question difficulty, and instructions.
- The respondents for the pretest and for the actual survey should be drawn from the same population.
- Pretests are best done by personal interviews, even if the actual survey is to be conducted by mail, telephone, or electronic means, because interviewers can observe respondents' reactions and attitudes.

Pretesting

- After the necessary changes have been made, another pretest could be conducted by mail, telephone, or electronic means if those methods are to be used in the actual survey.
- A variety of interviewers should be used for pretests.
- The pretest sample size varies from 15 to 30 respondents for each wave.
- Protocol analysis and debriefing are two commonly used procedures in pretesting.
- Finally, the responses obtained from the pretest should be coded and analyzed.

Observational Forms

Department Store Project

- **Who:** Purchasers, browsers, males, females, parents with children, or children alone.
- **What:** Products/brands considered, products/brands purchased, size, price of package inspected, or influence of children or other family members.
- **When:** Day, hour, date of observation.
- **Where:** Inside the store, checkout counter, or type of department within the store.
- **Why:** Influence of price, brand name, package size, promotion, or family members on the purchase.
- **Way:** Personal observer disguised as sales clerk, undisguised personal observer, hidden camera, or obtrusive mechanical device.

Questionnaire Design Checklist

- Step 1. Specify The Information Needed
- Step 2. Type of Interviewing Method
- Step 3. Individual Question Content
- Step 4. Overcome Inability and Unwillingness to Answer
- Step 5. Choose Question Structure
- Step 6. Choose Question Wording
- Step 7. Determine the Order of Questions
- Step 8. Form and Layout
- Step 9. Reproduce the Questionnaire
- Step 10. Pretest



Questionnaire Design Checklist (cont.)

Step 1. Specify the Information Needed

- 1. Ensure that the information obtained fully addresses all the components of the problem. Review components of the problem and the approach, particularly the research questions, hypotheses, and specification of information needed.
- 2. Prepare a set of dummy tables.
- 3. Have a clear idea of the target population.

Step 2. Type of Interviewing Method

- 1. Review the type of interviewing method determined based on considerations discussed in Chapter 6.



Questionnaire Design Checklist (cont.)

Step 3. Individual Question Content

1. Is the question necessary?
2. Are several questions needed instead of one to obtain the required information in an unambiguous manner?
3. Do not use double-barreled questions.



Questionnaire Design Checklist (cont.)

Step 4. Overcoming Inability and Unwillingness

1. Is the respondent informed?
2. If respondents are not likely to be informed, filter questions that measure familiarity, product use, and past experience should be asked before questions about the topics themselves.
3. Can the respondent remember?
4. Avoid errors of omission, telescoping, and creation.
5. Questions which do not provide the respondent with cues can underestimate the actual occurrence of an event.
6. Can the respondent articulate?

Questionnaire Design Checklist (cont.)

Step 4. Overcoming Inability and Unwillingness

7. Minimize the effort required of the respondents.
8. Is the context in which the questions are asked appropriate?
9. Make the request for information seem legitimate.
10. If the information is sensitive:
 - a. Place sensitive topics at the end of the questionnaire.
 - b. Preface the question with a statement that the behavior of interest is common.
 - c. Ask the question using the third-person technique.
 - d. Hide the question in a group of other questions which respondents are willing to answer.
 - e. Provide response categories rather than asking for specific figures.
 - f. Use randomized techniques, if appropriate.

Questionnaire Design Checklist (cont.)

Step 5. Choosing Question Structure

1. Open-ended questions are useful in exploratory research and as opening questions.
2. Use structured questions whenever possible.
3. In multiple-choice questions, the response alternatives should include the set of all possible choices and should be mutually exclusive.
4. In a dichotomous question, if a substantial proportion of the respondents can be expected to be neutral, include a neutral alternative.
5. Consider the use of the split ballot technique to reduce order bias in dichotomous and multiple-choice questions.
6. If the response alternatives are numerous, consider using more than one question to reduce the information processing demands on the respondents.



Questionnaire Design Checklist (cont.)

Step 6. Choosing Question Wording

1. Define the issue in terms of who, what, when, where, why, and way (the six Ws).
2. Use ordinary words. Words should match the vocabulary level of the respondents.
3. Avoid ambiguous words: usually, normally, frequently, often, regularly, occasionally, sometimes, etc.
4. Avoid leading questions that clue the respondent to what the answer should be.
5. Avoid implicit alternatives that are not explicitly expressed in the options.
6. Avoid implicit assumptions.
7. Respondent should not have to make generalizations or compute estimates.
8. Use positive and negative statements.

Questionnaire Design Checklist (cont.)

Step 7. Determine the Order of Questions

1. The opening questions should be interesting, simple, and non-threatening.
2. Qualifying questions should serve as the opening questions.
3. Basic information should be obtained first, followed by classification, and, finally, identification information.
4. Difficult, or sensitive questions should be placed late in the sequence.
5. General questions should precede the specific questions.
6. Questions should be asked in a logical order.
7. Branching questions should be designed carefully to cover all possible contingencies.
8. The question being branched should be placed as close as possible to the question causing the branching, and (2) the branching questions should be ordered so that the respondents cannot anticipate what additional information will be required.

Questionnaire Design Checklist (cont.)

Step 8. Form and Layout

1. Divide a questionnaire into several parts.
2. Questions in each part should be numbered.
3. The questionnaire should be pre-coded.
4. The questionnaires themselves should be numbered serially.



Questionnaire Design Checklist (cont.)

Step 9. Reproduction of the Questionnaire

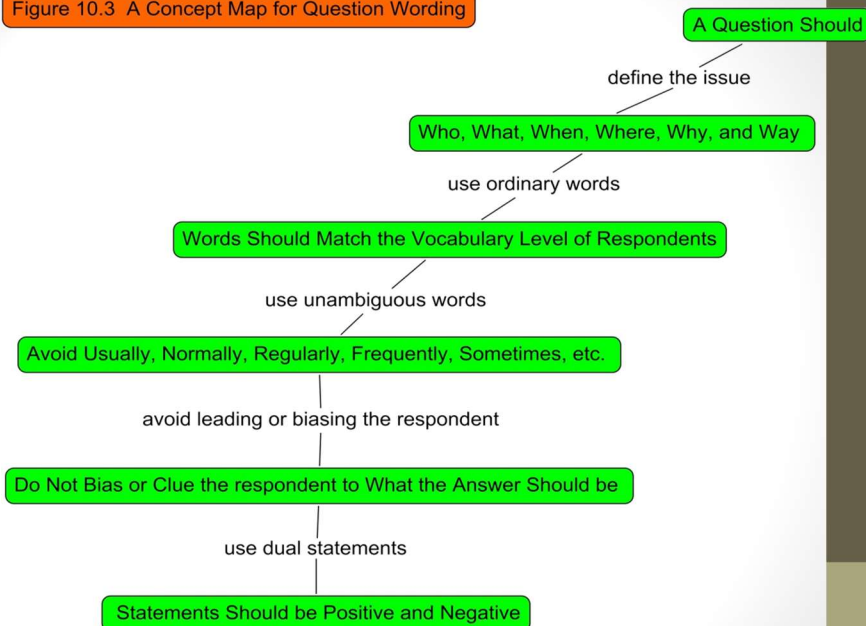
1. The questionnaire should have a professional appearance.
2. Booklet format should be used for long questionnaires.
3. Each question should be reproduced on a single page (or double-page spread).
4. Vertical response columns should be used.
5. Grids are useful when there are a number of related questions which use the same set of response categories.
6. The tendency to crowd questions to make the questionnaire look shorter should be avoided.
7. Directions or instructions for individual questions should be placed as close to the questions as possible.

Questionnaire Design Checklist (cont.)

Step 10. Pretesting

1. Pretesting should be done always.
2. All aspects of the questionnaire should be tested, including question content, wording, sequence, form and layout, question difficulty, and instructions.
3. The respondents in the pretest should be similar to those in the actual survey.
4. Begin the pretest by using personal interviews.
5. Pretest should also be conducted by mail or telephone if those methods are to be used in the actual survey.
6. A variety of interviewers should be used for pretests.
7. The pretest sample size is small, 15 to 30, respondents for the initial testing.
8. Use protocol analysis and debriefing to identify problems.
9. After each significant revision of the questionnaire, another pretest should be conducted, using a different sample of respondents.
10. The responses obtained from the pretest should be coded and analyzed.

Figure 10.3 A Concept Map for Question Wording

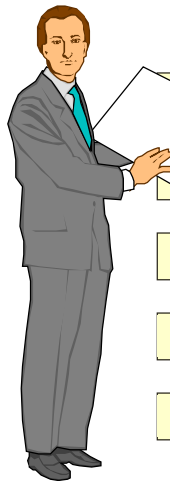


SAMPLING AND DATA COLLECTION

Sample Vs. Census

Type of Study	Conditions Favoring the Use of	
	Sample	Census
1. Budget	Small	Large
2. Time available	Short	Long
3. Population size	Large	Small
4. Variance in the characteristic	Small	Large
5. Cost of sampling errors	Low	High
6. Cost of nonsampling errors	High	Low
7. Nature of measurement	Destructive	Nondestructive
8. Attention to individual cases	Yes	No

The Sampling Design Process



Define the Population

Determine the Sampling Frame

Select Sampling Technique(s)

Determine the Sample Size

Execute the Sampling Process

Define the Target Population

The target population is the collection of elements or objects that possess the information sought by the researcher and about which inferences are to be made. The target population should be defined in terms of elements, sampling units, extent, and time.

- An **element** is the object about which or from which the information is desired, e.g., the respondent.
- A **sampling unit** is an element, or a unit containing the element, that is available for selection at some stage of the sampling process.
- **Extent** refers to the geographical boundaries.
- **Time** is the time period under consideration.

Define the Target Population

Important qualitative factors in determining the sample size are:

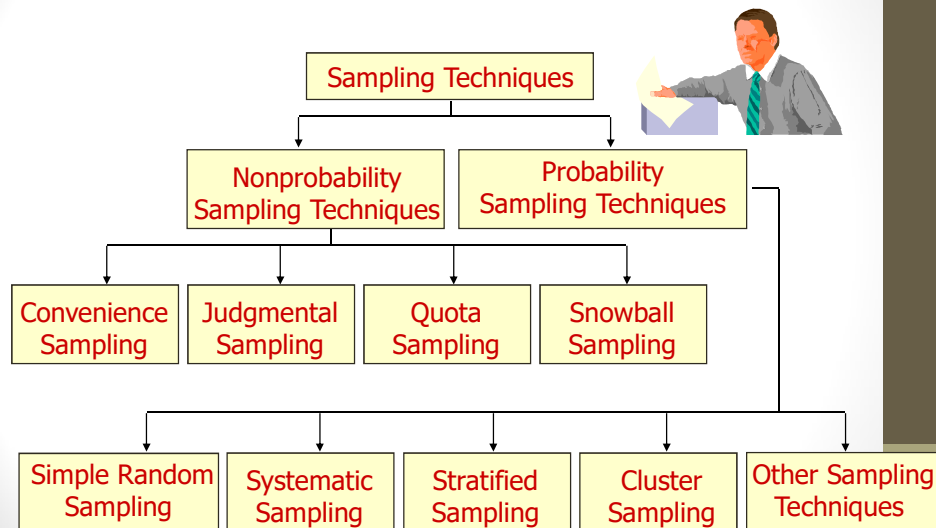
- the importance of the decision
- the nature of the research
- the number of variables
- the nature of the analysis
- sample sizes used in similar studies
- incidence rates
- completion rates
- resource constraints

Sample Sizes Used in Marketing Research Studies

Type of Study	Minimum Size	Typical Range
Problem identification research (e.g. market potential)	500	1,000-2,500
Problem-solving research (e.g. pricing)	200	300-500
Product tests	200	300-500
Test marketing studies	200	300-500
TV, radio, or print advertising (per commercial or ad tested)	150	200-300
Test-market audits	10 stores	10-20 stores
Focus groups	2 groups	6-15 groups



Classification of Sampling Techniques



Convenience Sampling

Convenience sampling attempts to obtain a sample of convenient elements. Often, respondents are selected because they happen to be in the right place at the right time.

- Use of students, and members of social organizations
- Mall intercept interviews without qualifying the respondents
- Department stores using charge account lists
- “People on the street” interviews

A Graphical Illustration of Convenience Sampling

A	B	C	D	E
1	6	11	16	21
2	7	12	17	22
3	8	13	18	23
4	9	14	19	24
5	10	15	20	25

Group D happens to assemble at a convenient time and place. So all the elements in this Group are selected. The resulting sample consists of elements 16, 17, 18, 19 and 20. Note, no elements are selected from group A, B, C and E.

Judgmental Sampling

Judgmental sampling is a form of convenience sampling in which the population elements are selected based on the judgment of the researcher.

- Test markets
- Purchase engineers selected in industrial marketing research
- Bellwether precincts selected in voting behavior research
- Expert witnesses used in court

Graphical Illustration of Judgmental Sampling

A	B	C	D	E
1	6	11	16	21
2	7	12	17	22
3	8	13	18	23
4	9	14	19	24
5	10	15	20	25

The researcher considers groups B, C and E to be typical and convenient. Within each of these groups one or two elements are selected based on typicality and convenience. The resulting sample consists of elements 8, 10, 11, 13, and 24. Note, no elements are selected from groups A and D.

Quota Sampling

Quota sampling may be viewed as two-stage restricted judgmental sampling.

- The first stage consists of developing control categories, or quotas, of population elements.
- In the second stage, sample elements are selected based on convenience or judgment.

Control Population Variable composition

Sex	Percentage
Male	48
Female	52
	<u>100</u>

Sample composition

Percentage	Number
48	480
52	520
<u>100</u>	<u>1000</u>

A Graphical Illustration of Quota Sampling

A	B	C	D	E
1	6	11	16	21
2	7	12	17	22
3	8	13	18	23
4	9	14	19	24
5	10	15	20	25

A quota of one element from each group, A to E, is imposed. Within each group, one element is selected based on judgment or convenience. The resulting sample consists of elements 3, 6, 13, 20 and 22. Note, one element is selected from each column or group.

Snowball Sampling

In **snowball sampling**, an initial group of respondents is selected, usually at random.

- After being interviewed, these respondents are asked to identify others who belong to the target population of interest.
- Subsequent respondents are selected based on the referrals.

A Graphical Illustration of Snowball Sampling

Random
Selection

Referrals

A	B	C	D	E
1	6	11	16	21
2	7	12	17	22
3	8	13	18	23
4	9	14	19	24
5	10	15	20	25

Elements 2 and 9 are selected randomly from groups A and B. Element 2 refers elements 12 and 13. Element 9 refers element 18. The resulting sample consists of elements 2, 9, 12, 13, and 18. Note, there are no elements from group E.

Simple Random Sampling

- Each element in the population has a known and equal probability of selection.
- Each possible sample of a given size (n) has a known and equal probability of being the sample actually selected.
- This implies that every element is selected independently of every other element.

Graphical Illustration of Simple Random Sampling

A	B	C	D	E
1	6	11	16	21
2	7	12	17	22
3	8	13	18	23
4	9	14	19	24
5	10	15	20	25

Select five random numbers from 1 to 25. The resulting sample consists of population elements 3, 7, 9, 16, and 24. Note, there is no element from Group C.

Systematic Sampling

- The sample is chosen by selecting a random starting point and then picking every i th element in succession from the sampling frame.
- The sampling interval, i , is determined by dividing the population size N by the sample size n and rounding to the nearest integer.
- When the ordering of the elements is related to the characteristic of interest, systematic sampling increases the representativeness of the sample.

Systematic Sampling

- If the ordering of the elements produces a cyclical pattern, systematic sampling may decrease the representativeness of the sample.

For example, there are 100,000 elements in the population and a sample of 1,000 is desired. In this case the sampling interval, i , is 100. A random number between 1 and 100 is selected. If, for example, this number is 23, the sample consists of elements 23, 123, 223, 323, 423, 523, and so on.

Graphical Illustration of Systematic Sampling

A	B	C	D	E
1	6	11	16	21
2	7	12	17	22
3	8	13	18	23
4	9	14	19	24
5	10	15	20	25

Select a random number between 1 and 5, say 2.
The resulting sample consists of population 2, $(2+5=) 7$, $(2+5 \times 2=) 12$, $(2+5 \times 3=) 17$, and $(2+5 \times 4=) 22$.
Note, all the elements are selected from a single row.

Stratified Sampling

- A two-step process in which the population is partitioned into subpopulations, or strata.
- The strata should be mutually exclusive and collectively exhaustive in that every population element should be assigned to one and only one stratum and no population elements should be omitted.
- Next, elements are selected from each stratum by a random procedure, usually SRS.
- A major objective of stratified sampling is to increase precision without increasing cost.

Stratified Sampling

- The elements within a stratum should be as homogeneous as possible, but the elements in different strata should be as heterogeneous as possible.
- The stratification variables should also be closely related to the characteristic of interest.
- Finally, the variables should decrease the cost of the stratification process by being easy to measure and apply.

Stratified Sampling

- In **proportionate stratified sampling**, the size of the sample drawn from each stratum is proportionate to the relative size of that stratum in the total population.
- In **disproportionate stratified sampling**, the size of the sample from each stratum is proportionate to the relative size of that stratum and to the standard deviation of the distribution of the characteristic of interest among all the elements in that stratum.

Graphical Illustration of Stratified Sampling

A	B	C	D	E
1	6	11	16	21
2	7	12	17	22
3	8	13	18	23
4	9	14	19	24
5	10	15	20	25

Randomly select a number from 1 to 5 for each stratum, A to E. The resulting sample consists of population elements 4, 7, 13, 19 and 21. Note, one element is selected from each column.

Cluster Sampling

- The target population is first divided into mutually exclusive and collectively exhaustive subpopulations, or clusters.
- Then a random sample of clusters is selected, based on a probability sampling technique such as SRS.
- For each selected cluster, either all the elements are included in the sample (one-stage) or a sample of elements is drawn probabilistically (two-stage).

Cluster Sampling

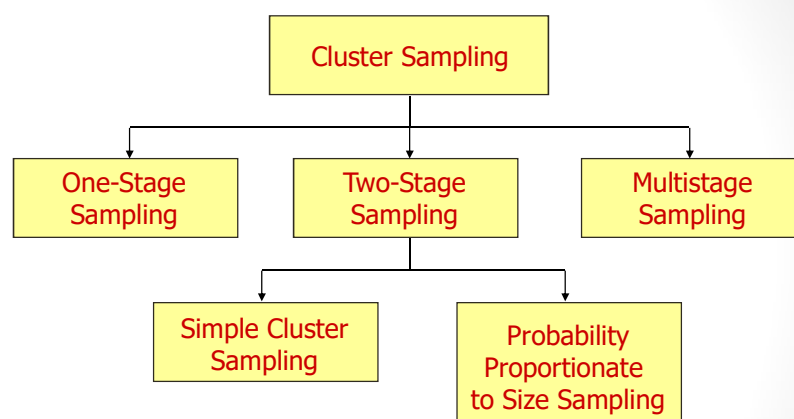
- Elements within a cluster should be as heterogeneous as possible, but clusters themselves should be as homogeneous as possible. Ideally, each cluster should be a small-scale representation of the population.
- In **probability proportionate to size sampling**, the clusters are sampled with probability proportional to size. In the second stage, the probability of selecting a sampling unit in a selected cluster varies inversely with the size of the cluster.

A Graphical Illustration of Cluster Sampling (2-Stage)

A	B	C	D	E
1	6	11	16	21
2	7	12	17	22
3	8	13	18	23
4	9	14	19	24
5	10	15	20	25

Randomly select 3 clusters, B, D and E.
 Within each cluster, randomly select one or two elements.
 The resulting sample consists of population elements 7, 18, 20, 21, and 23.
 Note, no elements are selected from clusters A and C.

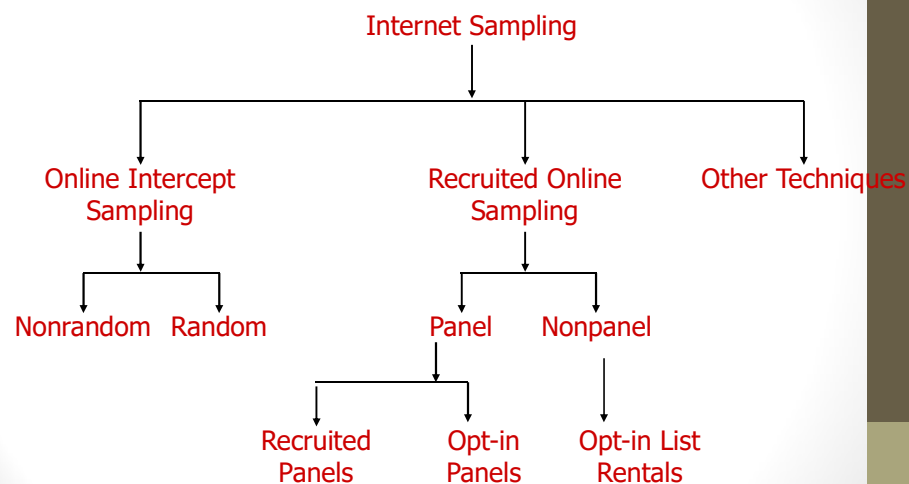
Types of Cluster Sampling



Strengths and Weaknesses of Basic Sampling Techniques

	Strengths	Weaknesses
Nonprobability Sampling		
Convenience sampling	Least expensive, least time-consuming, most convenient	Selection bias, sample not representative, not recommended for descriptive or causal research
Judgmental sampling	Low cost, convenient, not time-consuming	Does not allow generalization, subjective
Quota sampling	Sample can be controlled for certain characteristics	Selection bias, no assurance of representativeness
Snowball sampling	Can estimate rare characteristics	Time-consuming
Probability sampling		
Simple random sampling (SRS)	Easily understood, results projectable	Difficult to construct sampling frame, expensive, lower precision, no assurance of representativeness
Systematic sampling	Can increase representativeness, easier to implement than SRS, sampling frame not necessary	Can decrease representativeness
Stratified sampling	Include all important subpopulations, precision	Difficult to select relevant stratification variables, not feasible to stratify on many variables, expensive
Cluster sampling	Easy to implement, cost effective	Imprecise, difficult to compute and interpret results

A Classification of Internet Sampling



Procedures for Drawing Probability Samples

Simple Random Sampling

Simple Random Sampling

1. Select a suitable sampling frame.
2. Each element is assigned a number from 1 to N (pop. size).
3. Generate n (sample size) different random numbers between 1 and N.
4. The numbers generated denote the elements that should be included in the sample.

Procedures for Drawing Probability Samples (cont.)

Systematic Sampling

1. Select a suitable sampling frame.
2. Each element is assigned a number from 1 to N (pop. size).
3. Determine the sampling interval $i: i = N/n$. If i is a fraction, round to the nearest integer.
4. Select a random number, r , between 1 and i , as explained in simple random sampling.
5. The elements with the following numbers will comprise the systematic random sample: $r, r+i, r+2i, r+3i, r+4i, \dots, r+(n-1)i$.

Procedures for Drawing Probability Samples (cont.)

Stratified Sampling

1. Select a suitable frame.
2. Select the stratification variable(s) and the number of strata, H.
3. Divide the entire population into H strata. Based on the classification variable, each element of the population is assigned to one of the H strata.
4. In each stratum, number the elements from 1 to N_h (the pop. size of stratum h).
5. Determine the sample size of each stratum, n_h , based on proportionate or disproportionate stratified sampling, where

$$\sum_{h=1}^H n_h = n$$

6. In each stratum, select a simple random sample of size n_h

Procedures for Drawing Probability Samples (cont.)

Cluster Sampling

1. Assign a number from 1 to N to each element in the population.
2. Divide the population into C clusters of which c will be included in the sample.
3. Calculate the sampling interval i , $i=N/c$ (round to nearest integer).
4. Select a random number r between 1 and i , as explained in simple random sampling.
5. Identify elements with the following numbers:
 $r, r+i, r+2i, \dots, r+(c-1)i$.
6. Select the clusters that contain the identified elements.
7. Select sampling units within each selected cluster based on SRS or systematic sampling.
8. Remove clusters exceeding sampling interval i . Calculate new population size N^* , number of clusters to be selected $c^*=c-1$, and new sampling interval i^* .

Procedures for Drawing Probability Samples (cont.)

Cluster Sampling

Repeat the process until each of the remaining clusters has a population less than the sampling interval. If b clusters have been selected with certainty, select the remaining $c-b$ clusters according to steps 1 through 7. The fraction of units to be sampled with certainty is the overall sampling fraction $= n/N$. Thus, for clusters selected with certainty, we would select $n_s = (n/N)(N_1 + N_2 + \dots + N_b)$ units. The units selected from clusters selected under two-stage sampling will therefore be $n^* = n - n_s$.

Choosing Nonprobability Vs. Probability Sampling

Factors	Conditions Favoring the Use of Nonprobability sampling	Probability sampling
Nature of research	Exploratory	Conclusive
Relative magnitude of sampling and nonsampling errors	Nonsampling errors are larger	Sampling errors are larger
Variability in the population	Homogeneous (low)	Heterogeneous (high)
Statistical considerations	Unfavorable	Favorable
Operational considerations	Favorable	Unfavorable

Tennis' Systematic Sampling Returns a Smash

Tennis magazine conducted a mail survey of its subscribers to gain a better understanding of its market. Systematic sampling was employed to select a sample of 1,472 subscribers from the publication's domestic circulation list. If we assume that the subscriber list had 1,472,000 names, the sampling interval would be 1,000 ($1,472,000/1,472$). A number from 1 to 1,000 was drawn at random. Beginning with that number, every 1,000th subscriber was selected.

A brand-new dollar bill was included with the questionnaire as an incentive to respondents. An alert postcard was mailed one week before the survey. A second, follow-up, questionnaire was sent to the whole sample ten days after the initial questionnaire. There were 76 post office returns, so the net effective mailing was 1,396. Six weeks after the first mailing, 778 completed questionnaires were returned, yielding a response rate of 56%.

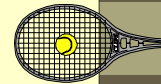
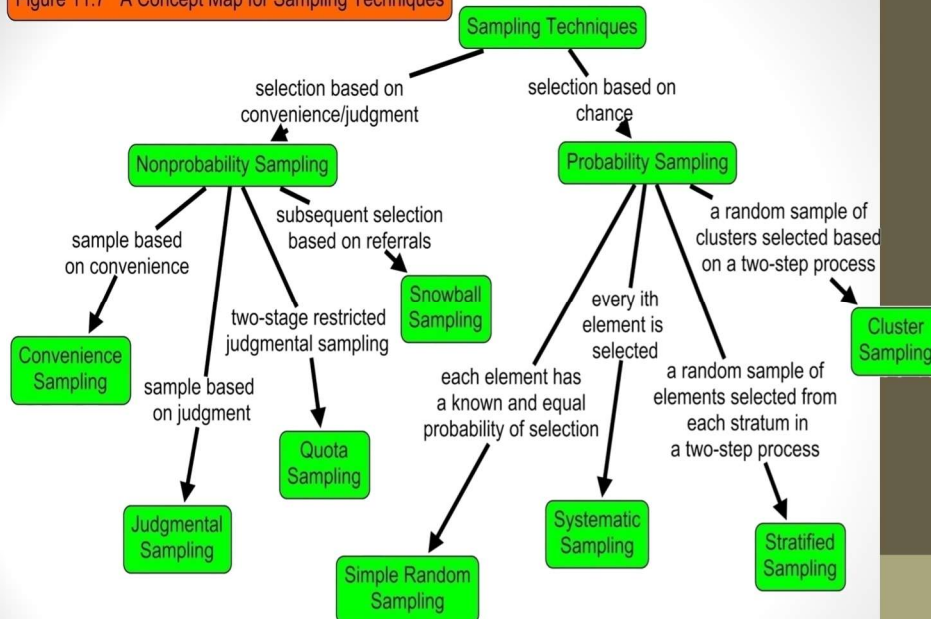


Figure 11.7 A Concept Map for Sampling Techniques



SAMPLE SIZE DETERMINATION

Definitions and Symbols

- **Parameter:** A **parameter** is a summary description of a fixed characteristic or measure of the target population. A parameter denotes the true value which would be obtained if a census rather than a sample was undertaken.
- **Statistic:** A **statistic** is a summary description of a characteristic or measure of the sample. The sample statistic is used as an estimate of the population parameter.
- **Finite Population Correction:** The **finite population correction** (fpc) is a correction for overestimation of the variance of a population parameter, e.g., a mean or proportion, when the sample size is 10% or more of the population size.

Definitions and Symbols

- **Precision level:** When estimating a population parameter by using a sample statistic, the **precision level** is the desired size of the estimating interval. This is the maximum permissible difference between the sample statistic and the population parameter.
- **Confidence interval:** The **confidence interval** is the range into which the true population parameter will fall, assuming a given level of confidence.
- **Confidence level:** The **confidence level** is the probability that a confidence interval will include the population parameter.

Symbols for Population and Sample Variables

Variable	Population	Sample
Mean	μ	$\bar{X}$
Proportion	Π	p
Variance	σ^2	s^2
Standard deviation	σ	s
Size	N	n
Standard error of the mean	$\sigma_{\bar{x}}$	$S_{\bar{x}}$
Standard error of the proportion	σ_p	S_p
Standardized variate (z)	$(X-\mu)/\sigma$	$(X-\bar{X})/S$
Coefficient of variation (CV)	σ/μ	$S/\bar{X}$

The Confidence Interval Approach

Calculation of the confidence interval involves determining a distance below ($\bar{X}_L$) and above ($\bar{X}_U$) the population mean (μ), which contains a specified area of the normal curve.

The z values corresponding to $\bar{X}_L$ and $\bar{X}_U$ may be calculated as

$$z_L = \frac{\bar{X}_L - \mu}{\sigma_{\bar{X}}}$$

$$z_U = \frac{\bar{X}_U - \mu}{\sigma_{\bar{X}}}$$

where $Z_L = -z$ and $Z_U = +z$. Therefore, the lower value of $\bar{X}$ is

$$\bar{X}_L = \mu - z\sigma_{\bar{X}}$$

and the upper value of $\bar{X}$ is

$$\bar{X}_U = \mu + z\sigma_{\bar{X}}$$

The Confidence Interval Approach

Note that μ is estimated by $\bar{X}$. The confidence interval is given by

$$\bar{X} \pm z\sigma_{\bar{X}}$$

We can now set a 95% confidence interval around the sample mean of \$182. As a first step, we compute the standard error of the mean:

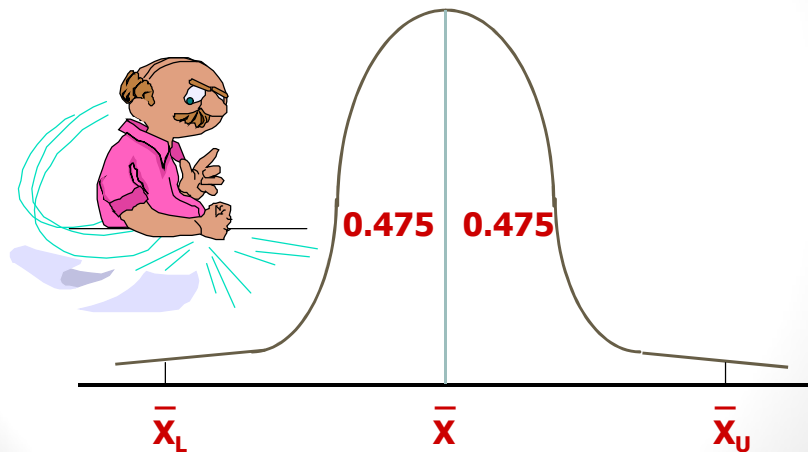
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = 55/\sqrt{300} = 3.18$$

From Table 2 in the Appendix of Statistical Tables, it can be seen that the central 95% of the normal distribution lies within ± 1.96 z values. The 95% confidence interval is given by

$$\begin{aligned}\bar{X} &\pm 1.96\sigma_{\bar{X}} \\ &= 182.00 \pm 1.96(3.18) \\ &= 182.00 \pm 6.23\end{aligned}$$

Thus the 95% confidence interval ranges from \$175.77 to \$188.23. The probability of finding the true population mean to be within \$175.77 and \$188.23 is 95%.

95% Confidence Interval



Sample Size Determination for Means and Proportions

Steps	Means	Proportions
1. Specify the level of precision.	$D = \pm \$5.00$	$D = p - \pi = \pm 0.05$
2. Specify the confidence level (CL).	CL = 95%	CL = 95%
3. Determine the z value associated with CL.	z value is 1.96	z value is 1.96
4. Determine the standard deviation of the population.	Estimate σ : $\sigma = 55$	Estimate π : $\pi = 0.64$
5. Determine the sample size using the formula for the standard error.	$n = \sigma^2 z^2 / D^2 = 465$	$n = \pi(1-\pi) z^2 / D^2 = 355$
6. If the sample size represents 10% of the population, apply the finite population correction.	$n_c = nN / (N+n-1)$	$n_c = nN / (N+n-1)$
7. If necessary, reestimate the confidence interval by employing s to estimate σ .	$= \bar{X} \pm z s_{\bar{X}}$	$= p \pm z s_p$
8. If precision is specified in relative rather than absolute terms, determine the sample size by substituting for D.	$D = R\mu$ $n = CV^2 z^2 / R^2$	$D = R\pi$ $n = z^2 (1-\pi) / (R^2 \pi)$

Sample Size for Estimating Multiple Parameters

	Variable Mean Household Monthly Expense On		
	Department store shopping	Clothes	Gifts
Confidence level	95%	95%	95%
z value	1.96	1.96	1.96
Precision level (D)	\$5	\$5	\$4
Standard deviation of the population (σ)	\$55	\$40	\$30
Required sample size (n)	465	246	217



Adjusting the Statistically Determined Sample Size

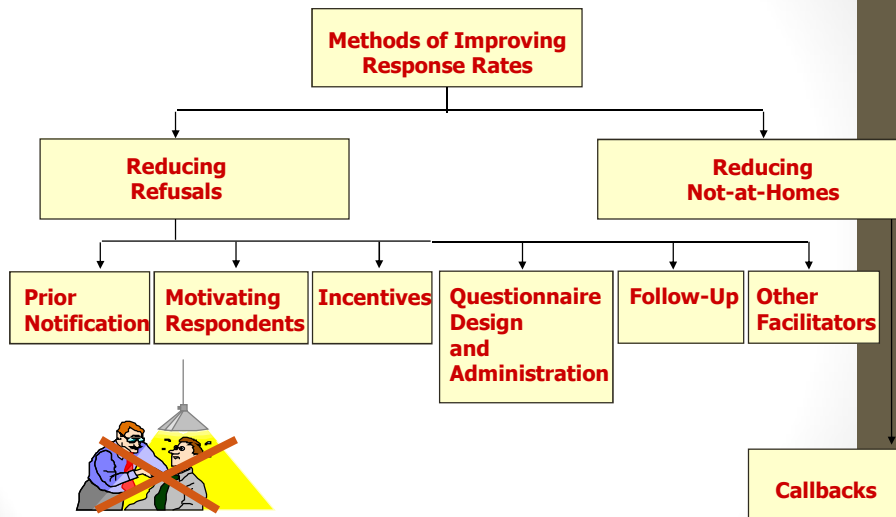
Incidence rate refers to the rate of occurrence or the percentage, of persons eligible to participate in the study.

In general, if there are c qualifying factors with an incidence of $Q_1, Q_2, Q_3, \dots, Q_c$, each expressed as a proportion:

$$\text{Incidence rate} = Q_1 \times Q_2 \times Q_3 \times \dots \times Q_c$$

$$\text{Initial sample size} = \frac{\text{Final sample size}}{\text{Incidence rate} \times \text{Completion rate}}$$

Improving Response Rates



Arbitron Responds to Low Response Rates

Arbitron, a major marketing research supplier, was trying to improve response rates in order to get more meaningful results from its surveys. Arbitron created a special cross-functional team of employees to work on the response rate problem. Their method was named the "breakthrough method," and the whole Arbitron system concerning the response rates was put in question and changed. The team suggested six major strategies for improving response rates:

1. Maximize the effectiveness of placement/follow-up calls.
2. Make materials more appealing and easy to complete.
3. Increase Arbitron name awareness.
4. Improve survey participant rewards.
5. Optimize the arrival of respondent materials.
6. Increase usability of returned diaries.



Eighty initiatives were launched to implement these six strategies. As a result, response rates improved significantly. However, in spite of those encouraging results, people at Arbitron remain very cautious. They know that they are not done yet and that it is an everyday fight to keep those response rates high.

Adjusting for Nonresponse

- **Subsampling of Nonrespondents** – the researcher contacts a subsample of the nonrespondents, usually by means of telephone or personal interviews.
- In **replacement**, the nonrespondents in the current survey are replaced with nonrespondents from an earlier, similar survey. The researcher attempts to contact these nonrespondents from the earlier survey and administer the current survey questionnaire to them, possibly by offering a suitable incentive.

Adjusting for Nonresponse

- In **substitution**, the researcher substitutes for nonrespondents other elements from the sampling frame that are expected to respond. The sampling frame is divided into subgroups that are internally homogeneous in terms of respondent characteristics but heterogeneous in terms of response rates. These subgroups are then used to identify substitutes who are similar to particular nonrespondents but dissimilar to respondents already in the sample.

Adjusting for Nonresponse

- **Subjective Estimates** – When it is no longer feasible to increase the response rate by subsampling, replacement, or substitution, it may be possible to arrive at subjective estimates of the nature and effect of nonresponse bias. This involves evaluating the likely effects of nonresponse based on experience and available information.
- **Trend analysis** is an attempt to discern a trend between early and late respondents. This trend is projected to nonrespondents to estimate where they stand on the characteristic of interest.

Use of Trend Analysis in Adjusting for Nonresponse

	Percentage Response	Average Dollar Expenditure	Percentage of Previous Wave's Response
First Mailing	12	412	—
Second Mailing	18	325	79
Third Mailing	13	277	85
Nonresponse	(57)	(230)	91
Total	100	275	



Adjusting for Nonresponse

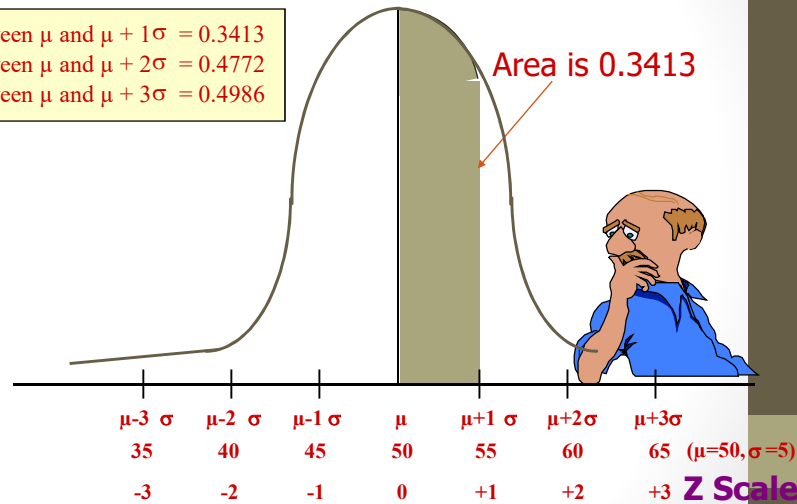
- **Weighting** attempts to account for nonresponse by assigning differential weights to the data depending on the response rates. For example, in a survey the response rates were 85, 70, and 40%, respectively, for the high-, medium-, and low-income groups. In analyzing the data, these subgroups are assigned weights inversely proportional to their response rates. That is, the weights assigned would be $(100/85)$, $(100/70)$, and $(100/40)$, respectively, for the high-, medium-, and low-income groups.

Adjusting for Nonresponse

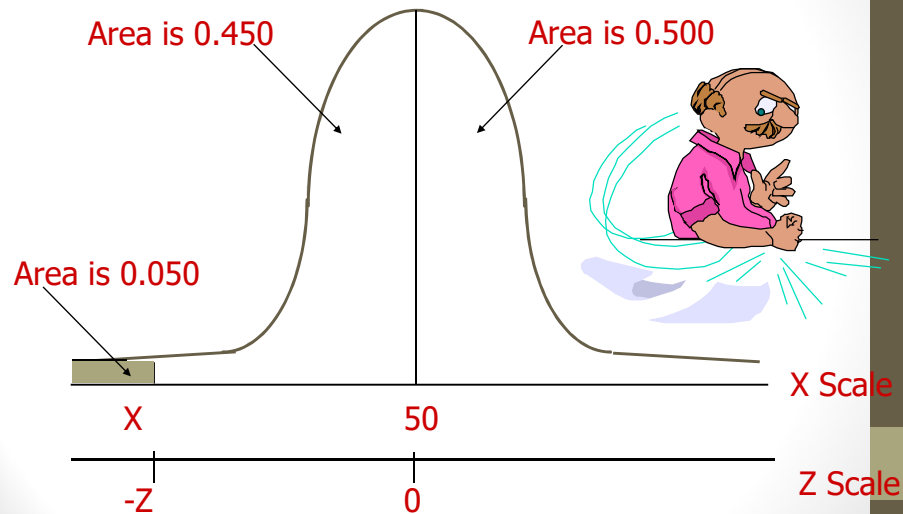
- **Imputation** involves imputing, or assigning, the characteristic of interest to the nonrespondents based on the similarity of the variables available for both nonrespondents and respondents. For example, a respondent who does not report brand usage may be imputed the usage of a respondent with similar demographic characteristics.

Finding Probabilities Corresponding to Known Values

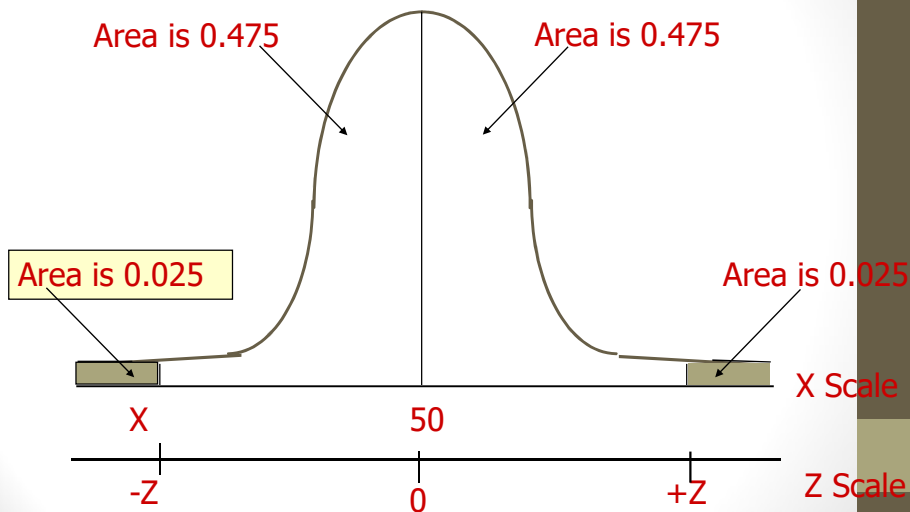
Area between μ and $\mu + 1\sigma = 0.3413$
 Area between μ and $\mu + 2\sigma = 0.4772$
 Area between μ and $\mu + 3\sigma = 0.4986$



Finding Probabilities Corresponding to Known Values



Finding Values Corresponding to Known Probabilities: Confidence Interval



Opinion Place Bases Its Opinions on 1000 Respondents

Marketing research firms are now turning to the Web to conduct online research. Recently, four leading market research companies (ASI Market Research, Custom Research, Inc., M/A/R/C Research, and Roper Search Worldwide) partnered with Digital Marketing Services (DMS), Dallas, to conduct custom research on AOL.

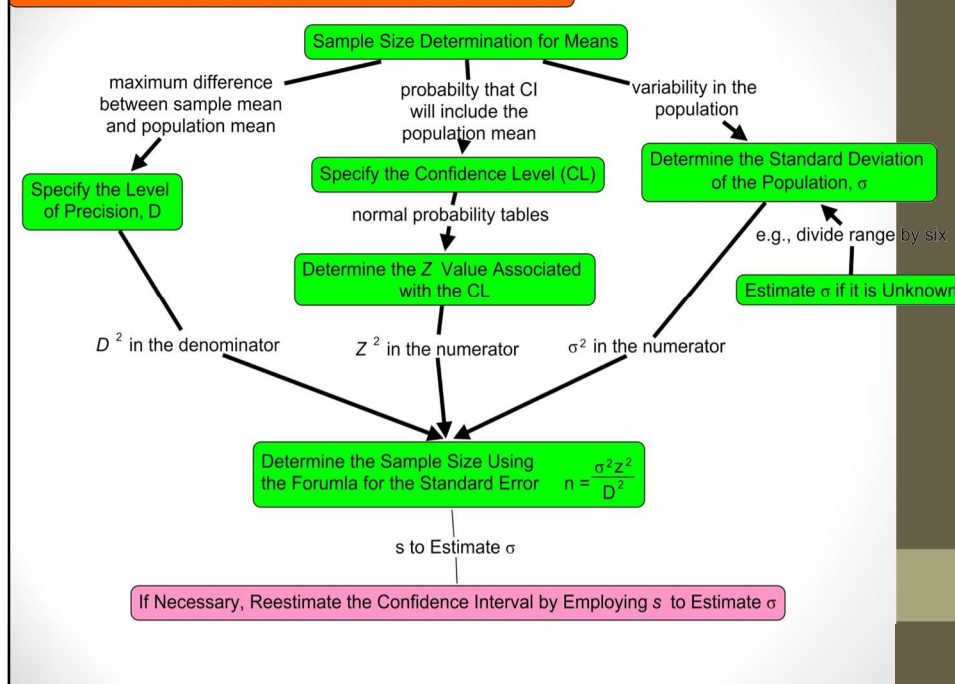
DMS and AOL will conduct online surveys on AOL's *Opinion Place*, with an average base of 1,000 respondents by survey. This sample size was determined based on statistical considerations as well as sample sizes used in similar research conducted by traditional methods. AOL will give reward points (that can be traded in for prizes) to respondents. Users will not have to submit their e-mail addresses. The surveys will help measure response to advertisers' online campaigns. The primary objective of this research is to gauge consumers' attitudes and other subjective information that can help media buyers plan their campaigns.

Opinion Place Bases Its Opinions on 1000 Respondents



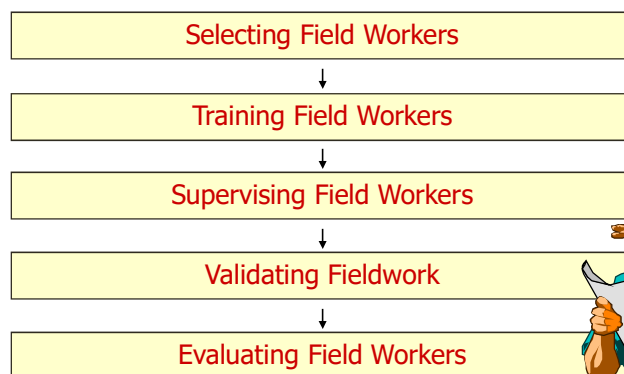
Another advantage of online surveys is that you are sure to reach your target (sample control) and that they are quicker to turn around than traditional surveys like mall intercepts or in-home interviews. They also are cheaper (DMS charges \$20,000 for an online survey, while it costs between \$30,000 and \$40,000 to conduct a mall-intercept survey of 1,000 respondents).

Figure 12.3 A Concept Map for Determining Sample Size for Means



FIELDWORK

Fieldwork/Data Collection Process



Selection of Field Workers

The researcher should:

- Develop job specifications for the project, taking into account the mode of data collection.
- Decide what characteristics the field workers should have.
- Recruit appropriate individuals.

General Qualifications of Field Workers

- **Healthy.** Field workers must have the stamina required to do the job.
- **Outgoing.** The interviewers should be able to establish rapport with the respondents.
- **Communicative.** Effective speaking and listening skills are a great asset.
- **Pleasant appearance.** If the field worker's physical appearance is unpleasant or unusual, the data collected may be biased.
- **Educated.** Interviewers must have good reading and writing skills.
- **Experienced.** Experienced interviewers are likely to do a better job.

Training of Field Workers

- **Making the Initial Contact** – Interviewers should be trained to make opening remarks that will convince potential respondents that their participation is important.
- **Asking the Questions**
 1. Be thoroughly familiar with the questionnaire.
 2. Ask the questions in the order in which they appear in the questionnaire.
 3. Use the exact wording given in the questionnaire.
 4. Read each question slowly.
 5. Repeat questions that are not understood.
 6. Ask every applicable question.
 7. Follow instructions, skip patterns, probe carefully.

Training of Field Workers

- **Probing** – Some commonly used probing techniques:
 1. Repeating the question.
 2. Repeating the respondent's reply.
 3. Using a pause or silent probe.
 4. Boosting or reassuring the respondent.
 5. Eliciting clarification.
 6. Using objective/neutral questions or comments.

Commonly Used Probes and Abbreviations



Standard Interviewer's Probe	Abbreviation
Any other reason?	(AO?)
Any others?	(Other?)
Anything else?	(AE or Else?)
Could you tell me more about your thinking on that?	(Tell more)
How do you mean?	(How mean?)
Repeat question	(RQ)
What do you mean?	(What mean?)
Which would be closer to the way you feel?	(Which closer?)
Why do you feel that way?	(Why?)
Would you tell me what you have in mind?	(What in mind?)

Training of Field Workers

- **Recording the Answers** – Guidelines for recording answers to unstructured questions:
 1. Record responses during the interview.
 2. Use the respondent's own words.
 3. Do not summarize or paraphrase the respondent's answers.
 4. Include everything that pertains to the question objectives.
 5. Include all probes and comments.
 6. Repeat the response as it is written down.
- **Terminating the Interview** – The respondent should be left with a positive feeling about the interview.

Guidelines on Interviewer Training: The Council of American Survey Research Organizations

Training should be conducted under the direction of supervisory personnel and should cover the following:

- 1) The research process: how a study is developed, implemented & reported.
 - 2) Importance of interviewers; need for honesty, objectivity & professionalism.
 - 3) Confidentiality of the respondent & client.
 - 4) Familiarity with market research terminology.
 - 5) Importance of following the exact wording & recording responses verbatim.
 - 6) Purpose & use of probing & clarifying techniques.
 - 7) The reason for & use of classification & respondent information questions.
 - 8) A review of samples of instructions & questionnaires.
 - 9) Importance of the respondent's positive feelings about survey research.
- An interviewer must be trained in the interviewing techniques outlined above.

Guidelines on Supervision: The Council of American Survey Research Organizations

All research projects should be properly supervised. It is the data collection agency's responsibility to:

- 1) Properly supervise interviews.
- 2) See that an agreed-upon proportion of interviewers' telephone calls are monitored.
- 3) Be available to report on the status of the project daily to the project director, unless otherwise instructed.
- 4) Keep all studies, materials, and findings confidential.
- 5) Notify concerned parties if the anticipated schedule is not met.
- 6) Attend all interviewer briefings.
- 7) Keep current & accurate records of the interviewing progress.
- 8) Make sure all interviewers have all materials in time.
- 9) Edit each questionnaire.
- 10) Provide consistent & positive feedback to the interviewers.
- 11) Not falsify any work.



Guidelines on Interviewing: The Council of American Survey Research Organizations

Each interviewer is to follow these techniques for good interviewing:

- 1) Provide his or her full name, if asked by the respondent, as well as a phone number for the research firm.
- 2) Read each question exactly as written. Report any problems to the supervisor as soon as possible.
- 3) Read the questions in the order indicated on the questionnaire, following the proper skip sequences.
- 4) Clarify any question by the respondent in a neutral way.
- 5) Not mislead respondents as to the length of the interview.
- 6) Not reveal the ultimate client's identity unless instructed to do so.
- 7) Keep a tally on and the reason for each terminated interview.
- 8) Remain neutral, do not indicate (dis)agreement with the respondent.

Guidelines on Interviewing: The Council of American Survey Research Organizations

- 9) Speak slowly & distinctly.
- 10) Record all replies verbatim, not paraphrased.
- 11) Avoid unnecessary conversation with the respondent.
- 12) Probe & clarify in a neutral manner for additional comments on all open-ended questions, unless otherwise indicated.
- 13) Write neatly & legibly.
- 14) Check all work for thoroughness before turning in to the supervisor.
- 15) When terminating a respondent, do it neutrally.
- 16) Keep all studies, materials, and findings confidential.
- 17) Not falsify any interviews or any answers to any question.
- 18) Thank the respondent for participating in the study.



Supervision of Field Workers

Supervision of field workers means making sure that they are following the procedures and techniques in which they were trained. Supervision involves quality control and editing, sampling control, control of cheating, and central office control.

- **Quality Control and Editing** – This requires checking to see if the field procedures are being properly implemented.
- **Sampling Control** – The supervisor attempts to ensure that the interviewers are strictly following the sampling plan.
- **Control of Cheating** – Cheating can be minimized through proper training, supervision, and validation.
- **Central Office Control** – Supervisors provide quality and cost-control information to the central office.

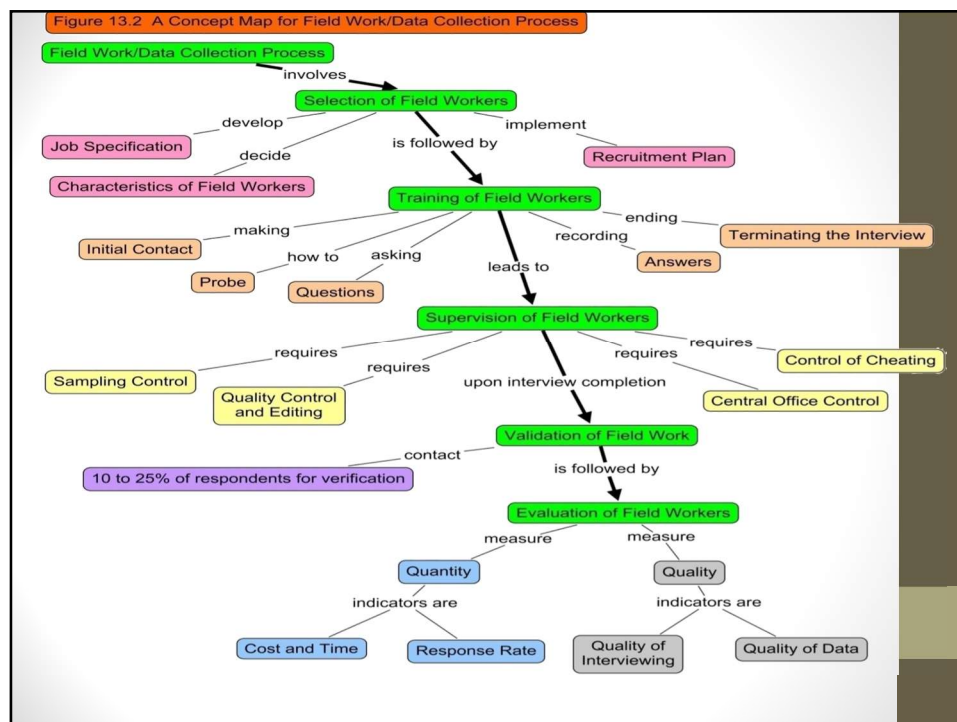
Validation of Fieldwork

Validation:

- The supervisors call 10 - 25% of the respondents to inquire whether the field workers actually conducted the interviews.
- The supervisors ask about the length and quality of the interview, reaction to the interviewer, and basic demographic data.
- The demographic information is cross-checked against the information reported by the interviewers on the questionnaires.

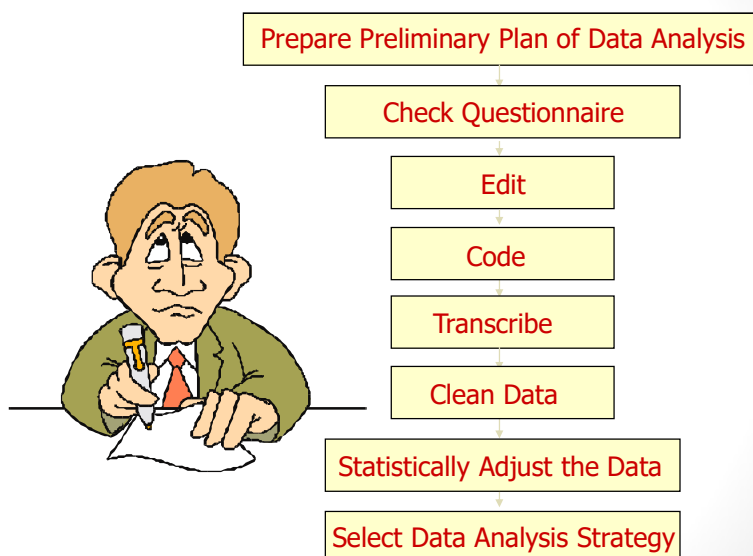
Evaluation of Field Workers

- **Cost and Time.** The interviewers can be compared in terms of the total cost (salary and expenses) per completed interview.
- **Response Rates.** It is important to monitor response rates on a timely basis so that corrective action can be taken if these rates are too low.
- **Quality of Interviewing.** To evaluate interviewers on the quality of interviewing, the supervisor must directly observe the interviewing process.
- **Quality of Data.** The completed questionnaires of each interviewer should be evaluated for the quality of data.



DATA PREPARATION

Data Preparation Process



Questionnaire Checking

A questionnaire returned from the field may be unacceptable for several reasons.

- Parts of the questionnaire may be incomplete.
- The pattern of responses may indicate that the respondent did not understand or follow the instructions.
- The responses show little variance.
- One or more pages are missing.
- The questionnaire is received after the preestablished cutoff date.
- The questionnaire is answered by someone who does not qualify for participation.

Editing

Treatment of Unsatisfactory Results

- **Returning to the Field** – The questionnaires with unsatisfactory responses may be returned to the field, where the interviewers recontact the respondents.
- **Assigning Missing Values** – If returning the questionnaires to the field is not feasible, the editor may assign missing values to unsatisfactory responses.
- **Discarding Unsatisfactory Respondents** – In this approach, the respondents with unsatisfactory responses are simply discarded.

Coding

Coding means assigning a code, usually a number, to each possible response to each question. The code includes an indication of the column position (field) and data record it will occupy.

Coding Questions

- **Fixed field codes**, which mean that the number of records for each respondent is the same and the same data appear in the same column(s) for all respondents, are highly desirable.
- If possible, standard codes should be used for missing data. Coding of structured questions is relatively simple, since the response options are predetermined.
- In questions that permit a large number of responses, each possible response option should be assigned a separate column.

Coding

Guidelines for Coding Unstructured Questions:

- Category codes should be mutually exclusive and collectively exhaustive.
- Only a few (10% or less) of the responses should fall into the "other" category.
- Category codes should be assigned for critical issues even if no one has mentioned them.
- Data should be coded to retain as much detail as possible.

Codebook

A **codebook** contains coding instructions and the necessary information about variables in the data set. A codebook generally contains the following information:

- column number
- record number
- variable number
- variable name
- question number
- instructions for coding

Coding Questionnaires

- The respondent code and the record number appear on each record in the data.
- The first record contains the additional codes: project code, interviewer code, date and time codes, and validation code.
- It is a good practice to insert blanks between parts.

Codebook Excerpt				
Column Number	Variable Number	Variable Name	Question Number	Coding Instructions
1	1	ID		1 to 20 as coded
2	2	Preference	1	Input the number circled. 1=Weak Preference 7=Strong Preference
3	3	Quality	2	Input the number circled. 1=Poor 7=Excellent
4	4	Quantity	3	Input the number circled. 1=Poor 7=Excellent
5	5	Value	4	Input the number circled. 1=Poor 7=Excellent
6	6	Service	5	Input the number circled. 1=Poor 7=Excellent

Codebook Excerpt (Cont.)				
Column Number	Variable Number	Variable Name	Question Number	Coding Instructions
7	7	Income	6	Input the number circled. 1 = Less than \$20,000 2 = \$20,000 to 34,999 3 = \$35,000 to 49,999 4 = \$50,000 to 74,999 5 = \$75,000 to 99,999 6 = \$100,00 or more

Example of Questionnaire Coding

Finally, in this part of the questionnaire we would like to ask you some background information for classification purposes.

PART D

Record #

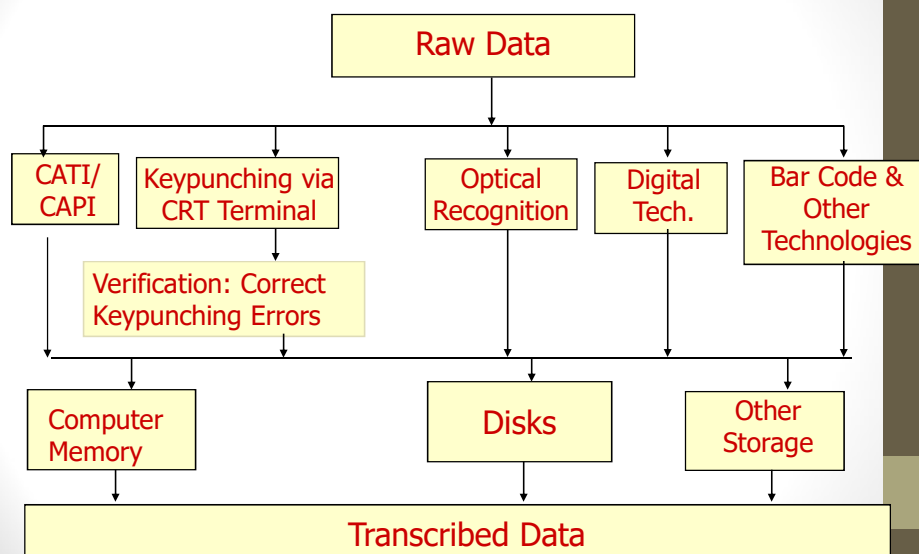
1. This questionnaire was answered by
 1. ☐ Primarily the male head of household
 2. ☐ Primarily the female head of household
 3. ☐ Jointly by the male and female heads of household

(29)
 2. Marital Status
 1. ☐ Married
 2. ☐ Never Married
 3. ☐ Divorced/Separated/Widowed

(30)
 3. What is the total number of family members living at home? _____ (31 - 32)
 4. Number of children living at home:
 - a. Under six years _____ (33)
 - b. Over six years _____ (34)
 5. Number of children not living at home _____ (35)
 6. Number of years of formal education which you (and your spouse, if applicable) have completed. (please circle)
- | | High School | College | Graduate | |
|-----------|----------------------|-------------|---------------------------|---------|
| a. You | 8 or less 9 10 11 12 | 13 14 15 16 | 17 18 19 20 21 22 or more | (36-37) |
| b. Spouse | 8 or less 9 10 11 12 | 13 14 15 16 | 17 18 19 20 21 22 or more | (37-38) |
7. a. Your age: _____ (40-41)
 - b. Age of spouse (if applicable) _____ (42-43)
 8. If employed please indicate your household's occupations by checking the appropriate category.

	44 Male Head	45 Female Head
1. Professional and technical	_____	_____
2. Managers and administrators	_____	_____
3. Sales workers	_____	_____
4. Clerical and kindred workers	_____	_____
5. Craftsman/operative /laborers	_____	_____
6. Homemakers	_____	_____
7. Others (please specify)	_____	_____
8. Not applicable	_____	_____
 9. Is your place of residence presently owned by household?
 1. Owned _____ (46)
 2. Rented _____
 10. How many years have you been residing in the greater Atlanta area? _____ years. (47-48)

Data Transcription



Data Cleaning Consistency Checks

Consistency checks identify data that are out of range, logically inconsistent, or have extreme values.

- Computer packages like SPSS, SAS, EXCEL and MINITAB can be programmed to identify out-of-range values for each variable and print out the respondent code, variable code, variable name, record number, column number, and out-of-range value.
- Extreme values should be closely examined.

Data Cleaning Treatment of Missing Responses

- **Substitute a Neutral Value** – A neutral value, typically the mean response to the variable, is substituted for the missing responses.
- **Substitute an Imputed Response** – The respondents' pattern of responses to other questions are used to impute or calculate a suitable response to the missing questions.
- In **casewise deletion**, cases, or respondents, with any missing responses are discarded from the analysis.
- In **pairwise deletion**, instead of discarding all cases with any missing values, the researcher uses only the cases or respondents with complete responses for each calculation.

Statistically Adjusting the Data Weighting

- In **weighting**, each case or respondent in the database is assigned a weight to reflect its importance relative to other cases or respondents.
- Weighting is most widely used to make the sample data more representative of a target population on specific characteristics.
- Yet another use of weighting is to adjust the sample so that greater importance is attached to respondents with certain characteristics.

Statistically Adjusting the Data

Use of Weighting for Representativeness

Years of Education	Sample Percentage	Population Percentage	Weight
Elementary School			
0 to 7 years	2.49	4.23	1.70
8 years	1.26	2.19	1.74
High School			
1 to 3 years	6.39	8.65	1.35
4 years	25.39	29.24	1.15
College			
1 to 3 years	22.33	29.42	1.32
4 years	15.02	12.01	0.80
5 to 6 years	14.94	7.36	0.49
7 years or more	12.18	6.90	0.57
Totals	100.00	100.00	

Statistically Adjusting the Data – Variable Respecification

- **Variable respecification** involves the transformation of data to create new variables or modify existing variables.
- e.g., the researcher may create new variables that are composites of several other variables.
- Dummy variables are used for respecifying categorical variables. The general rule is that to respecify a categorical variable with K categories, $K-1$ dummy variables are needed.

Statistically Adjusting the Data – Variable Respecification

Product Usage Category	Original Variable Code	Dummy Variable Code		
		X_1	X_2	X_3
Nonusers	1	1	0	0
Light users	2	0	1	0
Medium users	3	0	0	1
Heavy users	4	0	0	0

Note that $X_1 = 1$ for nonusers and 0 for all others. Likewise, $X_2 = 1$ for light users and 0 for all others, and $X_3 = 1$ for medium users and 0 for all others. In analyzing the data, X_1 , X_2 , and X_3 are used to represent all user/nonuser groups.

Statistically Adjusting the Data – Scale Transformation and Standardization

Scale transformation involves a manipulation of scale values to ensure comparability with other scales or otherwise make the data suitable for analysis.

A more common transformation procedure is **standardization**. Standardized scores, Z_i , may be obtained as: $Z_i = (X_i - \bar{X})/s_x$

Selecting a Data Analysis Strategy

Earlier Steps (1, 2, & 3) of the Marketing Research Process



Known Characteristics of the Data



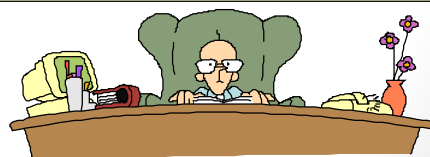
Properties of Statistical Techniques



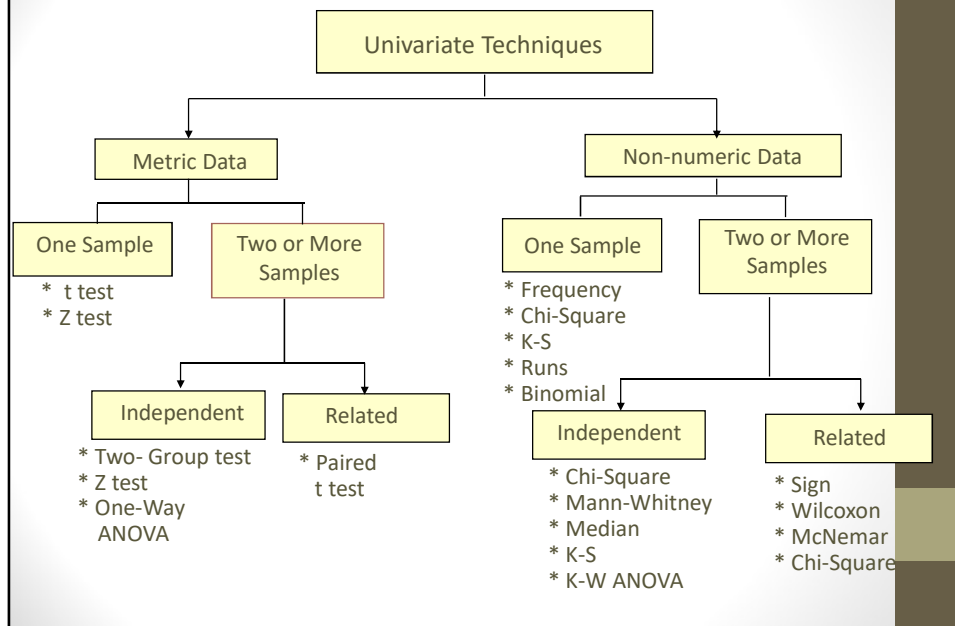
Background and Philosophy of the Researcher



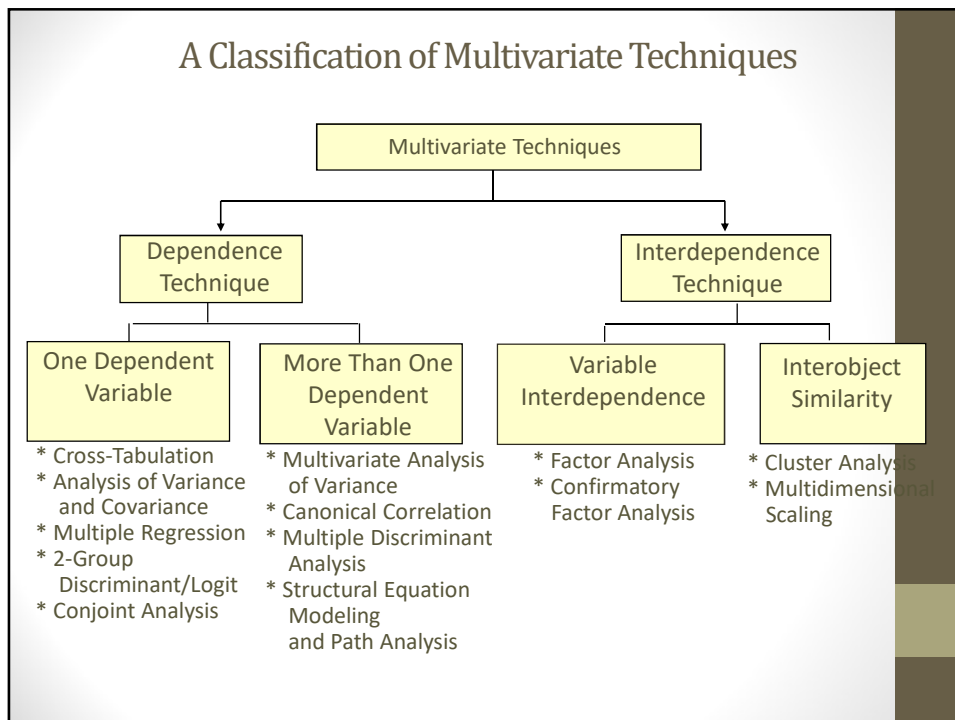
Data Analysis Strategy



A Classification of Univariate Techniques



A Classification of Multivariate Techniques



Nielsen's Internet Survey: Does It Carry Any Weight?



The Nielsen Media Research Company, a longtime player in television-related marketing research, has come under fire from the various TV networks for its surveying techniques. Additionally, in another potentially large, new revenue business, Internet surveying, Nielsen is encountering serious questions concerning the validity of its survey results. Due to the tremendous impact of electronic commerce on the business world, advertisers need to know how many people are doing business on the Internet in order to decide if it would be lucrative to place their ads online.

Nielsen performed a survey for CommerceNet, a group of companies that includes Sun Microsystems and American Express, to help determine the number of total users on the Internet.

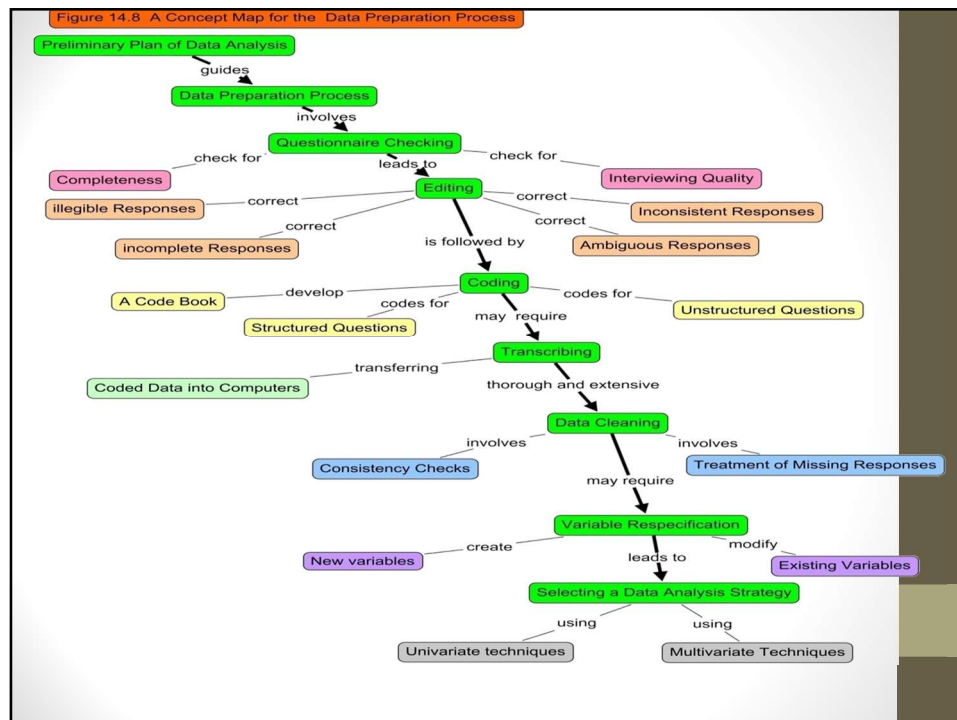
Nielsen's Internet Survey: Does It Carry Any Weight?

Statisticians believe the numbers reported by Nielsen may be incorrect in that the weighting used to help match the sample to the population may be flawed. Weighting must be used to prevent research from being skewed toward one demographic segment. Nielsen weighted for gender but not for education which may have skewed the population toward educated adults.



Nielsen's Internet Survey: Does It Carry Any Weight?

Nielsen then weighted the survey by age and income after they had already weighted it for gender. Statisticians also feel that this is incorrect because weighting must occur simultaneously, not in separate calculations. Nielsen does not believe the concerns about their sample are legitimate and feel that they have not erred in weighting the survey. However, due to the fact that most third parties have not endorsed Nielsen's methods, the validity of their research remains to be established.



DATA ANALYTIC METHODS AND HYPOTHESIS TESTING

Internet Usage Data

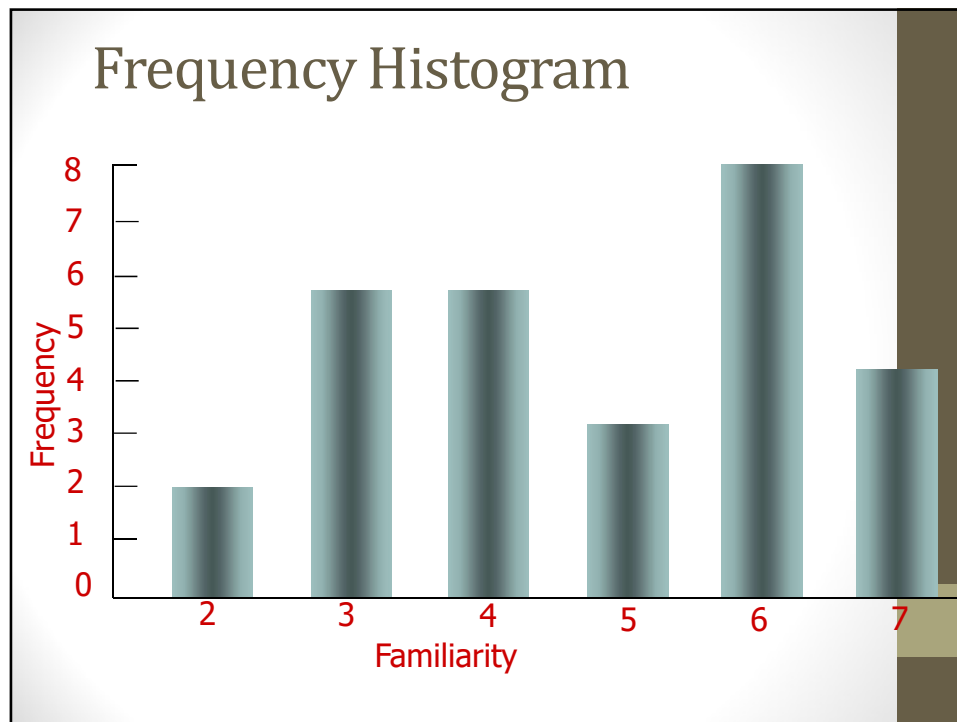
Respondent Number	Sex	Familiarity	Internet Usage	Attitude Toward Internet	Technology	Usage of Internet Shopping	Banking
1	1.00	7.00	14.00	7.00	6.00	1.00	1.00
2	2.00	2.00	2.00	3.00	3.00	2.00	2.00
3	2.00	3.00	3.00	4.00	3.00	1.00	2.00
4	2.00	3.00	3.00	7.00	5.00	1.00	2.00
5	1.00	7.00	13.00	7.00	7.00	1.00	1.00
6	2.00	4.00	6.00	5.00	4.00	1.00	2.00
7	2.00	2.00	2.00	4.00	5.00	2.00	2.00
8	2.00	3.00	6.00	5.00	4.00	2.00	2.00
9	2.00	3.00	6.00	6.00	4.00	1.00	2.00
10	1.00	9.00	15.00	7.00	6.00	1.00	2.00
11	2.00	4.00	3.00	4.00	3.00	2.00	2.00
12	2.00	5.00	4.00	6.00	4.00	2.00	2.00
13	1.00	6.00	9.00	6.00	5.00	2.00	1.00
14	1.00	6.00	8.00	3.00	2.00	2.00	2.00
15	1.00	6.00	5.00	5.00	4.00	1.00	2.00
16	2.00	4.00	3.00	4.00	3.00	2.00	2.00
17	1.00	6.00	9.00	5.00	3.00	1.00	1.00
18	1.00	4.00	4.00	5.00	4.00	1.00	2.00
19	1.00	7.00	14.00	6.00	6.00	1.00	1.00
20	2.00	6.00	6.00	6.00	4.00	2.00	2.00
21	1.00	6.00	9.00	4.00	2.00	2.00	2.00
22	1.00	5.00	5.00	5.00	4.00	2.00	1.00
23	2.00	3.00	2.00	4.00	2.00	2.00	2.00
24	1.00	7.00	15.00	6.00	6.00	1.00	1.00
25	2.00	6.00	6.00	5.00	3.00	1.00	2.00
26	1.00	6.00	13.00	6.00	6.00	1.00	1.00
27	2.00	5.00	4.00	5.00	5.00	1.00	1.00
28	2.00	4.00	2.00	3.00	2.00	2.00	2.00
29	1.00	4.00	4.00	5.00	3.00	1.00	2.00
30	1.00	3.00	3.00	7.00	5.00	1.00	2.00

Frequency Distribution

- In a **frequency distribution**, one variable is considered at a time.
- A frequency distribution for a variable produces a table of frequency counts, percentages, and cumulative percentages for all the values associated with that variable.

Frequency of Familiarity with the Internet

<u>Value label</u>	<u>Value</u>	<u>Frequency (n)</u>	<u>Percentage</u>	<u>Valid Percentage</u>	<u>Cumulative Percentage</u>
Not so familiar	1	0	0.0	0.0	0.0
	2	2	6.7	6.9	6.9
	3	6	20.0	20.7	27.6
	4	6	20.0	20.7	48.3
	5	3	10.0	10.3	58.6
	6	8	26.7	27.6	86.2
Very familiar	7	4	13.3	13.8	100.0
Missing	9	1	3.3		
	TOTAL	30	100.0	100.0	



Statistics Associated with Frequency Distribution: Measures of Location

- The **mean**, or average value, is the most commonly used measure of central tendency. The mean, $\bar{X}$, is given by

$$\bar{X} = \sum_{i=1}^n X_i / n$$

Where,

X_i = Observed values of the variable X

n = Number of observations (sample size)

- The **mode** is the value that occurs most frequently. It represents the highest peak of the distribution. The mode is a good measure of location when the variable is inherently categorical or has otherwise been grouped into categories.

Statistics Associated with Frequency Distribution: Measures of Location

- The **median** of a sample is the middle value when the data are arranged in ascending or descending order. If the number of data points is even, the median is usually estimated as the midpoint between the two middle values – by adding the two middle values and dividing their sum by 2. The median is the 50th percentile.

Statistics Associated with Frequency Distribution: Measures of Variability

- The **range** measures the spread of the data. It is simply the difference between the largest and smallest values in the sample.

$$\text{Range} = X_{\text{largest}} - X_{\text{smallest}}$$

- The **interquartile range** is the difference between the 75th and 25th percentile. For a set of data points arranged in order of magnitude, the p^{th} percentile is the value that has $p\%$ of the data points below it and $(100 - p)\%$ above it.

Statistics Associated with Frequency Distribution: Measures of Variability

- The **variance** is the mean squared deviation from the mean. The variance can never be negative.
- The **standard deviation** is the square root of the variance.

$$s_x = \sqrt{\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}}$$

- The **coefficient of variation** is the ratio of the standard deviation to the mean expressed as a percentage, and is a unitless measure of relative variability.

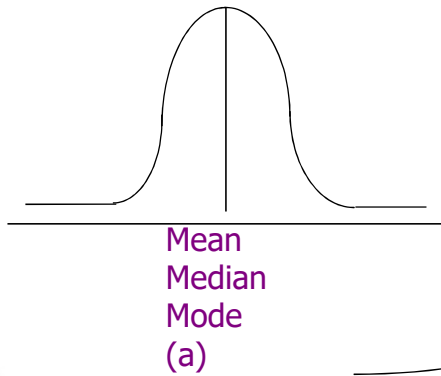
$$CV = s_x / \bar{X}$$

Statistics Associated with Frequency Distribution: Measures of Shape

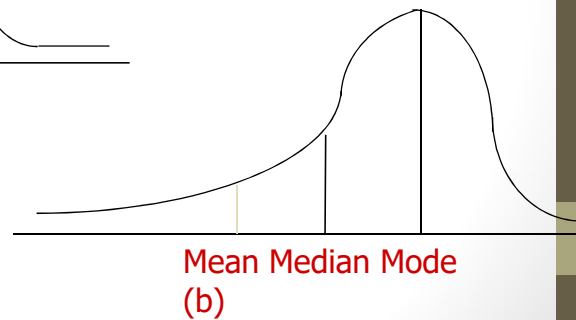
- **Skewness.** The tendency of the deviations from the mean to be larger in one direction than in the other. It can be thought of as the tendency for one tail of the distribution to be heavier than the other.
- **Kurtosis** is a measure of the relative peakedness or flatness of the curve defined by the frequency distribution. The kurtosis of a normal distribution is zero. If the kurtosis is positive, then the distribution is more peaked than a normal distribution. A negative value means that the distribution is flatter than a normal distribution.

Skewness of a Distribution

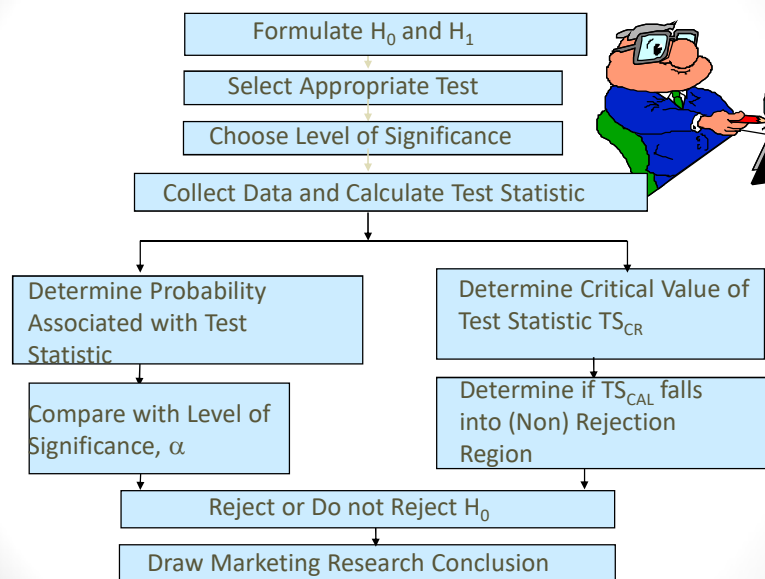
Symmetric Distribution



Skewed Distribution



Steps Involved in Hypothesis Testing



Logic of Testing the Null Hypothesis

- The null hypothesis represents the original state of nature; that is, nothingness
- If the null hypothesis is true, the probability of finding a test statistic out of the confidence interval is α
- α is smaller enough to be considered unlikely.
- If the test statistic that is found is actually out of the confidence interval, this test statistics is considered surprising.
 - If the null hypothesis is true, this test statistic is unlikely to be found.
 - If this test statistic is found, the null hypothesis is unlikely to be true.
- There is nothing about the alternative hypothesis.

[261]

A General Procedure for Hypothesis Testing

Step 1: Formulate the Hypothesis

- A **null hypothesis** is a statement of the status quo, one of no difference or no effect. If the null hypothesis is not rejected, no changes will be made.
- An **alternative hypothesis** is one in which some difference or effect is expected. Accepting the alternative hypothesis will lead to changes in opinions or actions.
- The null hypothesis refers to a specified value of the population parameter (e.g., μ , σ , π), not a sample statistic (e.g., $\bar{X}$).

A General Procedure for Hypothesis Testing

Step 1: Formulate the Hypothesis

- A null hypothesis may be rejected, but it can never be accepted based on a single test. In classical hypothesis testing, there is no way to determine whether the null hypothesis is true.
- In marketing research, the null hypothesis is formulated in such a way that its rejection leads to the acceptance of the desired conclusion. The alternative hypothesis represents the conclusion for which evidence is sought.

$$H_0: \pi \leq 0.40$$

$$H_1: \pi > 0.40$$

A General Procedure for Hypothesis Testing

Step 1: Formulate the Hypothesis

- The test of the null hypothesis is a **one-tailed test**, because the alternative hypothesis is expressed directionally. If that is not the case, then a **two-tailed test** would be required, and the hypotheses would be expressed as:

$$H_0: \pi = 0.40$$

$$H_1: \pi \neq 0.40$$

A General Procedure for Hypothesis Testing

Step 2: Select an Appropriate Test

- The **test statistic** measures how close the sample has come to the null hypothesis.
- The test statistic often follows a well-known distribution, such as the normal, t , or chi-square distribution.
- In our example, the **z statistic**, which follows the standard normal distribution, would be appropriate.

$$z = \frac{p - \pi}{\sigma_p}$$

where

$$\sigma_p = \sqrt{\frac{\pi(1 - \pi)}{n}}$$

A General Procedure for Hypothesis Testing

Step 3: Choose a Level of Significance

Type I Error

- **Type I error** occurs when the sample results lead to the rejection of the null hypothesis when it is in fact true.
- The probability of type I error (α) is also called the **level of significance**.

Type II Error

- **Type II error** occurs when, based on the sample results, the null hypothesis is not rejected when it is in fact false.
- The probability of type II error is denoted by β .
- Unlike α , which is specified by the researcher, the magnitude of β depends on the actual value of the population parameter (proportion).

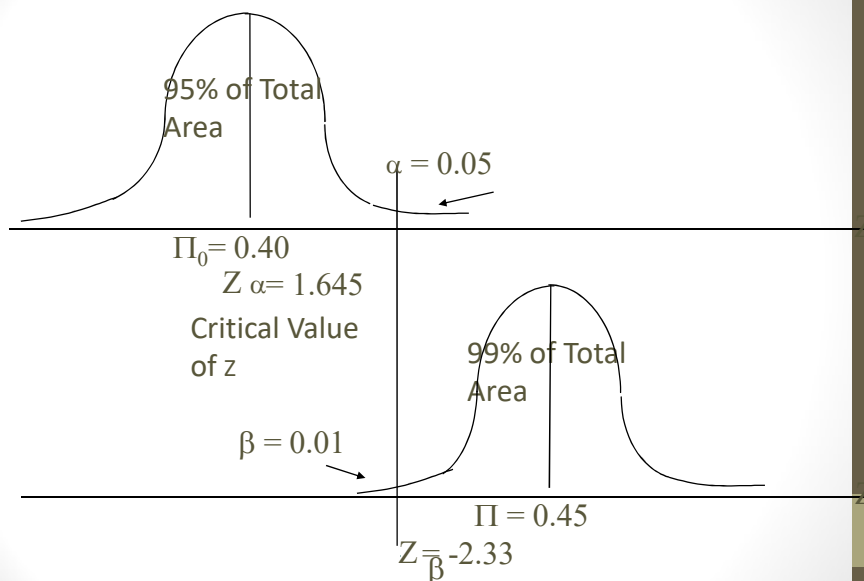
A General Procedure for Hypothesis Testing

Step 3: Choose a Level of Significance

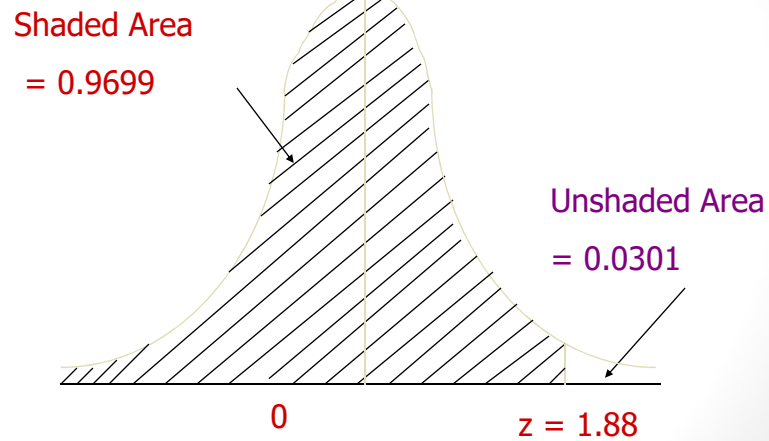
Power of a Test

- The **power of a test** is the probability ($1 - \beta$) of rejecting the null hypothesis when it is false and should be rejected.
- Although β is unknown, it is related to α . An extremely low value of α (e.g., $\alpha = 0.001$) will result in intolerably high β errors.
- So it is necessary to balance the two types of errors.

Probabilities of Type I & Type II Error



Probability of z with a One-Tailed Test



A General Procedure for Hypothesis Testing Step 4: Collect Data and Calculate Test Statistic

- The required data are collected and the value of the test statistic computed.
- In our example, the value of the sample proportion is $\hat{p} = 17/30 = 0.567$.
- The value of σ_p can be determined as follows:

$$\begin{aligned}\sigma_p &= \sqrt{\frac{\pi(1 - \pi)}{n}} \\ &= \sqrt{\frac{(0.40)(0.6)}{30}} \\ &= 0.089\end{aligned}$$

A General Procedure for Hypothesis Testing

Step 4: Collect Data and Calculate Test Statistic

The test statistic z can be calculated as follows:

$$\begin{aligned}
 z &= \frac{\hat{p} - \pi}{\sigma_p} \\
 &= \frac{0.567 - 0.40}{0.089} \\
 &= 1.88
 \end{aligned}$$

A General Procedure for Hypothesis Testing

Step 5: Determine the Probability (Critical Value)

- Using standard normal tables (Table 2 of the Statistical Appendix), the probability of obtaining a z value of 1.88 can be calculated (see Figure 15.5).
- The shaded area between $-\infty$ and 1.88 is 0.9699. Therefore, the area to the right of $z = 1.88$ is $1.0000 - 0.9699 = 0.0301$.
- Alternatively, the critical value of z , which will give an area to the right side of the critical value of 0.05, is between 1.64 and 1.65 and equals 1.645.
- Note, in determining the critical value of the test statistic, the area to the right of the critical value is either α or $\alpha/2$. It is α for a one-tail test and $\alpha/2$ for a two-tail test.

A General Procedure for Hypothesis Testing Steps 6 & 7:
Compare the Probability (Critical Value) & Making the Decision

- If the probability associated with the calculated or observed value of the test statistic (TSCAL) is less than the level of significance (α), the null hypothesis is rejected.
- The probability associated with the calculated or observed value of the test statistic is 0.0301. This is the probability of getting a p value of 0.567 when $\Pi = 0.40$. This is less than the level of significance of 0.05. Hence, the null hypothesis is rejected.
- Alternatively, if the absolute calculated value of the test statistic ($|TS_{CAL}|$) is greater than the absolute critical value of the test statistic ($|TS_{CR}|$), the null hypothesis is rejected.

A General Procedure for Hypothesis Testing Steps 6 & 7:
Compare the Probability (Critical Value) & Making the Decision

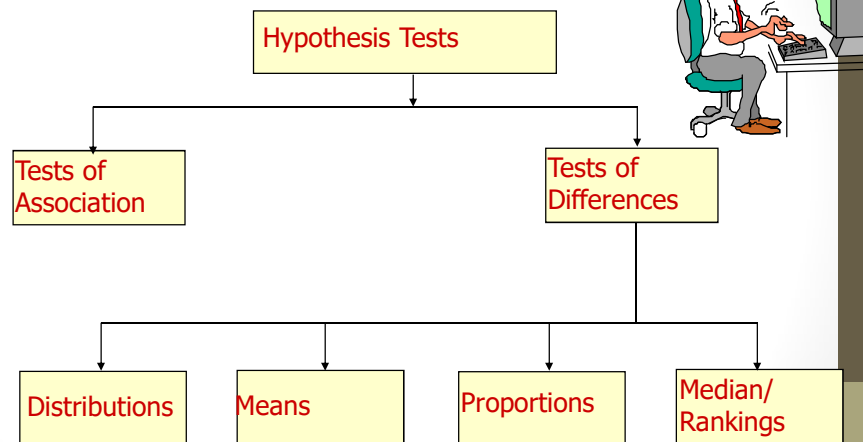
- The calculated value of the test statistic $z = 1.88$ lies in the rejection region, beyond the value of 1.645. Again, the same conclusion to reject the null hypothesis is reached.
- Note that the two ways of testing the null hypothesis are equivalent but mathematically opposite in the direction of comparison.
- If the probability of $TS_{CAL} < \text{significance level } (\alpha)$ then reject H_0 but if $|TS_{CAL}| > |TS_{CR}|$ then reject H_0 .

A General Procedure for Hypothesis Testing

Step 8: Marketing Research Conclusion

- The conclusion reached by hypothesis testing must be expressed in terms of the marketing research problem.
- In our example, we conclude that there is evidence that the proportion of Internet users who shop via the Internet is significantly greater than 0.40. Hence, the recommendation to the department store would be to introduce the new Internet shopping service.

A Broad Classification of Hypothesis Tests



Cross-Tabulation

- While a frequency distribution describes one variable at a time, a **cross-tabulation** describes two or more variables simultaneously.
- Cross-tabulation results in tables that reflect the joint distribution of two or more variables with a limited number of categories or distinct values, e.g., Table 15.3.

Gender and Internet Usage

Gender			
Internet Usage	Male	Female	Row Total
Light (1)	5	10	15
Heavy (2)	10	5	15
Column Total	15	15	



Two Variables Cross-Tabulation

- Since two variables have been cross-classified, percentages could be computed either columnwise, based on column totals (Table 15.4), or rowwise, based on row totals (Table 15.5).
- The general rule is to compute the percentages in the direction of the independent variable, across the dependent variable. The correct way of calculating percentages is as shown in Table 15.4.

Internet Usage by Gender

Table 15.4

Internet Usage	Gender	
	Male	Female
Light	33.3%	66.7%
Heavy	66.7%	33.3%
Column total	100%	100%



Gender by Internet Usage

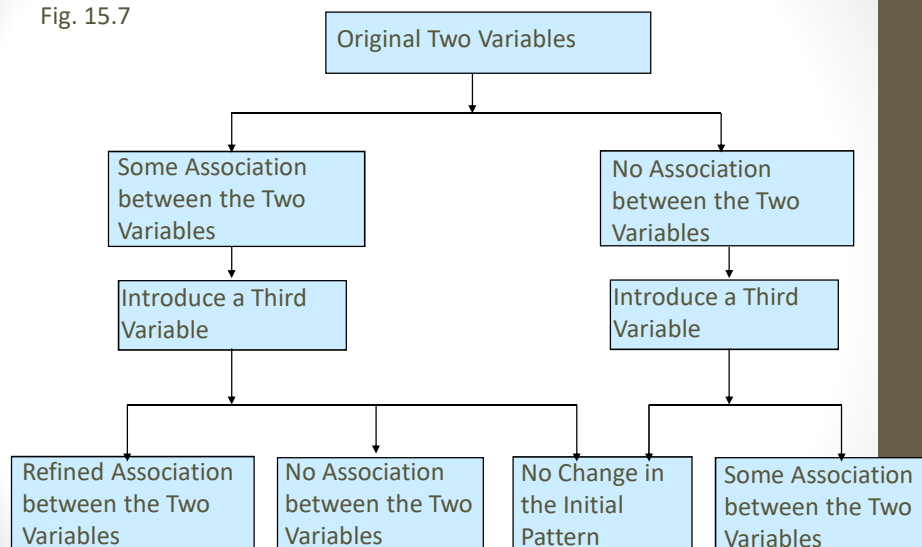
Table 15.5

Gender	Internet Usage		Total
	Light	Heavy	
Male	33.3%	66.7%	100.0%
Female	66.7%	33.3%	100.0%



Introduction of a Third Variable in Cross-Tabulation

Fig. 15.7



Three Variables Cross-Tabulation Refine an Initial Relationship

As shown in Figure 15.7, the introduction of a third variable can result in four possibilities:

- As can be seen from Table 15.6, 52% of unmarried respondents fell in the high-purchase category, as opposed to 31% of the married respondents. Before concluding that unmarried respondents purchase more fashion clothing than those who are married, a third variable, the buyer's sex, was introduced into the analysis.
- As shown in Table 15.7, in the case of females, 60% of the unmarried fall in the high-purchase category, as compared to 25% of those who are married. On the other hand, the percentages are much closer for males, with 40% of the unmarried and 35% of the married falling in the high purchase category.
- Hence, the introduction of sex (third variable) has refined the relationship between marital status and purchase of fashion clothing (original variables). Unmarried respondents are more likely to fall in the high purchase category than married ones, and this effect is much more pronounced for females than for males.

Purchase of Fashion Clothing by Marital Status

Purchase of Fashion Clothing	Current Marital Status	
	Married	Unmarried
High	31%	52%
Low	69%	48%
Column	100%	100%
Number of respondents	700	300



Purchase of Fashion Clothing by Marital Status

Purchase of Fashion Clothing	Sex			
	Male		Female	
	Married	Not Married	Married	Not Married
High	35%	40%	25%	60%
Low	65%	60%	75%	40%
Column totals	100%	100%	100%	100%
Number of cases	400	120	300	180



Three Variables Cross-Tabulation Initial Relationship was Spurious

- Table 15.8 shows that 32% of those with college degrees own an expensive automobile, as compared to 21% of those without college degrees. Realizing that income may also be a factor, the researcher decided to reexamine the relationship between education and ownership of expensive automobiles in light of income level.
- In Table 15.9, the percentages of those with and without college degrees who own expensive automobiles are the same for each of the income groups. When the data for the high income and low income groups are examined separately, the association between education and ownership of expensive automobiles disappears, indicating that the initial relationship observed between these two variables was spurious.

Ownership of Expensive Automobiles by Education Level

Table 15.8

Own Expensive Automobile	Education	
	College Degree	No College Degree
Yes	32%	21%
No	68%	79%
Column totals	100%	100%
Number of cases	250	750



Ownership of Expensive Automobiles by Education Level and Income Levels

Table 15.9

Own Expensive Automobile	Income			
	Low Income		High Income	
	College Degree	No College Degree	College Degree	No College Degree
Yes	20%	20%	40%	40%
No	80%	80%	60%	60%
Column totals	100%	100%	100%	100%
Number of respondents	100	700	150	50

Three Variables Cross-Tabulation Reveal Suppressed Association

- Table 15.10 shows no association between desire to travel abroad and age.
- When sex was introduced as the third variable, Table 15.11 was obtained. Among men, 60% of those under 45 indicated a desire to travel abroad, as compared to 40% of those 45 or older. The pattern was reversed for women, where 35% of those under 45 indicated a desire to travel abroad as opposed to 65% of those 45 or older.
- Since the association between desire to travel abroad and age runs in the opposite direction for males and females, the relationship between these two variables is masked when the data are aggregated across sex, as in Table 15.10.
- But when the effect of sex is controlled, as in Table 15.11, the suppressed association between desire to travel abroad and age is revealed for the separate categories of males and females.

Desire to Travel Abroad by Age

Table 15.10

Desire to Travel Abroad	Age	
	Less than 45	45 or More
Yes	50%	50%
No	50%	50%
Column totals	100%	100%
Number of respondents	500	500



Desire to Travel Abroad by Age and Gender

Table 15.11

Desire to Travel Abroad	Sex			
	Male Age		Female Age	
	< 45	>=45	<45	>=45
Yes	60%	40%	35%	65%
No	40%	60%	65%	35%
Column totals	100%	100%	100%	100%
Number of Cases	300	300	200	200



Three Variables Cross-Tabulations

No Change in Initial Relationship

- Consider the cross-tabulation of family size and the tendency to eat out frequently in fast-food restaurants, as shown in Table 15.12. No association is observed.
- When income was introduced as a third variable in the analysis, Table 15.13 was obtained. Again, no association was observed.

Eating Frequently in Fast-Food Restaurants by Family Size

Table 15.12

Eat Frequently in Fast-Food Restaurants	Family Size	
	Small	Large
Yes	65%	65%
No	35%	35%
Column totals	100%	100%
Number of cases	500	500



Eating Frequently in Fast-Food Restaurants by Family Size and Income

Table 15.13

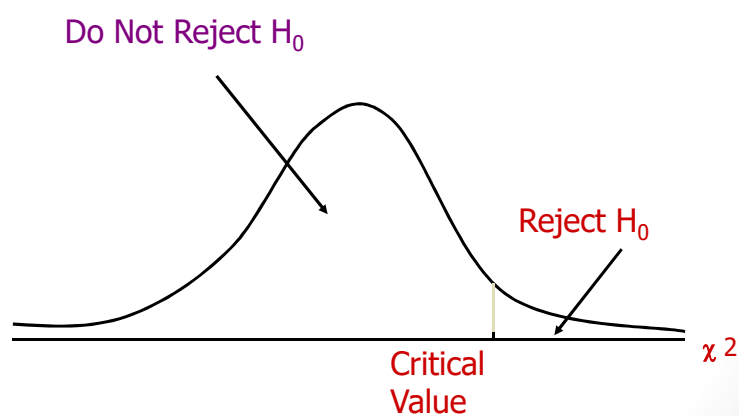
	Income				
Eat Frequently in Fast-Food Restaurants	Low			High	
	Family size			Family size	
	Small	Large		Small	Large
Yes	65%	65%		65%	65%
No	35%	35%		35%	35%
Column totals	100%	100%		100%	100%
Number of respondents	250	250		250	250



Statistics Associated with Cross-Tabulation Chi-Square

- To determine whether a systematic association exists, the probability of obtaining a value of chi-square as large or larger than the one calculated from the cross-tabulation is estimated.
- An important characteristic of the chi-square statistic is the number of degrees of freedom (df) associated with it. That is, $df = (r - 1) \times (c - 1)$.
- The null hypothesis (H_0) of no association between the two variables will be rejected only when the calculated value of the test statistic is greater than the critical value of the chi-square distribution with the appropriate degrees of freedom, as shown in Figure 15.8.

Chi-square Distribution



Statistics Associated with Cross-Tabulation Chi-Square

- The **chi-square statistic** (χ^2) is used to test the statistical significance of the observed association in a cross-tabulation.
- The expected frequency for each cell can be calculated by using a simple formula:

$$f_e = \frac{n_r n_c}{n}$$

where

n_r	= total number in the row
n_c	= total number in the column
n	= total sample size

Statistics Associated with Cross-Tabulation Chi-Square

For the data in Table 15.3, the expected frequencies for the cells going from left to right and from top to bottom, are:

$$\frac{15 \times 15}{30} = 7.50 \quad \frac{15 \times 15}{30} = 7.50$$

$$\frac{15 \times 15}{30} = 7.50 \quad \frac{15 \times 15}{30} = 7.50$$

Then the value of χ^2 is calculated as follows:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Statistics Associated with Cross-Tabulation Chi-Square

For the data in Table 15.3, the value of χ^2 is calculated as:

$$= \frac{(5 - 7.5)^2}{7.5} + \frac{(10 - 7.5)^2}{7.5} + \frac{(10 - 7.5)^2}{7.5} + \frac{(5 - 7.5)^2}{7.5}$$

$$= 0.833 + 0.833 + 0.833 + 0.833$$

$$= 3.333$$

Statistics Associated with Cross-Tabulation Chi-Square

- The **chi-square distribution** is a skewed distribution whose shape depends solely on the number of degrees of freedom. As the number of degrees of freedom increases, the chi-square distribution becomes more symmetrical.
- Table 3 in the Statistical Appendix contains upper-tail areas of the chi-square distribution for different degrees of freedom. For 1 degree of freedom, the probability of exceeding a chi-square value of 3.841 is 0.05.
- For the cross-tabulation given in Table 15.3, there are $(2-1) \times (2-1) = 1$ degree of freedom. The calculated chi-square statistic had a value of 3.333. Since this is less than the critical value of 3.841, the null hypothesis of no association can not be rejected indicating that the association is not statistically significant at the 0.05 level.

Statistics Associated with Cross-Tabulation Phi Coefficient

- The **phi coefficient** (ϕ) is used as a measure of the strength of association in the special case of a table with two rows and two columns (a 2 x 2 table).
- The phi coefficient is proportional to the square root of the chi-square statistic

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

- It takes the value of 0 when there is no association, which would be indicated by a chi-square value of 0 as well. When the variables are perfectly associated, phi assumes the value of 1 and all the observations fall just on the main or minor diagonal.

Statistics Associated with Cross-Tabulation Contingency Coefficient

- While the phi coefficient is specific to a 2 x 2 table, the **contingency coefficient** (C) can be used to assess the strength of association in a table of any size.

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

- The contingency coefficient varies between 0 and 1.
- The maximum value of the contingency coefficient depends on the size of the table (number of rows and number of columns). For this reason, it should be used only to compare tables of the same size.

Statistics Associated with Cross-Tabulation

Cramer's V

- **Cramer's V** is a modified version of the phi correlation coefficient, ϕ , and is used in tables larger than 2 x 2.

or
$$V = \sqrt{\frac{\phi^2}{\min(r-1), (c-1)}}$$

$$V = \sqrt{\frac{\chi^2/n}{\min(r-1), (c-1)}}$$

Statistics Associated with Cross-Tabulation

Lambda Coefficient

- **Asymmetric lambda** measures the percentage improvement in predicting the value of the dependent variable, given the value of the independent variable.
- Lambda also varies between 0 and 1. A value of 0 means no improvement in prediction. A value of 1 indicates that the prediction can be made without error. This happens when each independent variable category is associated with a single category of the dependent variable.
- Asymmetric lambda is computed for each of the variables (treating it as the dependent variable).
- A **symmetric lambda** is also computed, which is a kind of average of the two asymmetric values. The symmetric lambda does not make an assumption about which variable is dependent. It measures the overall improvement when prediction is done in both directions.

Statistics Associated with Cross-Tabulation Other Statistics

- Other statistics like tau b, tau c, and gamma are available to measure association between two ordinal-level variables. Both tau b and tau c adjust for ties.
- **Tau b** is the most appropriate with square tables in which the number of rows and the number of columns are equal. Its value varies between +1 and -1.
- For a rectangular table in which the number of rows is different than the number of columns, **tau c** should be used.
- **Gamma** does not make an adjustment for either ties or table size. Gamma also varies between +1 and -1 and generally has a higher numerical value than tau b or tau c.

Cross-Tabulation in Practice

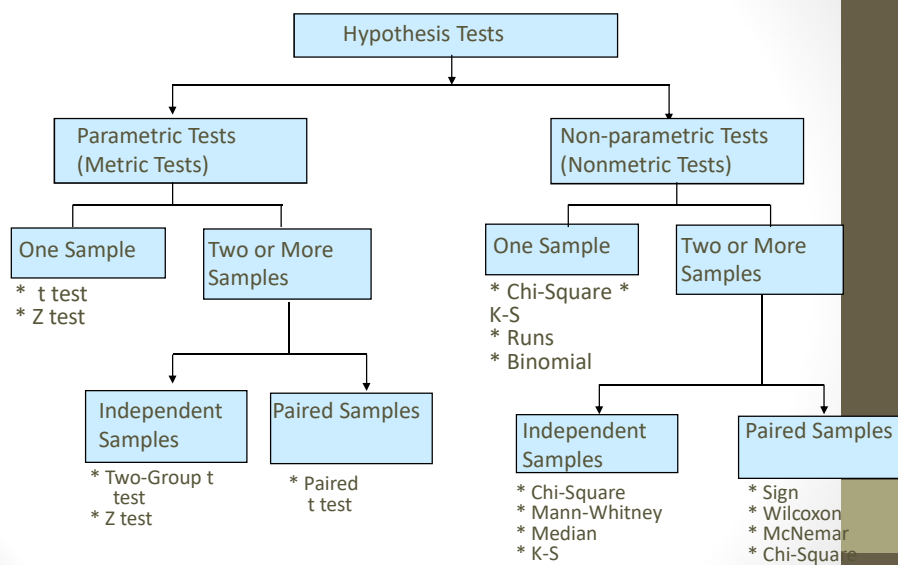
While conducting cross-tabulation analysis in practice, it is useful to proceed along the following steps.

1. Test the null hypothesis that there is no association between the variables using the chi-square statistic. If you fail to reject the null hypothesis, then there is no relationship.
2. If H_0 is rejected, then determine the strength of the association using an appropriate statistic (phi-coefficient, contingency coefficient, Cramer's V, lambda coefficient, or other statistics), as discussed earlier.
3. If H_0 is rejected, interpret the pattern of the relationship by computing the percentages in the direction of the independent variable, across the dependent variable.
4. If the variables are treated as ordinal rather than nominal, use tau b, tau c, or Gamma as the test statistic. If H_0 is rejected, then determine the strength of the association using the magnitude, and the direction of the relationship using the sign of the test statistic.

Hypothesis Testing Related to Differences

- **Parametric tests** assume that the variables of interest are measured on at least an interval scale.
- **Nonparametric tests** assume that the variables are measured on a nominal or ordinal scale.
- These tests can be further classified based on whether one or two or more samples are involved.
- The samples are **independent** if they are drawn randomly from different populations. For the purpose of analysis, data pertaining to different groups of respondents, e.g., males and females, are generally treated as independent samples.
- The samples are **paired** when the data for the two samples relate to the same group of respondents.

A Classification of Hypothesis Testing Procedures for Examining Differences



Parametric Tests

- The **t statistic** assumes that the variable is normally distributed and the mean is known (or assumed to be known) and the population variance is estimated from the sample.
- Assume that the random variable X is normally distributed, with mean and unknown population variance that is estimated by the sample variance s^2 .
- Then, $t = (\bar{X} - \mu) / \frac{s}{\sqrt{n}}$ is t distributed with $n - 1$ degrees of freedom.
- The **t distribution** is similar to the normal distribution in appearance. Both distributions are bell-shaped and symmetric. As the number of degrees of freedom increases, the t distribution approaches the normal distribution.

Hypothesis Testing Using the t Statistic

1. Formulate the null (H_0) and the alternative (H_1) hypotheses.
2. Select the appropriate formula for the t statistic.
3. Select a significance level, α , for testing H_0 . Typically, the 0.05 level is selected.
4. Take one or two samples and compute the mean and standard deviation for each sample.
5. Calculate the t statistic assuming H_0 is true.

Hypothesis Testing Using the t Statistic

6. Calculate the degrees of freedom and estimate the probability of getting a more extreme value of the statistic from Table 4. (Alternatively, calculate the critical value of the t statistic.)
7. If the probability computed in step 5 is smaller than the significance level selected in step 2, reject H_0 . If the probability is larger, do not reject H_0 . (Alternatively, if the absolute value of the calculated t statistic in step 4 is larger than the absolute critical value determined in step 5, reject H_0 .) Failure to reject H_0 does not necessarily imply that H_0 is true. It only means that the true state is not significantly different than that assumed by H_0 .
8. Express the conclusion reached by the t test in terms of the marketing research problem.

One Sample : t Test

For the data in Table 15.2, suppose we wanted to test the hypothesis that the mean familiarity rating exceeds 4.0, the neutral value on a 7-point scale. A significance level of $\alpha = 0.05$ is selected. The hypotheses may be formulated as:

$$H_0: \mu \leq 4.0$$

$$H_1: \mu > 4.0$$

$$t = (\bar{X} - \mu) / s_{\bar{X}}$$

$$s_{\bar{X}} = s / \sqrt{n}$$

$$s_{\bar{X}} = 1.579 / \sqrt{29}$$

$$= 1.579 / 5.385 = 0.293$$

$$t = (4.724 - 4.0) / 0.293 = 0.724 / 0.293 = 2.471$$

One Sample : t Test

The degrees of freedom for the t statistic to test the hypothesis about one mean are $n - 1$. In this case, $n - 1 = 29 - 1$ or 28. From Table 4 in the Statistical Appendix, the probability of getting a more extreme value than 2.471 is less than 0.05. (Alternatively, the critical t value for 28 degrees of freedom and a significance level of 0.05 is 1.7011, which is less than the calculated value.) Hence, the null hypothesis is rejected. The familiarity level does exceed 4.0.

One Sample : Z Test

Note that if the population standard deviation was assumed to be known as 1.5, rather than estimated from the sample, a **z test** would be appropriate. In this case, the value of the z statistic would be:

$$z = (\bar{X} - \mu) / \sigma_{\bar{X}}$$

where

$$\sigma_{\bar{X}} = 1.5 / \sqrt{29} = 1.5 / 5.385 = 0.279$$

and

$$z = (4.724 - 4.0) / 0.279 = 0.724 / 0.279 = 2.595$$

One Sample : Z Test

- From Table 2 in the Statistical Appendix, the probability of getting a more extreme value of z than 2.595 is less than 0.05. (Alternatively, the critical z value for a one-tailed test and a significance level of 0.05 is 1.645, which is less than the calculated value.) Therefore, the null hypothesis is rejected, reaching the same conclusion arrived at earlier by the t test.
- The procedure for testing a null hypothesis with respect to a proportion was illustrated earlier in this chapter when we introduced hypothesis testing.

Two Independent Samples Means

- In the case of means for two independent samples, the hypotheses take the following form.

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

- The two populations are sampled and the means and variances computed based on samples of sizes n_1 and n_2 . If both populations are found to have the same variance, a pooled variance estimate is computed from the two sample variances as follows:

$$s^2 = \frac{\sum_{i=1}^{n_1} (X_{i1} - \bar{X}_1)^2 + \sum_{i=1}^{n_2} (X_{i2} - \bar{X}_2)^2}{n_1 + n_2 - 2} \quad \text{or} \quad s^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}$$

Two Independent Samples Means

The standard deviation of the test statistic can be estimated as:

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

The appropriate value of t can be calculated as:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}}$$

The degrees of freedom in this case are $(n_1 + n_2 - 2)$.

Two Independent Samples F Test

An **F test** of sample variance may be performed if it is not known whether the two populations have equal variance. In this case, the hypotheses are:

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

Two Independent Samples F Statistic

The **F statistic** is computed from the sample variances as follows:

$$\text{where } F_{(n_1-1), (n_2-1)} = \frac{s_1^2}{s_2^2}$$

n_1 = size of sample 1

n_2 = size of sample 2

n_1-1 = degrees of freedom for sample 1

n_2-1 = degrees of freedom for sample 2

s_1^2 = sample variance for sample 1

s_2^2 = sample variance for sample 2

Using the data of Table 15.1, suppose we wanted to determine whether Internet usage was different for males as compared to females. A two-independent-samples *t* test was conducted. The results are presented in Table 15.14.

Two Independent-Samples *t* Tests





Summary Statistics					
	Number of Cases	Mean	Standard Deviation		
Male	15	9.333	1.137		
Female	15	3.867	0.435		
F Test for Equality of Variances					
	F value	2-tail probability			
	15.507	0.000			
t Test					
Equal Variances Assumed			Equal Variances Not Assumed		
t value	Degrees of freedom	2-tail probability	t value	Degrees of freedom	2-tail probability
-4.492	28	0.000	-4.492	18.014	0.000

Two Independent Samples Proportions

The case involving proportions for two independent samples is also illustrated using the data of Table 15.1, which gives the number of males and females who use the Internet for shopping. Is the proportion of respondents using the Internet for shopping the same for males and females?

The null and alternative hypotheses are:

$$H_0 : \pi_1 = \pi_2$$

$$H_1 : \pi_1 \neq \pi_2$$

A Z test is used as in testing the proportion for one sample. However, in this case the test statistic is given by:

$$Z = \frac{P_1 - P_2}{S_{P_1 - P_2}}$$

Two Independent Samples Proportions

In the test statistic, the numerator is the difference between the proportions in the two samples, P_1 and P_2 . The denominator is the standard error of the difference in the two proportions and is given by

$$S_{P_1 - P_2} = \sqrt{P(1-P) \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}$$

where

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

Two Independent Samples Proportions

A significance level of $\alpha = 0.05$ is selected. Given the data of Table 15.1, the test statistic can be calculated as:

$$P_1 - P_2 = (11/15) - (6/15)$$

$$= 0.733 - 0.400 = 0.333$$

$$P = (15 \times 0.733 + 15 \times 0.4) / (15 + 15) = 0.567$$

$$S_{P_1 - P_2} = \sqrt{0.567 \times 0.433 \left[\frac{1}{15} + \frac{1}{15} \right]} = 0.181$$

$$Z = 0.333 / 0.181 = 1.84$$

Two Independent Samples Proportions

Given a two-tail test, the area to the right of the critical value is 0.025. Hence, the critical value of the test statistic is 1.96. Since the calculated value is less than the critical value, the null hypothesis can not be rejected. Thus, the proportion of users (0.733 for males and 0.400 for females) is not significantly different for the two samples. Note that while the difference is substantial, it is not statistically significant due to the small sample sizes (15 in each group).

Paired Samples

The difference in these cases is examined by a **paired samples t test**. To compute t for paired samples, the paired difference variable, denoted by D , is formed and its mean and variance calculated. Then the t statistic is computed. The degrees of freedom are $n - 1$, where n is the number of pairs. The relevant formulas are:

$$H_0 : \mu_D = 0$$

$$H_1 : \mu_D \neq 0$$

$$t_{n-1} = \frac{\bar{D} - \mu_D}{\frac{s_D}{\sqrt{n}}}$$

continued...

Paired Samples

Where:

$$\bar{D} = \frac{\sum_{i=1}^n D_i}{n}$$

$$s_D = \sqrt{\frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n - 1}}$$

$$S_{\bar{D}} = \frac{s_D}{\sqrt{n}}$$

In the Internet usage example (Table 15.1), a paired t test could be used to determine if the respondents differed in their attitude toward the Internet and attitude toward technology. The resulting output is shown in Table 15.15.

Paired-Samples t Test

Table 15.15

Variable		Number of Cases	Mean	Standard Deviation	Standard Error		
Internet Attitude		30	5.167	1.234	0.225		
Technology Attitude		30	4.100	1.398	0.255		
Difference = Internet - Technology							
Difference Mean	Standard deviation	Standard error	Correlation	2-tail prob.	t value	Degrees of freedom	2-tail probability
1.067	0.828	0.1511	0.809	0.000	7.059	29	0.000



Nonparametric Tests

Nonparametric tests are used when the independent variables are nonmetric. Like parametric tests, nonparametric tests are available for testing variables from one sample, two independent samples, or two related samples.

Nonparametric Tests One Sample

- Sometimes the researcher wants to test whether the observations for a particular variable could reasonably have come from a particular distribution, such as the normal, uniform, or Poisson distribution.
- The **Kolmogorov-Smirnov (K-S) one-sample test** is one such goodness-of-fit test. The K-S compares the cumulative distribution function for a variable with a specified distribution. A_i denotes the cumulative relative frequency for each category of the theoretical (assumed) distribution, and O_i the comparable value of the sample frequency. The K-S test is based on the maximum value of the absolute difference between A_i and O_i . The test statistic is

$$K = \text{Max} |A_i - O_i|$$

Nonparametric Tests One Sample

- The decision to reject the null hypothesis is based on the value of K . The larger the K is, the more confidence we have that H_0 is false. For $\alpha = 0.05$, the critical value of K for large samples (over 35) is given by $1.36/\sqrt{n}$. Alternatively, K can be transformed into a normally distributed z statistic and its associated probability determined.
- In the context of the Internet usage example, suppose we wanted to test whether the distribution of Internet usage was normal. A K-S one-sample test is conducted, yielding the data shown in Table 15.16. Table 15.16 indicates that the probability of observing a K value of 0.222, as determined by the normalized z statistic, is 0.103. Since this is more than the significance level of 0.05, the null hypothesis can not be rejected, leading to the same conclusion. Hence, the distribution of Internet usage does not deviate significantly from the normal distribution.

K-S One-Sample Test for Normality of Internet Usage

Test Distribution - Normal				
Mean:		6.600		
Standard Deviation:		4.296		
Cases:		30		
Most Extreme Differences				
Absolute	Positive	Negative	K-S z	2-Tailed p
0.222	0.222	-0.142	1.217	0.103



Nonparametric Tests One Sample

- The **chi-square test** can also be performed on a single variable from one sample. In this context, the chi-square serves as a goodness-of-fit test.
- The **runs test** is a test of randomness for the dichotomous variables. This test is conducted by determining whether the order or sequence in which observations are obtained is random.
- The **binomial test** is also a goodness-of-fit test for dichotomous variables. It tests the goodness of fit of the observed number of observations in each category to the number expected under a specified binomial distribution.

Nonparametric Tests Two Independent Samples

- When the difference in the location of two populations is to be compared based on observations from two independent samples, and the variable is measured on an ordinal scale, the **Mann-Whitney U test** can be used.
- In the Mann-Whitney U test, the two samples are combined and the cases are ranked in order of increasing size.
- The test statistic, U , is computed as the number of times a score from sample or group 1 precedes a score from group 2.
- If the samples are from the same population, the distribution of scores from the two groups in the rank list should be random. An extreme value of U would indicate a nonrandom pattern, pointing to the inequality of the two groups.
- For samples of less than 30, the exact significance level for U is computed. For larger samples, U is transformed into a normally distributed z statistic. This z can be corrected for ties within ranks.

Nonparametric Tests Two Independent Samples

- We examine again the difference in the Internet usage of males and females. This time, though, the Mann-Whitney U test is used. The results are given in Table 15.17.
- One could also use the cross-tabulation procedure to conduct a chi-square test. In this case, we will have a 2×2 table. One variable will be used to denote the sample, and will assume the value 1 for sample 1 and the value of 2 for sample 2. The other variable will be the binary variable of interest.
- The **two-sample median test** determines whether the two groups are drawn from populations with the same median. It is not as powerful as the Mann-Whitney U test because it merely uses the location of each observation relative to the median, and not the rank, of each observation.
- The **Kolmogorov-Smirnov two-sample test** examines whether the two distributions are the same. It takes into account any differences between the two distributions, including the median, dispersion, and skewness.

Mann-Whitney U - Wilcoxon Rank Sum W Test Internet Usage by Gender

Table 15.17

Sex	Mean Rank	Cases
Male	20.93	15
Female	10.07	15
Total		30
		Corrected for ties
		2-tailed p
U	W	z
31.000	151.000	-3.406
		0.001

Note

U = Mann-Whitney test statistic
 W = Wilcoxon W Statistic
 z = U transformed into normally distributed z statistic



Nonparametric Tests Paired Samples

- The **Wilcoxon matched-pairs signed-ranks test** analyzes the differences between the paired observations, taking into account the magnitude of the differences.
- It computes the differences between the pairs of variables and ranks the absolute differences.
- The next step is to sum the positive and negative ranks. The test statistic, z , is computed from the positive and negative rank sums.
- Under the null hypothesis of no difference, z is a standard normal variate with mean 0 and variance 1 for large samples.

Nonparametric Tests Paired Samples

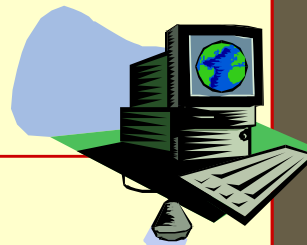
- The example considered for the paired t test, whether the respondents differed in terms of attitude toward the Internet and attitude toward technology, is considered again. Suppose we assume that both these variables are measured on ordinal rather than interval scales. Accordingly, we use the Wilcoxon test. The results are shown in Table 15.18.
- The **sign test** is not as powerful as the Wilcoxon matched-pairs signed-ranks test as it only compares the signs of the differences between pairs of variables without taking into account the ranks.
- In the special case of a binary variable where the researcher wishes to test differences in proportions, the McNemar test can be used. Alternatively, the chi-square test can also be used for binary variables.

Wilcoxon Matched-Pairs Signed-Rank Test Internet with Technology

Table 15.18

(Technology - Internet)	Cases	Mean rank
-Ranks	23	12.72
+Ranks	1	7.50
Ties	6	
Total	30	

$z = -4.207$	2-tailed $p = 0.0000$
--------------	-----------------------



A Summary of Hypothesis Tests Related to Differences

Table 15.19



Sample	Application	Level of Scaling	Test/Comments
One Sample	Proportion	Metric	Z test
One Sample	Distributions	Nonmetric	K-S and chi-square for goodness of fit Runs test for randomness Binomial test for goodness of fit for dichotomous variables
One Sample	Means	Metric	t test, if variance is unknown z test, if variance is known

A Summary of Hypothesis Tests Related to Differences

Table 15.19, cont.

Two Independent Samples			
Two independent samples	Distributions	Nonmetric	K-S two-sample test for examining the equivalence of two distributions
Two independent samples	Means	Metric	Two-group t test F test for equality of variances
Two independent samples	Proportions	Metric	z test
		Nonmetric	Chi-square test
Two independent samples	Rankings/Medians	Nonmetric	Mann-Whitney U test is more powerful than the median test

A Summary of Hypothesis Tests Related to Differences

Table 15.19, cont.

Paired Samples

Paired samples	Means	Metric	Paired t test
Paired samples	Proportions	Nonmetric	McNemar test for binary variables Chi-square test
Paired samples	Rankings/Medians	Nonmetric	Wilcoxon matched-pair: ranked-signs test is more powerful than the sign test

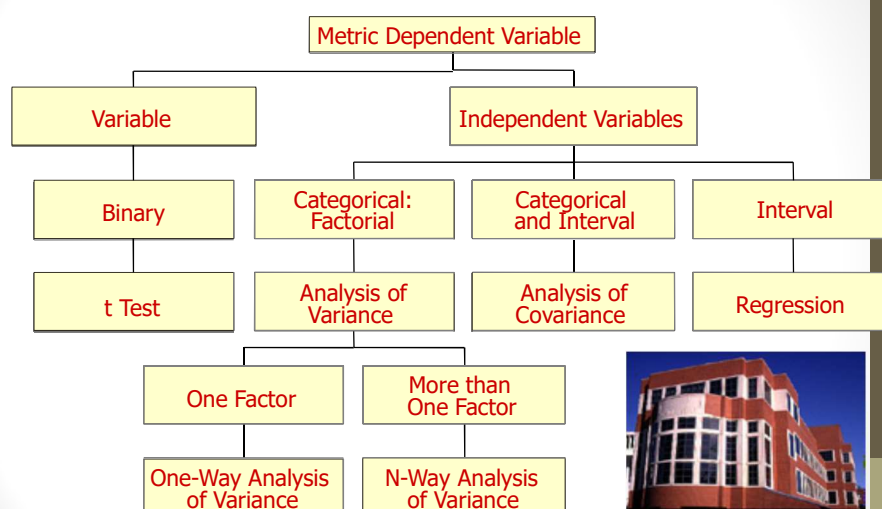
Analysis of Variance (ANOVA)

- **Analysis of variance (ANOVA)** is used as a test of means for two or more populations. The null hypothesis, typically, is that all means are equal.
- Analysis of variance must have a dependent variable that is metric (measured using an interval or ratio scale).
- There must also be one or more independent variables that are all categorical (nonmetric). Categorical independent variables are also called **factors**.

Relationship Among Techniques

- A particular combination of factor levels, or categories, is called a **treatment**.
- **One-way analysis of variance** involves only one categorical variable, or a single factor. In one-way analysis of variance, a treatment is the same as a factor level.
- If two or more factors are involved, the analysis is termed **n-way analysis of variance**.
- If the set of independent variables consists of both categorical and metric variables, the technique is called **analysis of covariance (ANCOVA)**. In this case, the categorical independent variables are still referred to as factors, whereas the metric-independent variables are referred to as **covariates**.

Relationship Amongst Test, Analysis of Variance, Analysis of Covariance, & Regression



One-Way Analysis of Variance

Researchers are often interested in examining the differences in the mean values of the dependent variable for several categories of a single independent variable or factor. For example:

- Do the various segments differ in terms of their volume of product consumption?
- Do the brand evaluations of groups exposed to different commercials vary?
- What is the effect of consumers' familiarity with the store (measured as high, medium, and low) on preference for the store?

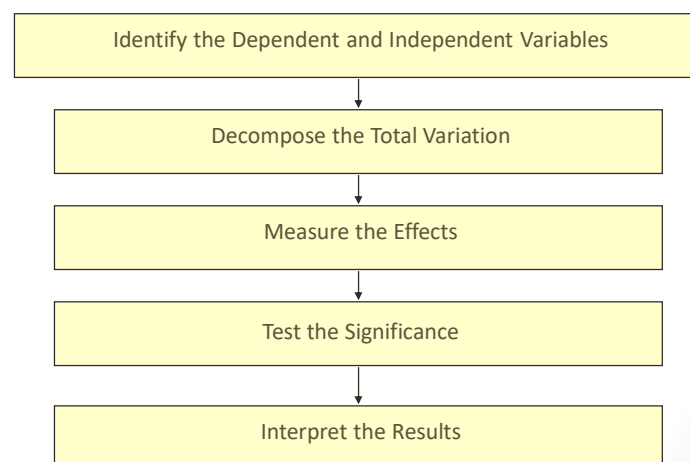
Statistics Associated with One-Way Analysis of Variance

- **η^2 (η^2)**. The strength of the effects of X (independent variable or factor) on Y (dependent variable) is measured by **η^2 (η^2)**. The value of η^2 varies between 0 and 1.
- **F statistic**. The null hypothesis that the category means are equal in the population is tested by an **F statistic** based on the ratio of mean square related to X and mean square related to error.
- **Mean square**. This is the sum of squares divided by the appropriate degrees of freedom.

Statistics Associated with One-Way Analysis of Variance

- SS_{between} . Also denoted as SS_X , this is the variation in Y related to the variation in the means of the categories of X . This represents variation between the categories of X , or the portion of the sum of squares in Y related to X .
- SS_{within} . Also referred to as SS_{error} , this is the variation in Y due to the variation within each of the categories of X . This variation is not accounted for by X .
- SS_Y . This is the total variation in Y .

Conducting One-Way ANOVA



Conducting One-Way Analysis of Variance

Decompose the Total Variation

The total variation in Y , denoted by SS_y , can be decomposed into two components:

$$SS_y = SS_{\text{between}} + SS_{\text{within}}$$

where the subscripts *between* and *within* refer to the categories of X .

SS_{between} is the variation in Y related to the variation in the means of the categories of X . For this reason, SS_{between} is also denoted as SS_x .

SS_{within} is the variation in Y related to the variation within each category of X . SS_{within} is not accounted for by X . Therefore it is referred to as SS_{error} .

Conducting One-Way Analysis of Variance

Decompose the Total Variation

The total variation in Y may be decomposed as:

$$SS_y = SS_x + SS_{\text{error}}$$

Where $SS_y = \sum_{i=1}^N (Y_i - \bar{Y})^2$

$$SS_x = \sum_{j=1}^c n (\bar{Y}_j - \bar{Y})^2$$

$$SS_{\text{error}} = \sum_{j=1}^c \sum_{i=1}^n (Y_{ij} - \bar{Y}_j)^2$$

Y_i = individual observation

$\bar{Y}_j$ = mean for category j

$\bar{Y}$ = mean over the whole sample, or grand mean

Y_{ij} = i th observation in the j th category

Decomposition of the Total Variation: One-Way ANOVA

Independent Variable					X
Categories					Total Sample
X_1	X_2	X_3	...	X_c	
Y_1	Y_1	Y_1		Y_1	Y_1
Y_2	Y_2	Y_2		Y_2	Y_2
$\vdots$					$\vdots$
Y_n	Y_n	Y_n		Y_n	Y_N
$\bar{Y}_1$	$\bar{Y}_2$	$\bar{Y}_3$		$\bar{Y}_c$	$\bar{Y}$

Within Category Variation = SS_{within}

Category Mean

Between Category Variation = SS_{between}

Total Variation = SS_y

Conducting One-Way Analysis of Variance

In analysis of variance, we estimate two measures of variation: within groups (SS_{within}) and between groups (SS_{between}). Thus, by comparing the Y variance estimates based on between-group and within-group variation, we can test the null hypothesis.

Measure the Effects

The strength of the effects of X on Y are measured as follows:

$$\eta^2 = SS_x / SS_y = (SS_y - SS_{\text{error}}) / SS_y$$

The value of η^2 varies between 0 and 1.

Conducting One-Way Analysis of Variance Test of Significance

In one-way analysis of variance, the interest lies in testing the null hypothesis that the category means are equal in the population.

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_c$$

Under the null hypothesis, SS_x and SS_{error} come from the same source of variation. In other words, the estimate of the population variance of Y ,

$$\begin{aligned} &= SS_x / (c - 1) \\ &= \text{Mean square due to } X \\ S_y^2 &= MS_x \end{aligned}$$

or

$$\begin{aligned} &= SS_{error} / (N - c) \\ S_y^2 &= \text{Mean square due to error} \\ S_y^2 &= MS_{error} \end{aligned}$$

Conducting One-Way Analysis of Variance Test of Significance

The null hypothesis may be tested by the **F statistic** based on the ratio between these two estimates:

$$F = \frac{SS_x / (c - 1)}{SS_{error} / (N - c)} = \frac{MS_x}{MS_{error}}$$

This statistic follows the **F distribution**, with $(c - 1)$ and $(N - c)$ degrees of freedom (df).

Conducting One-Way Analysis of Variance Interpret the Results

- If the null hypothesis of equal category means is not rejected, then the independent variable does not have a significant effect on the dependent variable.
- On the other hand, if the null hypothesis is rejected, then the effect of the independent variable is significant.
- A comparison of the category mean values will indicate the nature of the effect of the independent variable.

Illustrative Applications of One-Way Analysis of Variance

We illustrate the concepts discussed in this chapter using the data presented in Table 16.2.

The department store is attempting to determine the effect of in-store promotion (X) on sales (Y). For the purpose of illustrating hand calculations, the data of Table 16.2 are transformed in Table 16.3 to show the store sales (Y_{ij}) for each level of promotion.

The null hypothesis is that the category means are equal:

$$H_0: \mu_1 = \mu_2 = \mu_3$$

Effect of Promotion and Clientele on Sales

Table 16.2

Store Number	Coupon Level	In-Store Promotion	Sales	Clientele Rating
1	1.00	1.00	10.00	9.00
2	1.00	1.00	9.00	10.00
3	1.00	1.00	10.00	8.00
4	1.00	1.00	8.00	4.00
5	1.00	1.00	9.00	6.00
6	1.00	2.00	8.00	8.00
7	1.00	2.00	8.00	4.00
8	1.00	2.00	7.00	10.00
9	1.00	2.00	9.00	6.00
10	1.00	2.00	6.00	9.00
11	1.00	3.00	5.00	8.00
12	1.00	3.00	7.00	9.00
13	1.00	3.00	6.00	6.00
14	1.00	3.00	4.00	10.00
15	1.00	3.00	5.00	4.00
16	2.00	1.00	8.00	10.00
17	2.00	1.00	9.00	6.00
18	2.00	1.00	7.00	8.00
19	2.00	1.00	7.00	4.00
20	2.00	1.00	6.00	9.00
21	2.00	2.00	4.00	6.00
22	2.00	2.00	5.00	8.00
23	2.00	2.00	5.00	10.00
24	2.00	2.00	6.00	4.00
25	2.00	2.00	4.00	9.00
26	2.00	3.00	2.00	4.00
27	2.00	3.00	3.00	6.00
28	2.00	3.00	2.00	10.00
29	2.00	3.00	1.00	9.00
30	2.00	3.00	2.00	8.00



Illustrative Applications of One-Way Analysis of Variance

Table 16.3

EFFECT OF IN-STORE PROMOTION ON SALES

Store No.	Level of In-store Promotion		
	High	Medium	Low
	Normalized Sales		
1	10	8	5
2	9	8	7
3	10	7	6
4	8	9	4
5	9	6	5
6	8	4	2
7	9	5	3
8	7	5	2
9	7	6	1
10	6	4	2
Column Totals	83	62	37
Category means: $\overline{y}_j$	83/10 = 8.3	62/10 = 6.2	37/10 = 3.7
Grand mean, $\overline{Y}$	= (83 + 62 + 37)/30 = 6.067		

Illustrative Applications of One-Way Analysis of Variance

To test the null hypothesis, the various sums of squares are computed as follows:

$$\begin{aligned}
 SS_y &= (10-6.067)^2 + (9-6.067)^2 + (10-6.067)^2 + (8-6.067)^2 + (9-6.067)^2 \\
 &\quad + (8-6.067)^2 + (9-6.067)^2 + (7-6.067)^2 + (7-6.067)^2 + (6-6.067)^2 \\
 &\quad + (8-6.067)^2 + (8-6.067)^2 + (7-6.067)^2 + (9-6.067)^2 + (6-6.067)^2 \\
 &\quad + (4-6.067)^2 + (5-6.067)^2 + (5-6.067)^2 + (6-6.067)^2 + (4-6.067)^2 \\
 &\quad + (5-6.067)^2 + (7-6.067)^2 + (6-6.067)^2 + (4-6.067)^2 + (5-6.067)^2 \\
 &\quad + (2-6.067)^2 + (3-6.067)^2 + (2-6.067)^2 + (1-6.067)^2 + (2-6.067)^2 \\
 &= (3.933)^2 + (2.933)^2 + (3.933)^2 + (1.933)^2 + (2.933)^2 \\
 &\quad + (1.933)^2 + (2.933)^2 + (0.933)^2 + (0.933)^2 + (-0.067)^2 \\
 &\quad + (1.933)^2 + (1.933)^2 + (0.933)^2 + (2.933)^2 + (-0.067)^2 \\
 &\quad + (-2.067)^2 + (-1.067)^2 + (-1.067)^2 + (-0.067)^2 + (-2.067)^2 \\
 &\quad + (-1.067)^2 + (0.9333)^2 + (-0.067)^2 + (-2.067)^2 + (-1.067)^2 \\
 &\quad + (-4.067)^2 + (-3.067)^2 + (-4.067)^2 + (-5.067)^2 + (-4.067)^2 \\
 &= 185.867
 \end{aligned}$$

Illustrative Applications of One-Way Analysis of Variance

$$\begin{aligned}
 SS_x &= 10(8.3-6.067)^2 + 10(6.2-6.067)^2 + 10(3.7-6.067)^2 \\
 &= 10(2.233)^2 + 10(0.133)^2 + 10(-2.367)^2 \\
 &= 106.067
 \end{aligned}$$

$$\begin{aligned}
 SS_{error} &= (10-8.3)^2 + (9-8.3)^2 + (10-8.3)^2 + (8-8.3)^2 + (9-8.3)^2 \\
 &\quad + (8-8.3)^2 + (9-8.3)^2 + (7-8.3)^2 + (7-8.3)^2 + (6-8.3)^2 \\
 &\quad + (8-6.2)^2 + (8-6.2)^2 + (7-6.2)^2 + (9-6.2)^2 + (6-6.2)^2 \\
 &\quad + (4-6.2)^2 + (5-6.2)^2 + (5-6.2)^2 + (6-6.2)^2 + (4-6.2)^2 \\
 &\quad + (5-3.7)^2 + (7-3.7)^2 + (6-3.7)^2 + (4-3.7)^2 + (5-3.7)^2 \\
 &\quad + (2-3.7)^2 + (3-3.7)^2 + (2-3.7)^2 + (1-3.7)^2 + (2-3.7)^2 \\
 &= (1.7)^2 + (0.7)^2 + (1.7)^2 + (-0.3)^2 + (0.7)^2 \\
 &\quad + (-0.3)^2 + (0.7)^2 + (-1.3)^2 + (-1.3)^2 + (-2.3)^2 \\
 &\quad + (1.8)^2 + (1.8)^2 + (0.8)^2 + (2.8)^2 + (-0.2)^2 \\
 &\quad + (-2.2)^2 + (-1.2)^2 + (-1.2)^2 + (-0.2)^2 + (-2.2)^2 \\
 &\quad + (1.3)^2 + (3.3)^2 + (2.3)^2 + (0.3)^2 + (1.3)^2 \\
 &\quad + (-1.7)^2 + (-0.7)^2 + (-1.7)^2 + (-2.7)^2 + (-1.7)^2 \\
 &= 79.80
 \end{aligned}$$

Illustrative Applications of One-Way Analysis of Variance

It can be verified that

$$SS_y = SS_x + SS_{error}$$

as follows:

$$185.867 = 106.067 + 79.80$$

The strength of the effects of X on Y are measured as follows:

$$\begin{aligned} \eta^2 &= SS_x / SS_y \\ &= 106.067 / 185.867 \\ &= 0.571 \end{aligned}$$

In other words, 57.1% of the variation in sales (Y) is accounted for by in-store promotion (X), indicating a modest effect. The null hypothesis may now be tested.

$$\begin{aligned} F &= \frac{SS_x / (c - 1)}{SS_{error} / (N - c)} = \frac{MS_x}{MS_{error}} \\ F &= \frac{106.067 / (3 - 1)}{79.800 / (30 - 3)} \\ &= 17.944 \end{aligned}$$

Illustrative Applications of One-Way Analysis of Variance

- From Table 5 in the Statistical Appendix we see that for 2 and 27 degrees of freedom, the critical value of F is 3.35 for $\alpha = 0.05$. Because the calculated value of F is greater than the critical value, we reject the null hypothesis.
- We now illustrate the analysis of variance procedure using a computer program. The results of conducting the same analysis by computer are presented in Table 16.4.

One-Way ANOVA: Effect of In-Store Promotion on Store Sales

Table 16.4

Source of prob.	Sum of	df	Mean	F ratio	F
Variation	squares		square		
Between groups (Promotion)	106.067	2	53.033	17.944	0.000
Within groups (Error)	79.800	27	2.956		
TOTAL	185.867	29	6.409		
Cell means					
Level of Promotion	Count	Mean			
High (1)	10	8.300			
Medium (2)	10	6.200			
Low (3)	10	3.700			
TOTAL	30	6.067			



Assumptions in Analysis of Variance

The salient assumptions in analysis of variance can be summarized as follows:

1. Ordinarily, the categories of the independent variable are assumed to be fixed. Inferences are made only to the specific categories considered. This is referred to as the **fixed-effects model**.
2. The error term is **normally** distributed, with a **zero mean** and a **constant variance**. The error is not related to any of the categories of X .
3. The error terms are **uncorrelated**. If the error terms are correlated (i.e., the observations are not independent), the F ratio can be seriously distorted.

N-Way Analysis of Variance

In marketing research, one is often concerned with the effect of more than one factor simultaneously. For example:

- How do advertising levels (high, medium, and low) interact with price levels (high, medium, and low) to influence a brand's sale?
- Do educational levels (less than high school, high school graduate, some college, and college graduate) and age (less than 35, 35-55, more than 55) affect consumption of a brand?
- What is the effect of consumers' familiarity with a department store (high, medium, and low) and store image (positive, neutral, and negative) on preference for the store?

N-Way Analysis of Variance

Consider the simple case of two factors X_1 and X_2 having categories c_1 and c_2 . The total variation in this case is partitioned as follows:

$$SS_{total} = SS \text{ due to } X_1 + SS \text{ due to } X_2 + SS \text{ due to interaction of } X_1 \text{ and } X_2 + SS_{within}$$

or

$$SS_y = SS_{x_1} + SS_{x_2} + SS_{x_1x_2} + SS_{error}$$

The strength of the joint effect of two factors, called the overall effect, or **multiple η^2** , is measured as follows:

$$\text{multiple } \eta^2 = (SS_{x_1} + SS_{x_2} + SS_{x_1x_2}) / SS_y$$

N-Way Analysis of Variance

The **significance of the overall effect** may be tested by an F test, as follows:

$$\begin{aligned} F &= \frac{(SS_{x_1} + SS_{x_2} + SS_{x_1x_2})/df_n}{SS_{error}/df_d} \\ &= \frac{SS_{x_1x_2x_1x_2}/df_n}{SS_{error}/df_d} \\ &= \frac{MS_{x_1x_2x_1x_2}}{MS_{error}} \end{aligned}$$

where

df_n	=	degrees of freedom for the numerator
	=	$(c_1 - 1) + (c_2 - 1) + (c_1 - 1)(c_2 - 1)$
	=	$c_1c_2 - 1$
df_d	=	degrees of freedom for the denominator
	=	$N - c_1c_2$
MS	=	mean square

N-Way Analysis of Variance

If the overall effect is significant, the next step is to examine the **significance of the interaction effect**. Under the null hypothesis of no interaction, the appropriate F test is:

$$\begin{aligned} F &= \frac{SS_{x_1x_2}/df_n}{SS_{error}/df_d} \\ &= \frac{MS_{x_1x_2}}{MS_{error}} \end{aligned}$$

Where

$$\begin{aligned} df_n &= (c_1 - 1)(c_2 - 1) \\ df_d &= N - c_1c_2 \end{aligned}$$

N-Way Analysis of Variance

The **significance of the main effect of each factor** may be tested as follows for X_1 :

$$F = \frac{SS_{x_1}/df_n}{SS_{error}/df_d}$$

$$= \frac{MS_{x_1}}{MS_{error}}$$

where

$$df_n = c_1 - 1$$

$$df_d = N - c_1c_2$$

Two-Way Analysis of Variance

Table 16.5

Source of Variation	Sum of squares	df	Mean square	F	Sig. of F	ω^2
Main Effects						
Promotion	106.067	2	53.033	54.862	0.000	0.557
Coupon	53.333	1	53.333	55.172	0.000	0.280
Combined	159.400	3	53.133	54.966	0.000	
Two-way interaction	3.267	2	1.633	1.690	0.226	
Model	162.667	5	32.533	33.655	0.000	
Residual (error)	23.200	24	0.967			
TOTAL	185.867	29	6.409			

Two-Way Analysis of Variance

Table 16.5, cont.

Cell Means				
Promotion	Coupon	Count	Mean	
High	Yes	5	9.200	
High	No	5	7.400	
Medium	Yes	5	7.600	
Medium	No	5	4.800	
Low	Yes	5	5.400	
Low	No	5	2.000	
TOTAL		30		
Factor Level Means				
Promotion	Coupon	Count	Mean	
High		10	8.300	
Medium		10	6.200	
Low		10	3.700	
	Yes	15	7.400	
	No	15	4.733	
Grand Mean		30	6.067	

Analysis of Covariance

When examining the differences in the mean values of the dependent variable related to the effect of the controlled independent variables, it is often necessary to take into account the influence of uncontrolled independent variables. For example:

- In determining how different groups exposed to different commercials evaluate a brand, it may be necessary to control for prior knowledge.
- In determining how different price levels will affect a household's cereal consumption, it may be essential to take household size into account. We again use the data of Table 16.2 to illustrate analysis of covariance.
- Suppose that we wanted to determine the effect of in-store promotion and couponing on sales while controlling for the effect of clientele. The results are shown in Table 16.6.

Analysis of Covariance

Table 16.6

Source of Variation	Sum of Squares	df	Mean Square	F	Sig. of F
Covariance					
Cientele	0.838	1	0.838	0.862	0.363
Main effects					
Promotion	106.067	2	53.033	54.546	0.000
Coupon	53.333	1	53.333	54.855	0.000
Combined	159.400	3	53.133	54.649	0.000
2-Way Interaction					
Promotion* Coupon	3.267	2	1.633	1.680	0.208
Model	163.505	6	27.251	28.028	0.000
Residual (Error)	22.362	23	0.972		
TOTAL	185.867	29	6.409		
Covariate	Raw Coefficient				
Cientele	-0.078				

Issues in Interpretation

Important issues involved in the interpretation of ANOVA results include interactions, relative importance of factors, and multiple comparisons.

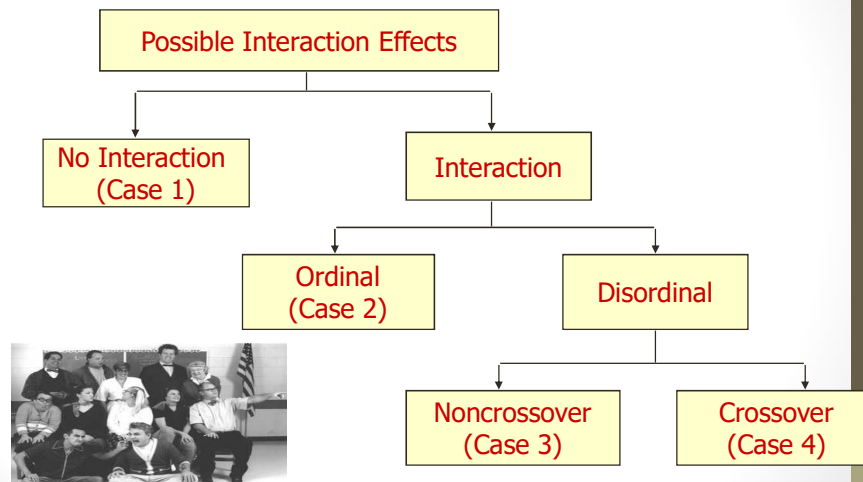
Interactions

- The different interactions that can arise when conducting ANOVA on two or more factors are shown in Figure 16.3.

Relative Importance of Factors

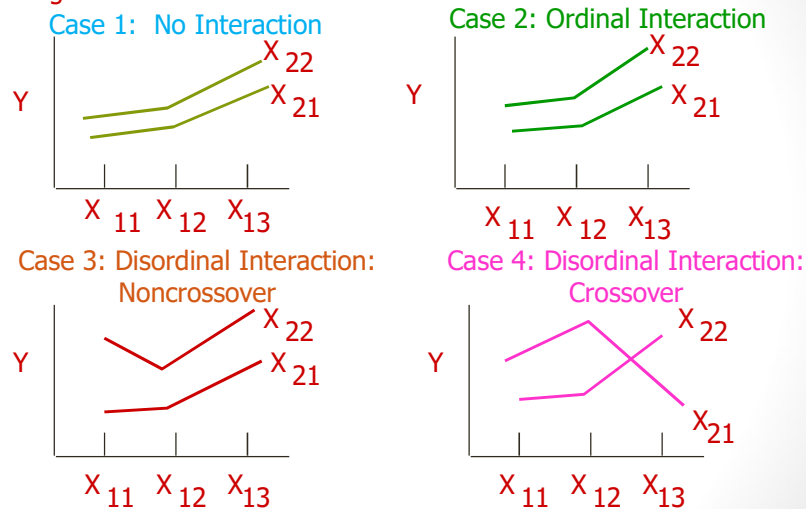
- Experimental designs are usually balanced, in that each cell contains the same number of respondents. This results in an orthogonal design in which the factors are uncorrelated. Hence, it is possible to determine unambiguously the relative importance of each factor in explaining the variation in the dependent variable.

A Classification of Interaction Effects



Patterns of Interaction

Fig. 16.4



Issues in Interpretation

- The most commonly used measure in ANOVA is **omega squared, ω^2** . This measure indicates what proportion of the variation in the dependent variable is related to a particular independent variable or factor. The relative contribution of a factor X is calculated as follows:

$$\omega_x^2 = \frac{SS_x - (df_x \times MS_{error})}{SS_{total} + MS_{error}}$$

- Normally, ω^2 is interpreted only for statistically significant effects. In Table 16.5, ω^2 associated with the level of in-store promotion is calculated as follows:

$$\begin{aligned}\omega_p^2 &= \frac{106.067 - (2 \times 0.967)}{185.867 + 0.967} \\ &= \frac{104.133}{186.834} \\ &= 0.557\end{aligned}$$

Issues in Interpretation

Note, in Table 16.5, that

$$\begin{aligned}SS_{total} &= 106.067 + 53.333 + 3.267 + 23.2 \\ &= 185.867\end{aligned}$$

Likewise, the ω^2 associated with couponing is:

$$\begin{aligned}\omega_c^2 &= \frac{53.333 - (1 \times 0.967)}{185.867 + 0.967} \\ &= \frac{52.366}{186.834} \\ &= 0.280\end{aligned}$$

As a guide to interpreting, a large experimental effect produces an index of 0.15 or greater, a medium effect produces an index of around 0.06, and a small effect produces an index of 0.01. In Table 16.5, while the effect of promotion and couponing are both large, the effect of promotion is much larger.

Issues in Interpretation - Multiple comparisons

- If the null hypothesis of equal means is rejected, we can only conclude that not all of the group means are equal. We may wish to examine differences among specific means. This can be done by specifying appropriate **contrasts**, or comparisons used to determine which of the means are statistically different.
- **A priori contrasts** are determined before conducting the analysis, based on the researcher's theoretical framework. Generally, a priori contrasts are used in lieu of the ANOVA F test. The contrasts selected are orthogonal (they are independent in a statistical sense).

Issues in Interpretation - Multiple Comparisons

- **A posteriori contrasts** are made after the analysis. These are generally **multiple comparison tests**. They enable the researcher to construct generalized confidence intervals that can be used to make pairwise comparisons of all treatment means. These tests, listed in order of decreasing power, include least significant difference, Duncan's multiple range test, Student-Newman-Keuls, Tukey's alternate procedure, honestly significant difference, modified least significant difference, and Scheffe's test. Of these tests, least significant difference is the most powerful, Scheffe's the most conservative.

Repeated Measures ANOVA

One way of controlling the differences between subjects is by observing each subject under each experimental condition (see Table 16.7). Since repeated measurements are obtained from each respondent, this design is referred to as within-subjects design or **repeated measures analysis of variance**. Repeated measures analysis of variance may be thought of as an extension of the paired-samples t test to the case of more than two related samples.

Decomposition of the Total Variation: Repeated Measures ANOVA

Table 16.7

		Independent Variable					X	Total Sample
		Subject No.	Categories					
			X_1	X_2	X_3	...	X_c	
Between People Variation =SS _{between people}	1	Y_{11}	Y_{12}	Y_{13}		Y_{1c}	Y_1	Total Variation =SS _y
	2	Y_{21}	Y_{22}	Y_{23}		Y_{2c}	Y_2	
		$\vdots$					$\vdots$	
	n	Y_{n1}	Y_{n2}	Y_{n3}		Y_{nc}	Y_N	
Category Mean			Y_1	Y_2	Y_3		Y_c	Y
$\downarrow$ Within People Category Variation = SS_{within people}								

Repeated Measures ANOVA

In the case of a single factor with repeated measures, the total variation, with $nc - 1$ degrees of freedom, may be split into between-people variation and within-people variation.

$$SS_{total} = SS_{between\ people} + SS_{within\ people}$$

The between-people variation, which is related to the differences between the means of people, has $n - 1$ degrees of freedom. The within-people variation has $n(c - 1)$ degrees of freedom. The within-people variation may, in turn, be divided into two different sources of variation. One source is related to the differences between treatment means, and the second consists of residual or error variation. The degrees of freedom corresponding to the treatment variation are $c - 1$, and those corresponding to residual variation are $(c - 1)(n - 1)$.

Repeated Measures ANOVA

Thus,

$$SS_{within\ people} = SS_x + SS_{error}$$

A test of the null hypothesis of equal means may now be constructed in the usual way:

$$F = \frac{SS_x / (c - 1)}{SS_{error} / (n - 1)(c - 1)} = \frac{MS_x}{MS_{error}}$$

So far we have assumed that the dependent variable is measured on an interval or ratio scale. If the dependent variable is nonmetric, however, a different procedure should be used.

Nonmetric Analysis of Variance

- **Nonmetric analysis of variance** examines the difference in the central tendencies of more than two groups when the dependent variable is measured on an ordinal scale.
- One such procedure is the ***k*-sample median test**. As its name implies, this is an extension of the median test for two groups, which was considered in Chapter 15.

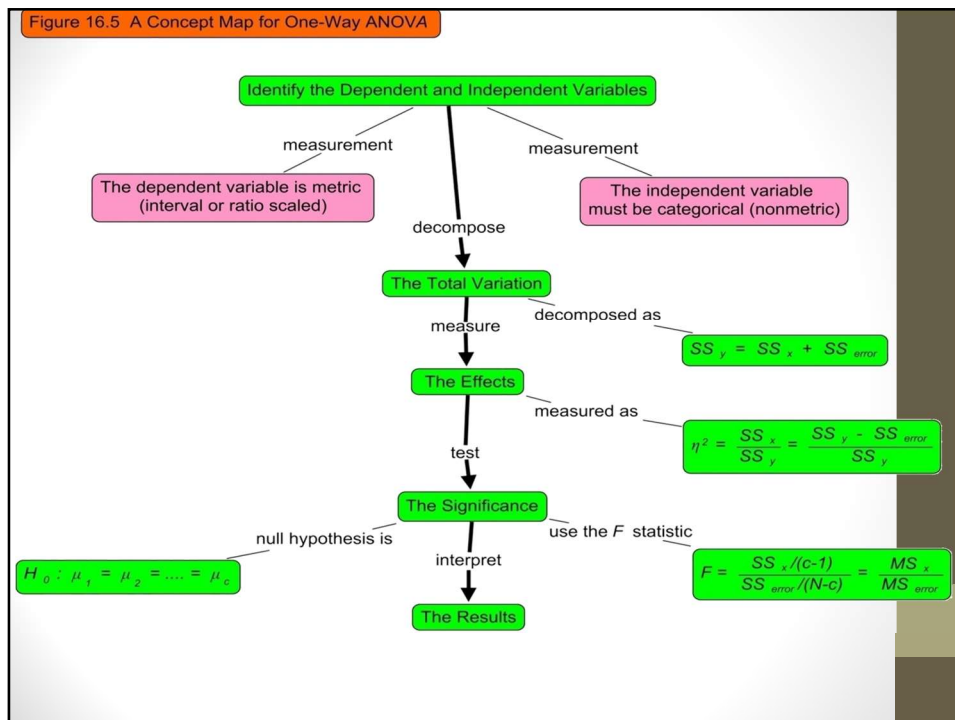
Nonmetric Analysis of Variance

- A more powerful test is the **Kruskal-Wallis one-way analysis of variance**. This is an extension of the Mann-Whitney test (Chapter 15). This test also examines the difference in medians. All cases from the *k* groups are ordered in a single ranking. If the *k* populations are the same, the groups should be similar in terms of ranks within each group. The rank sum is calculated for each group. From these, the Kruskal-Wallis *H* statistic, which has a chi-square distribution, is computed.
- The Kruskal-Wallis test is more powerful than the *k*-sample median test as it uses the rank value of each case, not merely its location relative to the median. However, if there are a large number of tied rankings in the data, the *k*-sample median test may be a better choice.

Multivariate Analysis of Variance

- **Multivariate analysis of variance (MANOVA)** is similar to analysis of variance (ANOVA), except that instead of one metric dependent variable, we have two or more.
- In MANOVA, the null hypothesis is that the vectors of means on multiple dependent variables are equal across groups.
- Multivariate analysis of variance is appropriate when there are two or more dependent variables that are correlated.

Figure 16.5 A Concept Map for One-Way ANOVA



Product Moment Correlation

- The **product moment correlation**, r , summarizes the strength of association between two metric (interval or ratio scaled) variables, say X and Y .
- It is an index used to determine whether a linear or straight-line relationship exists between X and Y .
- As it was originally proposed by Karl Pearson, it is also known as the *Pearson correlation coefficient*.
It is also referred to as *simple correlation*, *bivariate correlation*, or merely the *correlation coefficient*.

Product Moment Correlation

From a sample of n observations, X and Y , the product moment correlation, r , can be calculated as:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Division of the numerator and denominator by $(n-1)$ gives

$$\begin{aligned} r &= \frac{\sum_{i=1}^n \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}}{\sqrt{\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1} \sum_{i=1}^n \frac{(Y_i - \bar{Y})^2}{n-1}}} \\ &= \frac{COV_{xy}}{S_x S_y} \end{aligned}$$

Product Moment Correlation

- r varies between -1.0 and +1.0.
- The correlation coefficient between two variables will be the same regardless of their underlying units of measurement.

Explaining Attitude Toward the City of Residence



Table 17.1

Respondent No	Attitude Toward the City	Duration of Residence	Importance Attached to Weather
1	6	10	3
2	9	12	11
3	8	12	4
4	3	4	1
5	10	12	11
6	4	6	1
7	5	8	7
8	2	2	4
9	11	18	8
10	9	9	10
11	10	17	8
12	2	2	5

Product Moment Correlation

The **correlation coefficient** may be calculated as follows:

$$\bar{X} = (10 + 12 + 12 + 4 + 12 + 6 + 8 + 2 + 18 + 9 + 17 + 2)/12 \\ = 9.333$$

$$\bar{Y} = (6 + 9 + 8 + 3 + 10 + 4 + 5 + 2 + 11 + 9 + 10 + 2)/12 \\ = 6.583$$

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) &= (10 - 9.33)(6 - 6.58) + (12 - 9.33)(9 - 6.58) \\ &+ (12 - 9.33)(8 - 6.58) + (4 - 9.33)(3 - 6.58) \\ &+ (12 - 9.33)(10 - 6.58) + (6 - 9.33)(4 - 6.58) \\ &+ (8 - 9.33)(5 - 6.58) + (2 - 9.33)(2 - 6.58) \\ &+ (18 - 9.33)(11 - 6.58) + (9 - 9.33)(9 - 6.58) \\ &+ (17 - 9.33)(10 - 6.58) + (2 - 9.33)(2 - 6.58) \\ &= -0.3886 + 6.4614 + 3.7914 + 19.0814 \\ &+ 9.1314 + 8.5914 + 2.1014 + 33.5714 \\ &+ 38.3214 - 0.7986 + 26.2314 + 33.5714 \\ &= 179.6668 \end{aligned}$$

Product Moment Correlation

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})^2 &= (10 - 9.33)^2 + (12 - 9.33)^2 + (12 - 9.33)^2 + (4 - 9.33)^2 \\ &+ (12 - 9.33)^2 + (6 - 9.33)^2 + (8 - 9.33)^2 + (2 - 9.33)^2 \\ &+ (18 - 9.33)^2 + (9 - 9.33)^2 + (17 - 9.33)^2 + (2 - 9.33)^2 \\ &= 0.4489 + 7.1289 + 7.1289 + 28.4089 \\ &+ 7.1289 + 11.0889 + 1.7689 + 53.7289 \\ &+ 75.1689 + 0.1089 + 58.8289 + 53.7289 \\ &= 304.6668 \end{aligned}$$

$$\begin{aligned} \sum_{i=1}^n (Y_i - \bar{Y})^2 &= (6 - 6.58)^2 + (9 - 6.58)^2 + (8 - 6.58)^2 + (3 - 6.58)^2 \\ &+ (10 - 6.58)^2 + (4 - 6.58)^2 + (5 - 6.58)^2 + (2 - 6.58)^2 \\ &+ (11 - 6.58)^2 + (9 - 6.58)^2 + (10 - 6.58)^2 + (2 - 6.58)^2 \\ &= 0.3364 + 5.8564 + 2.0164 + 12.8164 \\ &+ 11.6964 + 6.6564 + 2.4964 + 20.9764 \\ &+ 19.5364 + 5.8564 + 11.6964 + 20.9764 \\ &= 120.9168 \end{aligned}$$

$$\text{Thus, } r = \frac{179.6668}{\sqrt{(304.6668)(120.9168)}} = 0.9361$$

Decomposition of the Total Variation

$$\begin{aligned}
 r^2 &= \frac{\text{Explained variation}}{\text{Total variation}} \\
 &= \frac{SS_x}{SS_y} \\
 &= \frac{\text{Total variation} - \text{Error variation}}{\text{Total variation}} \\
 &= \frac{SS_y - SS_{\text{error}}}{SS_y}
 \end{aligned}$$

Decomposition of the Total Variation

- When it is computed for a population rather than a sample, the product moment correlation is denoted by ρ , the Greek letter rho. The coefficient r is an estimator of ρ .
- The statistical significance of the relationship between two variables measured by using r can be conveniently tested. The hypotheses are:

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

Decomposition of the Total Variation

The test statistic is:

$$t = r \left[\frac{n-2}{1-r^2} \right]^{1/2}$$

which has a t distribution with $n - 2$ degrees of freedom.

For the correlation coefficient calculated based on the data given in Table 17.1,

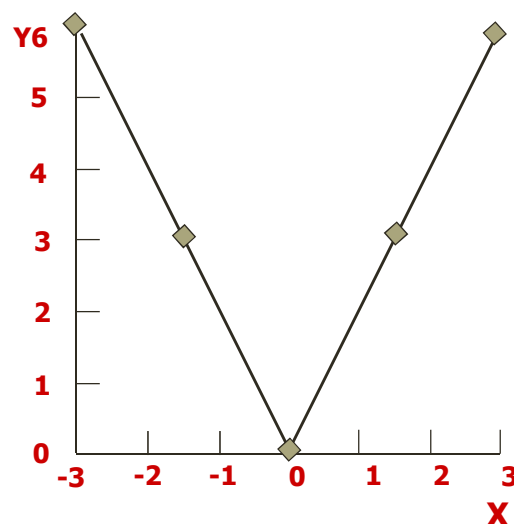
$$t = 0.9361 \left[\frac{12-2}{1 - (0.9361)^2} \right]^{1/2}$$

$$= 8.414$$

and the degrees of freedom = $12 - 2 = 10$. From the t distribution table (Table 4 in the Statistical Appendix), the critical value of t for a two-tailed test and $\alpha = 0.05$ is 2.228. Hence, the null hypothesis of no relationship between X and Y is rejected.

A Nonlinear Relationship for Which $r = 0$

Fig. 17.1



Partial Correlation

A **partial correlation coefficient** measures the association between two variables after controlling for, or adjusting for, the effects of one or more additional variables.

$$r_{xy \cdot z} = \frac{r_{xy} - (r_{xz})(r_{yz})}{\sqrt{1 - r_{xz}^2} \sqrt{1 - r_{yz}^2}}$$

- Partial correlations have an *order* associated with them. The order indicates how many variables are being adjusted or controlled.
- The simple correlation coefficient, r , has a zero-order, as it does not control for any additional variables while measuring the association between two variables.

Partial Correlation

- The coefficient $r_{xy \cdot z}$ is a **first-order partial correlation coefficient**, as it controls for the effect of one additional variable, Z .
- A **second-order partial correlation coefficient** controls for the effects of two variables, a third-order for the effects of three variables, and so on.
- The special case when a partial correlation is larger than its respective zero-order correlation involves a **suppressor effect**.

Part Correlation Coefficient

The **part correlation coefficient** represents the correlation between Y and X when the linear effects of the other independent variables have been removed from X but not from Y. The part correlation coefficient, $r_{y(x \cdot z)}$ is calculated as follows:

$$r_{y(x \cdot z)} = \frac{r_{xy} - r_{yz}r_{xz}}{\sqrt{1 - r_{xz}^2}}$$

The partial correlation coefficient is generally viewed as more important than the part correlation coefficient.

Nonmetric Correlation

- If the nonmetric variables are ordinal and numeric, Spearman's rho, ρ_s , and Kendall's tau, τ , are two measures of **nonmetric correlation**, which can be used to examine the correlation between them.
- Both these measures use rankings rather than the absolute values of the variables, and the basic concepts underlying them are quite similar. Both vary from -1.0 to +1.0 (see Chapter 15).
- In the absence of ties, **Spearman's** ρ_s yields a closer approximation to the Pearson product moment correlation coefficient, ρ , than Kendall's τ . In these cases, the absolute magnitude of τ tends to be smaller than Pearson's ρ .
- On the other hand, when the data contain a large number of tied ranks, **Kendall's** τ seems more appropriate.

Regression Analysis

Regression analysis examines associative relationships between a metric dependent variable and one or more independent variables in the following ways:

- Determine whether the independent variables explain a significant variation in the dependent variable: whether a relationship exists.
- Determine how much of the variation in the dependent variable can be explained by the independent variables: strength of the relationship.
- Determine the structure or form of the relationship: the mathematical equation relating the independent and dependent variables.
- Predict the values of the dependent variable.
- Control for other independent variables when evaluating the contributions of a specific variable or set of variables.
- Regression analysis is concerned with the nature and degree of association between variables and does not imply or assume any causality.

Statistics Associated with Bivariate Regression Analysis

- **Bivariate regression model.** The basic regression equation is $Y_i = \beta_0 + \beta_1 X_i + e_i$, where Y = dependent or criterion variable, X = independent or predictor variable, β_0 = intercept of the line, β_1 = slope of the line, and e_i is the error term associated with the i th observation.
- **Coefficient of determination.** The strength of association is measured by the coefficient of determination, r^2 . It varies between 0 and 1 and signifies the proportion of the total variation in Y that is accounted for by the variation in X .
- **Estimated or predicted value.** The estimated or predicted value of Y_i is $\hat{Y}_i = a + b x$, where $\hat{Y}_i$ is the predicted value of Y_i , and a and b are estimators of β_0 and β_1 , respectively.

Statistics Associated with Bivariate Regression Analysis

- **Regression coefficient.** The estimated parameter b is usually referred to as the non-standardized regression coefficient.
- **Scattergram.** A scatter diagram, or scattergram, is a plot of the values of two variables for all the cases or observations.
- **Standard error of estimate.** This statistic, SEE, is the standard deviation of the actual Y values from the predicted $\hat{Y}$ values.
- **Standard error.** The standard deviation of b , SE_b , is called the standard error.

Statistics Associated with Bivariate Regression Analysis

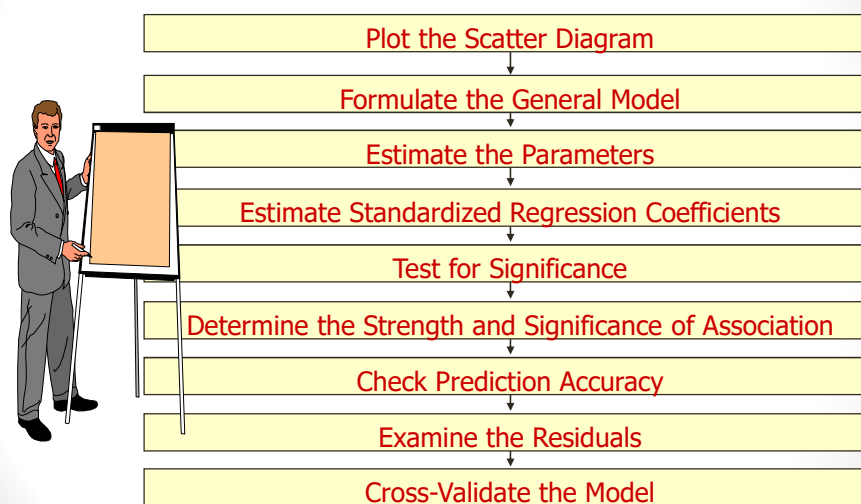
- **Standardized regression coefficient.** Also termed the beta coefficient or beta weight, this is the slope obtained by the regression of Y on X when the data are standardized.
- **Sum of squared errors.** The distances of all the points from the regression line are squared and added together to arrive at the sum of squared errors, which is a measure of total error, $\sum e_j^2$
- **t statistic.** A t statistic with $n - 2$ degrees of freedom can be used to test the null hypothesis that no linear relationship exists between X and Y , or $H_0: \beta = 0$, where $t = b / SE_b$

Conducting Bivariate Regression Analysis

Plot the Scatter Diagram

- A **scatter diagram**, or **scattergram**, is a plot of the values of two variables for all the cases or observations.
- The most commonly used technique for fitting a straight line to a scattergram is the **least-squares procedure**. In fitting the line, the least-squares procedure minimizes the sum of squared errors, $\sum e_j^2$.

Conducting Bivariate Regression Analysis



Conducting Bivariate Regression Analysis

Formulate the Bivariate Regression Model

In the bivariate regression model, the general form of a straight line is: $Y = \beta_0 + \beta_1 X$

where

Y = dependent or criterion variable

X = independent or predictor variable

β_0 = intercept of the line

β_1 = slope of the line

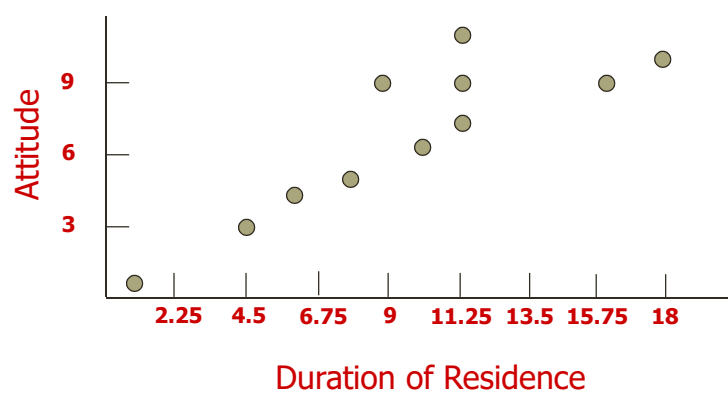
The regression procedure adds an error term to account for the probabilistic or stochastic nature of the relationship:

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

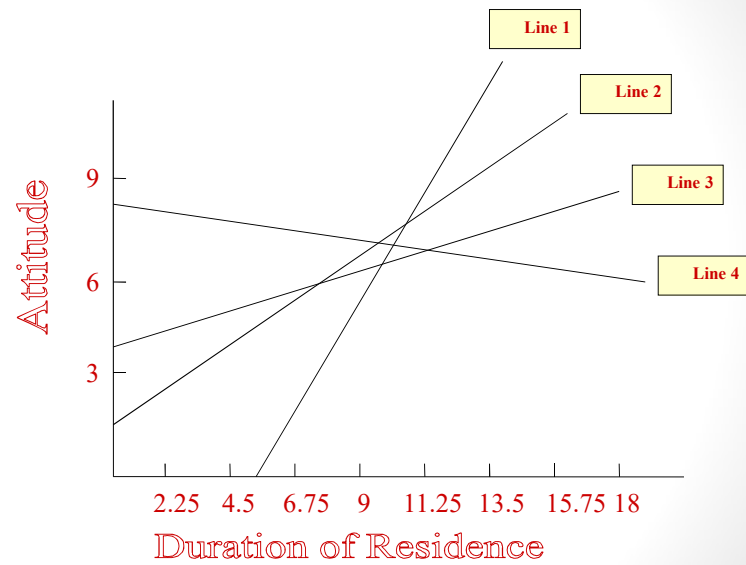
where e_i is the error term associated with the i th observation.

Plot of Attitude with Duration

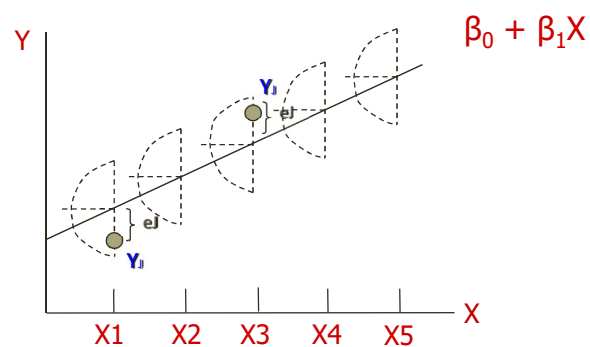
Fig. 17.3



Which Straight Line Is Best?



Bivariate Regression



Conducting Bivariate Regression Analysis

Estimate the Parameters

In most cases, β_0 and β_1 are unknown and are estimated from the sample observations using the equation

$$\hat{Y}_i = a + b x_i$$

where $\hat{Y}_i$ is the estimated or predicted value of Y_i , and a and b are estimators of β_0 and β_1 , respectively.

$$b = \frac{COV_{xy}}{S_x^2}$$

$$= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$= \frac{\sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^n X_i^2 - n\bar{X}^2}$$

Conducting Bivariate Regression Analysis

Estimate the Parameters

The intercept, a , may then be calculated using:

$$a = \bar{Y} - b \bar{X}$$

For the data in Table 17.1, the estimation of parameters may be illustrated as follows:

$$\sum_{i=1}^{12} X_i Y_i$$

$$= (10)(6) + (12)(9) + (12)(8) + (4)(3) + (12)(10) + (6)(4) \\ + (8)(5) + (2)(2) + (18)(11) + (9)(9) + (17)(10) + (2)(2) \\ = 917$$

$$\sum_{i=1}^{12} X_i^2 = 102 + 122 + 122 + 42 + 122 + 62 \\ + 82 + 22 + 182 + 92 + 172 + 22 \\ = 1350$$

Conducting Bivariate Regression Analysis

Estimate the Parameters

It may be recalled from earlier calculations of the simple correlation that:

$$\bar{X} = 9.333$$

$$\bar{Y} = 6.583$$

Given $n = 12$, b can be calculated as:

$$b = \frac{917 - (12)(9.333)(6.583)}{1350 - (12)(9.333)^2}$$

$$= 0.5897$$

$$a = \bar{Y} - b\bar{X}$$

$$= 6.583 - (0.5897)(9.333)$$

$$= 1.0793$$

Conducting Bivariate Regression Analysis

Estimate the Standardized Regression Coefficient

- **Standardization** is the process by which the raw data are transformed into new variables that have a mean of 0 and a variance of 1 (Chapter 14).
- When the data are standardized, the intercept assumes a value of 0.
- The term **beta coefficient** or **beta weight** is used to denote the standardized regression coefficient.

$$B_{yx} = B_{xy} = r_{xy}$$

- There is a simple relationship between the standardized and non-standardized regression coefficients:

$$B_{yx} = b_{yx} (S_x / S_y)$$

Conducting Bivariate Regression Analysis Test for Significance

The statistical significance of the linear relationship between X and Y may be tested by examining the hypotheses:

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

A t statistic with $n - 2$ degrees of freedom can be used, where

$$t = \frac{b}{SE_b}$$

SE_b denotes the standard deviation of b and is called the **standard error**.

Conducting Bivariate Regression Analysis Test for Significance

Using a computer program, the regression of attitude on duration of residence, using the data shown in Table 17.1, yielded the results shown in Table 17.2. The intercept, a , equals 1.0793, and the slope, b , equals 0.5897. Therefore, the estimated equation is:

$$\text{Attitude } (\hat{Y}) = 1.0793 + 0.5897 (\text{Duration of residence})$$

The standard error, or standard deviation of b is estimated as 0.07008, and the value of the t statistic as $t = 0.5897/0.0700 = 8.414$, with $n - 2 = 10$ degrees of freedom.

From Table 4 in the Statistical Appendix, we see that the critical value of t with 10 degrees of freedom and $\alpha = 0.05$ is 2.228 for a two-tailed test. Since the calculated value of t is larger than the critical value, the null hypothesis is rejected.

Conducting Bivariate Regression Analysis

Determine the Strength and Significance of Association

The total variation, SS_y , may be decomposed into the variation accounted for by the regression line, SS_{reg} , and the error or residual variation, SS_{error} or SS_{res} , as follows:

$$SS_y = SS_{reg} + SS_{res}$$

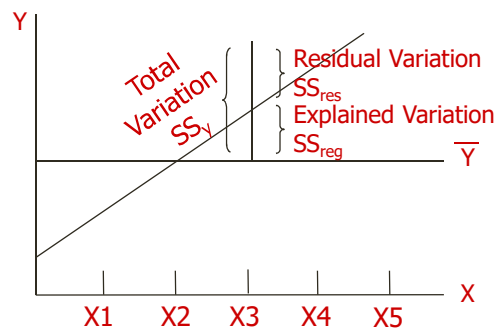
where

$$SS_y = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

$$SS_{reg} = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

$$SS_{res} = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Decomposition of the Total Variation in Bivariate Regression



Conducting Bivariate Regression Analysis

Determine the Strength and Significance of Association

The **strength of association** may then be calculated as follows:

$$r^2 = \frac{SS_{reg}}{SS_y}$$

$$= \frac{SS_y - SS_{res}}{SS_y}$$

To illustrate the calculations of r^2 , let us consider again the effect of attitude toward the city on the duration of residence. It may be recalled from earlier calculations of the simple correlation coefficient that:

$$SS_y = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

$$= 120.9168$$

Conducting Bivariate Regression Analysis

Determine the Strength and Significance of Association

The predicted values ($\hat{Y}$) can be calculated using the regression equation:

$$\text{Attitude (} \hat{Y} \text{)} = 1.0793 + 0.5897 (\text{Duration of residence})$$

For the first observation in Table 17.1, this value is:

$$(\hat{Y}) = 1.0793 + 0.5897 \times 10 = 6.9763.$$

For each successive observation, the predicted values are, in order, 8.1557, 8.1557, 3.4381, 8.1557, 4.6175, 5.7969, 2.2587, 11.6939, 6.3866, 11.1042, and 2.2587.

Conducting Bivariate Regression Analysis

Determine the Strength and Significance of Association

Therefore,

$$\begin{aligned}
 SS_{reg} &= \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \\
 &= (6.9763 - 6.5833)^2 + (8.1557 - 6.5833)^2 \\
 &\quad + (8.1557 - 6.5833)^2 + (3.4381 - 6.5833)^2 \\
 &\quad + (8.1557 - 6.5833)^2 + (4.6175 - 6.5833)^2 \\
 &\quad + (5.7969 - 6.5833)^2 + (2.2587 - 6.5833)^2 \\
 &\quad + (11.6939 - 6.5833)^2 + (6.3866 - 6.5833)^2 \\
 &\quad + (11.1042 - 6.5833)^2 + (2.2587 - 6.5833)^2 \\
 &= 0.1544 + 2.4724 + 2.4724 + 9.8922 + 2.4724 \\
 &\quad + 3.8643 + 0.6184 + 18.7021 + 26.1182 \\
 &\quad + 0.0387 + 20.4385 + 18.7021 \\
 &= 105.9524
 \end{aligned}$$

Conducting Bivariate Regression Analysis

Determine the Strength and Significance of Association

$$\begin{aligned}
 SS_{res} &= \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = (6 - 6.9763)^2 + (9 - 8.1557)^2 + (8 - 8.1557)^2 \\
 &\quad + (3 - 3.4381)^2 + (10 - 8.1557)^2 + (4 - 4.6175)^2 \\
 &\quad + (5 - 5.7969)^2 + (2 - 2.2587)^2 + (11 - 11.6939)^2 \\
 &\quad + (9 - 6.3866)^2 + (10 - 11.1042)^2 + (2 - 2.2587)^2 \\
 &= 14.9644
 \end{aligned}$$

It can be seen that $SS_y = SS_{reg} + SS_{res}$. Furthermore,

$$\begin{aligned}
 r^2 &= SS_{reg} / SS_y \\
 &= 105.9524 / 120.9168 \\
 &= 0.8762
 \end{aligned}$$

Conducting Bivariate Regression Analysis

Determine the Strength and Significance of Association

Another, equivalent test for examining the significance of the linear relationship between X and Y (significance of b) is the test for the significance of the coefficient of determination. The hypotheses in this case are:

$$H_0: R^2_{pop} = 0$$

$$H_1: R^2_{pop} > 0$$

Conducting Bivariate Regression Analysis

Determine the Strength and Significance of Association

The appropriate test statistic is the F statistic:

$$F = \frac{SS_{reg}}{SS_{res}/(n-2)}$$

which has an F distribution with 1 and $n - 2$ degrees of freedom. The F test is a generalized form of the t test (see Chapter 15). If a random variable is t distributed with n degrees of freedom, then t^2 is F distributed with 1 and n degrees of freedom. Hence, the F test for testing the significance of the coefficient of determination is equivalent to testing the following hypotheses:

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

or

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

Conducting Bivariate Regression Analysis

Determine the Strength and Significance of Association

From Table 17.2, it can be seen that:

$$r^2 = 105.9522 / (105.9522 + 14.9644)$$

$$= 0.8762$$

Which is the same as the value calculated earlier. The value of the F statistic is:

$$F = 105.9522 / (14.9644 / 10)$$

$$= 70.8027$$

with 1 and 10 degrees of freedom. The calculated F statistic exceeds the critical value of 4.96 determined from Table 5 in the Statistical Appendix. Therefore, the relationship is significant at $\alpha = 0.05$, corroborating the results of the t test.

α

Bivariate Regression



Multiple R	0.93608
R^2	0.87624
Adjusted R^2	0.86387
Standard Error	1.22329

ANALYSIS OF VARIANCE

	df	Sum of Squares	Mean Square
Regression	1	105.95222	105.95222
Residual	10	14.96444	1.49644
$F = 70.80266$		Significance of $F = 0.0000$	

VARIABLES IN THE EQUATION

Variable	b	SE_b	Beta (β)	T of T	Significance
Duration	0.58972	0.07008	0.93608	8.414	0.0000
(Constant)	1.07932	0.74335		1.452	0.1772

Conducting Bivariate Regression Analysis

Check Prediction Accuracy

To estimate the accuracy of predicted values, $\hat{Y}$, it is useful to calculate the standard error of estimate, SEE.

$$SEE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2}}$$

or

$$SEE = \sqrt{\frac{SS_{res}}{n-2}}$$

or more generally, if there are k independent variables,

$$SEE = \sqrt{\frac{SS_{res}}{n-k-1}}$$

For the data given in Table 17.2, the SEE is estimated as follows:

$$SEE = \sqrt{14.9644/(12-2)} \\ = 1.22329$$

Assumptions

- The error term is **normally** distributed. For each fixed value of X , the distribution of Y is normal.
- The means of all these normal distributions of Y , given X , lie on a **straight line** with slope b .
- The mean of the error term is **0**.
- The variance of the error term is **constant**. This variance does not depend on the values assumed by X .
- The error terms are **uncorrelated**. In other words, the observations have been drawn independently.

Multiple Regression

The general form of the **multiple regression model** is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k + \mathbf{e}$$

which is estimated by the following equation:

$$\hat{Y} = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_k X_k$$

As before, the coefficient a represents the intercept, but the b 's are now the partial regression coefficients.

Statistics Associated with Multiple Regression

- **Adjusted R^2 .** R^2 , coefficient of multiple determination, is adjusted for the number of independent variables and the sample size to account for the diminishing returns. After the first few variables, the additional independent variables do not make much contribution.
- **Coefficient of multiple determination.** The strength of association in multiple regression is measured by the square of the multiple correlation coefficient, R^2 , which is also called the coefficient of multiple determination.
- **F test.** The F test is used to test the null hypothesis that the coefficient of multiple determination in the population, R^2_{pop} , is zero. This is equivalent to testing the null hypothesis. The test statistic has an F distribution with k and $(n - k - 1)$ degrees of freedom.

Statistics Associated with Multiple Regression

- **Partial F test.** The significance of a partial regression coefficient, b_i , of X_i may be tested using an incremental F statistic. The incremental F statistic is based on the increment in the explained sum of squares resulting from the addition of the independent variable X_i to the regression equation after all the other independent variables have been included.
- **Partial regression coefficient.** The partial regression coefficient, b_1 , denotes the change in the predicted value, $\hat{Y}$, per unit change in X_1 when the other independent variables, X_2 to X_k , are held constant.

Conducting Multiple Regression Analysis

Partial Regression Coefficients

To understand the meaning of a partial regression coefficient, let us consider a case in which there are two independent variables, so that:

$$\hat{Y} = a + b_1X_1 + b_2X_2$$

- First, note that the relative magnitude of the partial regression coefficient of an independent variable is, in general, different from that of its bivariate regression coefficient.
- The interpretation of the partial regression coefficient, b_1 , is that it represents the expected change in Y when X_1 is changed by one unit but X_2 is held constant or otherwise controlled. Likewise, b_2 represents the expected change in Y for a unit change in X_2 , when X_1 is held constant. Thus, calling b_1 and b_2 partial regression coefficients is appropriate.

Conducting Multiple Regression Analysis Partial Regression Coefficients

- It can also be seen that the combined effects of X_1 and X_2 on Y are additive. In other words, if X_1 and X_2 are each changed by one unit, the expected change in Y would be $(b_1 + b_2)$.
- Suppose one was to remove the effect of X_2 from X_1 . This could be done by running a regression of X_1 on X_2 . In other words, one would estimate the equation $\hat{X}_1 = a + b X_2$ and calculate the residual $X_r = (X_1 - \hat{X}_1)$. The **partial regression coefficient**, b_1 , is then equal to the bivariate regression coefficient, b_r , obtained from the equation $\hat{Y} = a + b_r X_r$.

Conducting Multiple Regression Analysis Partial Regression Coefficients

- Extension to the case of k variables is straightforward. The partial regression coefficient, b_1 , represents the expected change in Y when X_1 is changed by one unit and X_2 through X_k are held constant. It can also be interpreted as the bivariate regression coefficient, b , for the regression of Y on the residuals of X_1 , when the effect of X_2 through X_k has been removed from X_1 .
- The relationship of the standardized to the non-standardized coefficients remains the same as before:

$$B_1 = b_1 (S_{x1}/S_y)$$

$$B_k = b_k (S_{xk}/S_y)$$

The estimated regression equation is:

$$(\hat{Y}) = 0.33732 + 0.48108 X_1 + 0.28865 X_2$$

Or

$$\text{Attitude} = 0.33732 + 0.48108 (\text{Duration}) + 0.28865 (\text{Importance})$$

Multiple Regression

Multiple R 0.97210
 R^2 0.94498
 Adjusted R^2 0.93276
 Standard Error 0.85974



ANALYSIS OF VARIANCE

	df	Sum of Squares	Mean Square
Regression	2	114.26425	57.13213
Residual	9	6.65241	0.73916
$F = 77.29364$ Significance of $F = 0.0000$			

VARIABLES IN THE EQUATION

Variable	b	SE_b	Beta (β)	T of T	Significance
IMPORTANCE	0.28865	0.08608	0.31382	3.353	0.0085
DURATION	0.48108	0.05895	0.76363	8.160	0.0000
(Constant)	0.33732	0.56736		0.595	0.5668

Conducting Multiple Regression Analysis Strength of Association

$$SS_y = SS_{reg} + SS_{res}$$

where

$$SS_y = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

$$SS_{reg} = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

$$SS_{res} = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Conducting Multiple Regression Analysis Strength of Association

The strength of association is measured by the square of the multiple correlation coefficient, R^2 , which is also called the **coefficient of multiple determination**.

$$R^2 = \frac{SS_{reg}}{SS_y}$$

R^2 is adjusted for the number of independent variables and the sample size by using the following formula:

$$\text{Adjusted } R^2 = R^2 - \frac{k(1 - R^2)}{n - k - 1}$$

Conducting Multiple Regression Analysis Significance Testing

$$H_0 : R^2_{pop} = 0$$

This is equivalent to the following null hypothesis:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0$$

The overall test can be conducted by using an F statistic:

$$F = \frac{SS_{reg}/k}{SS_{res}/(n - k - 1)}$$

$$= \frac{R^2/k}{(1 - R^2)/(n - k - 1)}$$

which has an F distribution with k and $(n - k - 1)$ degrees of freedom.

Conducting Multiple Regression Analysis Significance Testing

Testing for the significance of the β_i 's can be done in a manner similar to that in the bivariate case by using t tests. The significance of the partial coefficient for importance attached to weather may be tested by the following equation:

$$t = \frac{b}{SE_b}$$

which has a t distribution with $n - k - 1$ degrees of freedom.

Conducting Multiple Regression Analysis Examination of Residuals

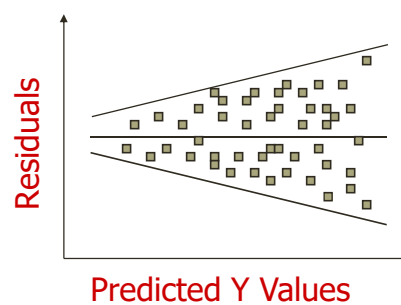
- A **residual** is the difference between the observed value of Y_i and the value predicted by the regression equation $\hat{Y}_i$.
- Scattergrams of the residuals, in which the residuals are plotted against the predicted values, $\hat{Y}_i$, time, or predictor variables, provide useful insights in examining the appropriateness of the underlying assumptions and regression model fit.
- The assumption of a normally distributed error term can be examined by constructing a histogram of the residuals.
- The assumption of constant variance of the error term can be examined by plotting the residuals against the predicted values of the dependent variable, $\hat{Y}_i$.

Conducting Multiple Regression Analysis

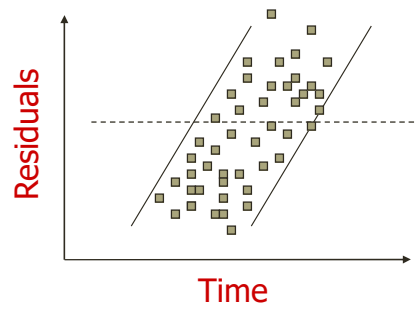
Examination of Residuals

- A plot of residuals against time, or the sequence of observations, will throw some light on the assumption that the error terms are uncorrelated.
- Plotting the residuals against the independent variables provides evidence of the appropriateness or inappropriateness of using a linear model. Again, the plot should result in a random pattern.
- To examine whether any additional variables should be included in the regression equation, one could run a regression of the residuals on the proposed variables.
- If an examination of the residuals indicates that the assumptions underlying linear regression are not met, the researcher can transform the variables in an attempt to satisfy the assumptions.

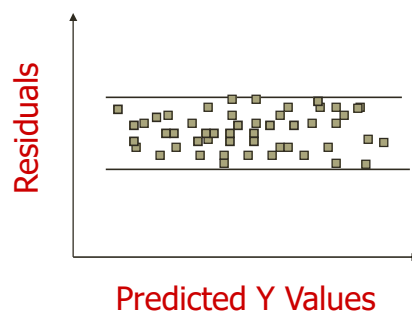
Residual Plot Indicating that Variance Is Not Constant



Residual Plot Indicating a Linear Relationship Between Residuals and Time



Plot of Residuals Indicating that a Fitted Model Is Appropriate



Stepwise Regression

The purpose of **stepwise regression** is to select, from a large number of predictor variables, a small subset of variables that account for most of the variation in the dependent or criterion variable. In this procedure, the predictor variables enter or are removed from the regression equation one at a time. There are several approaches to stepwise regression.

- **Forward inclusion.** Initially, there are no predictor variables in the regression equation. Predictor variables are entered one at a time, only if they meet certain criteria specified in terms of F ratio. The order in which the variables are included is based on the contribution to the explained variance.
- **Backward elimination.** Initially, all the predictor variables are included in the regression equation. Predictors are then removed one at a time based on the F ratio for removal.
- **Stepwise solution.** Forward inclusion is combined with the removal of predictors that no longer meet the specified criterion at each step.

Multicollinearity

- **Multicollinearity** arises when intercorrelations among the predictors are very high.
- Multicollinearity can result in several problems, including:
 - The partial regression coefficients may not be estimated precisely. The standard errors are likely to be high.
 - The magnitudes, as well as the signs of the partial regression coefficients, may change from sample to sample.
 - It becomes difficult to assess the relative importance of the independent variables in explaining the variation in the dependent variable.
 - Predictor variables may be incorrectly included or removed in stepwise regression.

Multicollinearity

- A simple procedure for adjusting for multicollinearity consists of using only one of the variables in a highly correlated set of variables.
- Alternatively, the set of independent variables can be transformed into a new set of predictors that are mutually independent by using techniques such as principal components analysis.
- More specialized techniques, such as ridge regression and latent root regression, can also be used.

Relative Importance of Predictors

Unfortunately, because the predictors are correlated, there is no unambiguous measure of relative importance of the predictors in regression analysis. However, several approaches are commonly used to assess the relative importance of predictor variables.

- **Statistical significance.** If the partial regression coefficient of a variable is not significant, as determined by an incremental F test, that variable is judged to be unimportant. An exception to this rule is made if there are strong theoretical reasons for believing that the variable is important.
- **Square of the simple correlation coefficient.** This measure, r^2 , represents the proportion of the variation in the dependent variable explained by the independent variable in a bivariate relationship.

Relative Importance of Predictors

- **Square of the partial correlation coefficient.** This measure, $R^2_{yxi.xjxk\ldots}$ is the coefficient of determination between the dependent variable and the independent variable, controlling for the effects of the other independent variables.
- **Square of the part correlation coefficient.** This coefficient represents an increase in R^2 when a variable is entered into a regression equation that already contains the other independent variables.
- **Measures based on standardized coefficients or beta weights.** The most commonly used measures are the absolute values of the beta weights, $|B_j|$, or the squared values, B_j^2 .
- **Stepwise regression.** The order in which the predictors enter or are removed from the regression equation is used to infer their relative importance.

Cross-Validation

- The regression model is estimated using the entire data set.
- The available data are split into two parts, the **estimation sample** and the **validation sample**. The estimation sample generally contains 50-90% of the total sample.
- The regression model is estimated using the data from the estimation sample only. This model is compared to the model estimated on the entire sample to determine the agreement in terms of the signs and magnitudes of the partial regression coefficients.
- The estimated model is applied to the data in the validation sample to predict the values of the dependent variable, $\hat{Y}_i$, for the observations in the validation sample.
- The observed values Y_i and the predicted values, $\hat{Y}_i$, in the validation sample are correlated to determine the simple r^2 . This measure, r^2 , is compared to R^2 for the total sample and to R^2 for the estimation sample to assess the degree of shrinkage.

Regression with Dummy Variables

Product Usage Category	Original Variable Code	Dummy Variable Code		
		D1	D2	D3
Nonusers.....	1	1	0	0
Light Users.....	2	0	1	0
Medium Users.....	3	0	0	1
Heavy Users.....	4	0	0	0

$$\hat{Y}_i = a + b_1D_1 + b_2D_2 + b_3D_3$$

- In this case, "heavy users" has been selected as a reference category and has not been directly included in the regression equation.
- The coefficient b_1 is the difference in predicted $\hat{Y}_i$ for nonusers, as compared to heavy users.

Analysis of Variance and Covariance with Regression

In regression with dummy variables, the predicted $\hat{Y}$ for each category is the mean of Y for each category.

Product Usage Category	Predicted Value $\hat{Y}$	Mean Value $\bar{Y}$
Nonusers.....	$a + b_1$	$a + b_1$
Light Users.....	$a + b_2$	$a + b_2$
Medium Users.....	$a + b_3$	$a + b_3$
Heavy Users.....	a	a

Analysis of Variance and Covariance with Regression

Given this equivalence, it is easy to see further relationships between dummy variable regression and one-way ANOVA.

Dummy Variable Regression

$$SS_{res} = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

$$SS_{reg} = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

$$R^2$$

Overall F test

One-Way ANOVA

$$= SS_{within} = SS_{error}$$

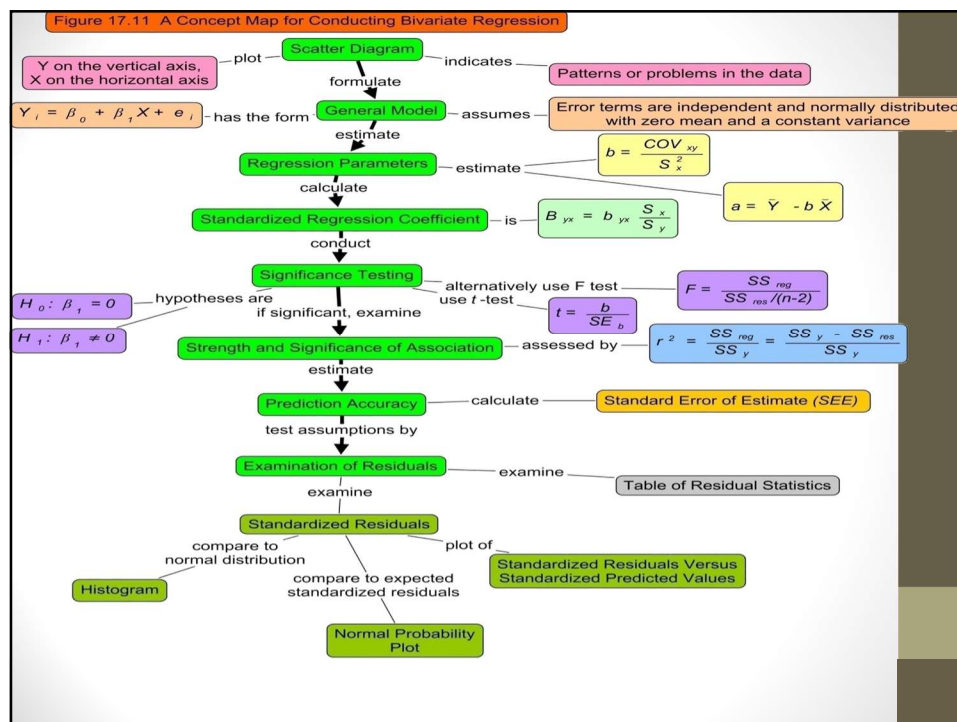
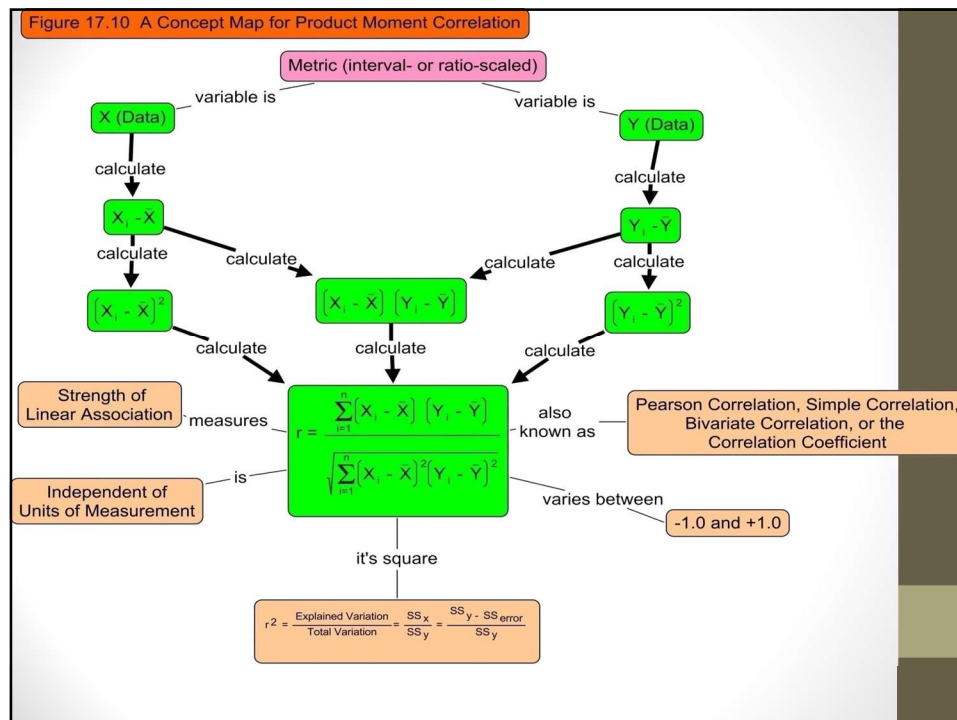
$$= SS_{between} = SS_x$$

$$= \eta^2$$

= F test

Complicated Models

- Mediated moderations
- Moderated mediations



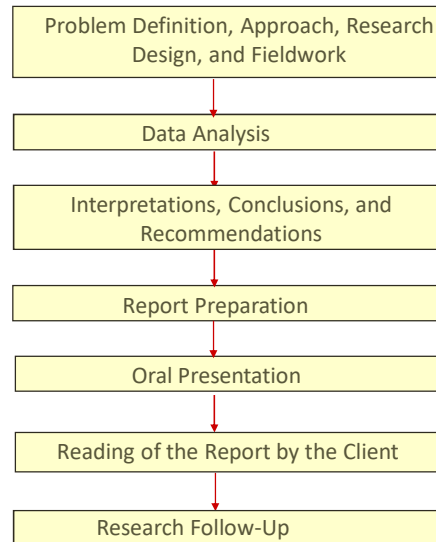
RESULTS PRESENTATION

Importance of the Report and Presentation

For the following reasons, the report and its presentation are important parts of the research project:

1. They are the tangible products of the research effort.
2. Management decisions are guided by the report and the presentation.
3. The involvement of many managers in the project is limited to the written report and the oral presentation.
4. Management's decision to undertake marketing research in the future or to use the particular research supplier again will be influenced by the perceived usefulness of the report and the presentation.

The Report Preparation and Presentation Process



Report Format

- I. Title page
- II. Letter of transmittal
- III. Letter of authorization
- IV. Table of contents
- V. List of tables
- VI. List of graphs
- VII. List of appendices
- VIII. List of exhibits
- IX. Executive summary
 - a. Major findings
 - b. Conclusions
 - c. Recommendations

Report Format

- X. Problem definition
 - a. Background to the problem
 - b. Statement of the problem
- XI. Approach to the problem
- XII. Research design
 - a. Type of research design
 - b. Information needs
 - c. Data collection from secondary sources
 - d. Data collection from primary sources
 - e. Scaling techniques
 - f. Questionnaire development and pretesting
 - g. Sampling techniques
 - h. Fieldwork

Report Format

- XIII. Data analysis
 - a. Methodology
 - b. Plan of data analysis
- XIV. Results
- XV. Limitations and caveats
- XVI. Conclusions and recommendations
- XVII. Exhibits
 - a. Questionnaires and forms
 - b. Statistical output
 - c. Lists

TNS-Global Guidelines: Title Page

Use client language in title – avoid "research-eze."

- "Practices Followed in Selecting Long-Distance Carriers" is better than "Long-Distance Service Study."
- "Customers' Reactions to an Expanded Financial/Insurance Relationship" is better than "Relationship Study."

TNS-Global Guidelines: Conclusions

Conclusions concerning, for example:

- customer behavior
- customer attitudes or perceptions
- the nature of the markets studied

Generally, in studies with samples designed to represent the market. Avoid interesting results that are not relevant to the conclusions.

- May be in the form of statement or paragraphs
- Use subheadings to identify conclusions covering different subjects or market segments

TNS-Global Guidelines: Recommendations

Recommendations regarding actions that should be taken or considered in light of the research results:

- Add/drop a product
- What to say in advertising—advertising positioning
- Market segments to select as primary targets
- How to price product
- Further research that should be considered

Should be related to the stated purpose of the research.

Sometimes omitted, for example:

- Client staff members want to author the recommendations
- Study designed merely to familiarize client with a market

Most clients are interested in our suggestions, in spite of the fact that we may not be familiar with internal financial issues and other internal corporate factors.

Report Writing

- **Readers.** A report should be written for a specific reader or readers: the marketing managers who will use the results.
- **Easy to follow.** The report should be easy to follow. It should be structured logically and written clearly.
- **Presentable and professional appearance.** The look of a report is important.
- **Objective.** Objectivity is a virtue that should guide report writing. The rule is, "Tell it like it is."
- **Reinforce text with tables and graphs.** It is important to reinforce key information in the text with tables, graphs, pictures, maps, and other visual devices.
- **Terse.** A report should be terse and concise. Yet, brevity should not be achieved at the expense of completeness.

Guidelines for Tables

- **Title and number.** Every table should have a number (1a) and title (1b).
- **Arrangement of data items.** The arrangement of data items in a table should emphasize the most significant aspect of the data.
- **Basis of measurement.** The basis or unit of measurement should be clearly stated (3a).
- **Leaders, rulings, spaces.** *Leaders*, dots or hyphens used to lead the eye horizontally, impart uniformity and improve readability (4a). Instead of ruling the table horizontally or vertically, white spaces (4b) are used to set off data items. Skipping lines after different sections of the data can also assist the eye. Horizontal rules (4c) are often used after the headings.
- **Explanations and comments: Headings, stubs, and footnotes.** Designations placed over the vertical columns are called headings (5a). Designations placed in the left-hand column are called stubs (5b). Information that cannot be incorporated in the table should be explained by footnotes (5c).
- **Sources of the data.** If the data contained in the table are secondary, the source of data should be cited (6a).

U.S. Auto Sales 2003 - 2007

TABLE 23.1
U.S. Automotive Sales 2003-2007

	2003	2004	2005	2006	2007
GM.....	4,756,403	4,707,416	4,517,730	4,124,645	3,866,620
Ford.....	3,811,000	3,623,000	3,443,000	3,051,000	2,836,000
Chrysler...	2,129,000	2,287,000	2,305,000	2,087,000	2,100,000
Honda.....	1,982,000	2,103,000	2,271,000	2,556,000	2,942,000
Toyota.....	1,558,000	1,575,000	1,575,000	1,682,000	1,788,000
Nissan.....	856,000	1,013,000	1,075,000	1,035,000	1,100,000
Other*	1,907,597	1,991,584	1,800,345	2,016,193	1,467,380
Total.....	17,000,000	17,300,000	16,987,075	16,551,838	16,100,000

* - includes all other producers
Source: Company Websites

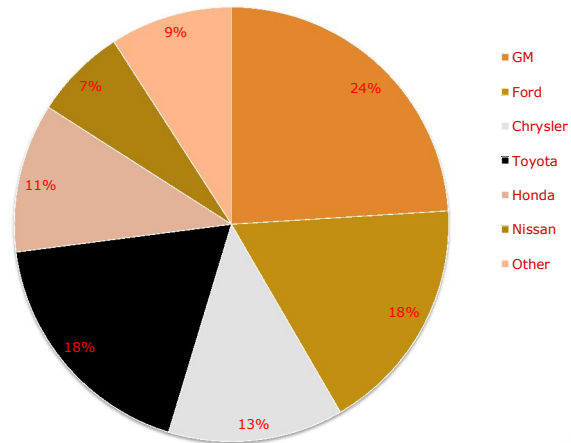
Guidelines for Graphs: Geographic and Other Maps

- **Geographic maps** can pertain to countries, states, counties, sales territories, and other divisions.
- Chapter 21 shows examples of **product-positioning maps**.

Guidelines for Graphs: Round or Pie Charts

- In a **pie chart**, the area of each section, as a percentage of the total area of the circle, reflects the percentage associated with the value of a specific variable.
- A pie chart is not useful for displaying relationships over time or relationships among several variables.
- As a general guideline, a pie chart should not require more than seven sections.

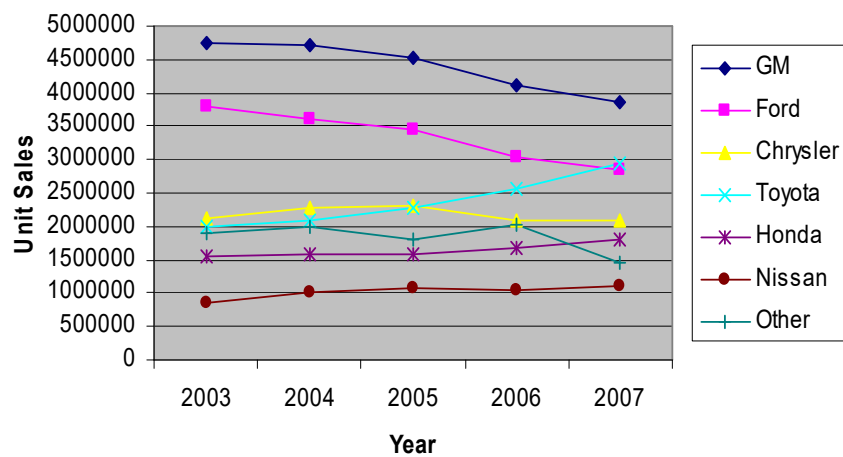
Pie Chart of 2007 U.S. Auto Sales



Guidelines for Graphs: Line Charts

- A **line chart** connects a series of data points using continuous lines.
- This is an attractive way of illustrating trends and changes over time.
- Several series can be compared on the same chart, and forecasts, interpolations, and extrapolations can be shown.

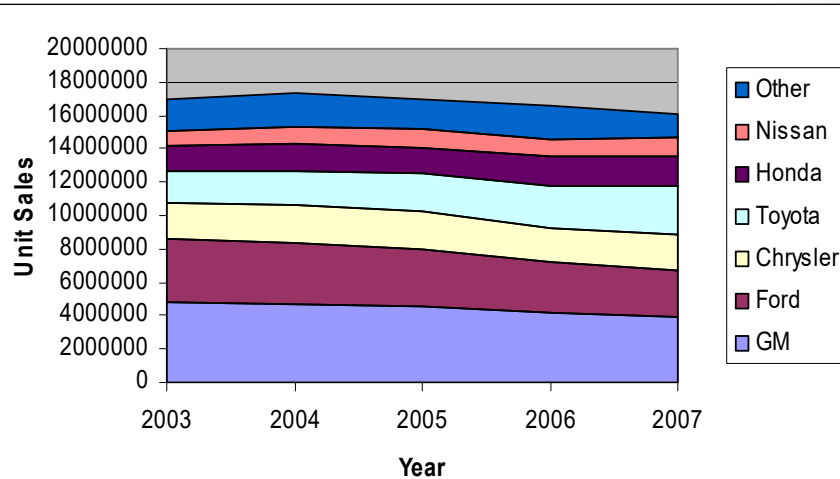
Line Chart of Total U.S. Auto Sales



Guidelines for Graphs: Line Charts

- A **stratum chart** is a set of line charts in which the data are successively aggregated over the series.
- Areas between the line charts display the magnitudes of the relevant variables.

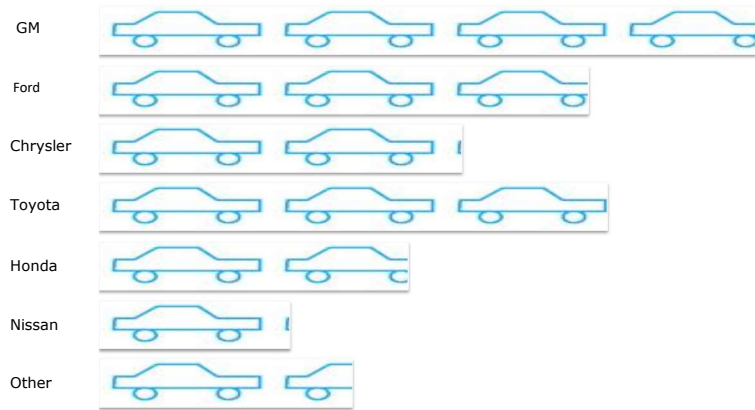
Stratum Chart of Total U.S. Auto Sales



Guidelines for Graphs: Pictographs

- A **pictograph** uses small pictures or symbols to display the data.
- Pictographs do not depict results precisely. Hence, caution should be exercised when using them.

Pictograph for 2007 U.S. Auto Sales

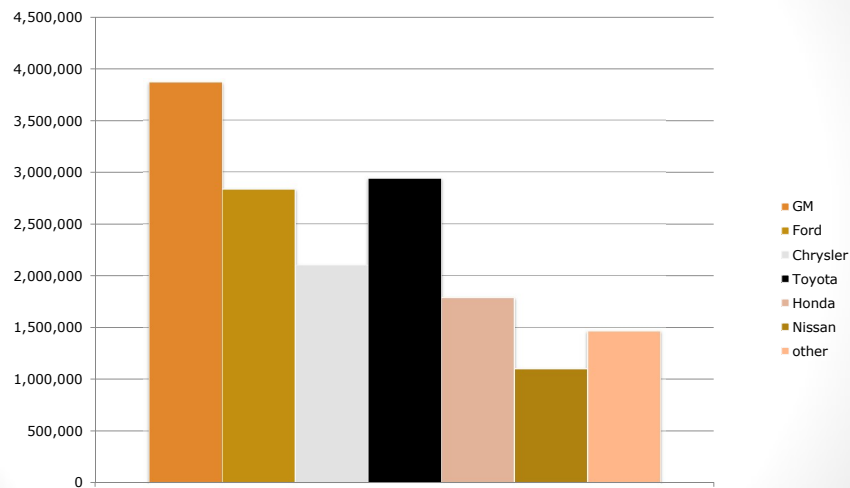


*Each Symbol Equals 1,000,000 Units

Guidelines for Graphs: Histograms and Bar Charts

- A **bar chart** displays data in various bars that may be positioned horizontally or vertically.
- The **histogram** is a vertical bar chart in which the height of the bars represents the relative or cumulative frequency of occurrence of a specific variable.

Histogram of 2007 U.S. Auto Sales



Guidelines for Graphs: Schematic Figures and Flow Charts

- **Schematic figures** and **flow charts** take on a number of different forms. They can be used to display the steps or components of a process, as in Figure 22.1.
- Another useful form of these charts is a **classification diagram**. Examples of classification charts for classifying secondary data were provided in Chapter 4 (Figures 4.1 to 4.4).
- An example of a flow chart for questionnaire design was given in Chapter 10 (Figure 10.2).

Oral Presentation

- The key to an effective presentation is preparation.
- A written script or detailed outline should be prepared following the format of the written report.
- The presentation must be geared to the audience.
- The presentation should be rehearsed several times before it is made to the management.
- Visual aids, such as tables and graphs, should be displayed with a variety of media.
- It is important to maintain eye contact and interact with the audience during the presentation.

Oral Presentation

- Filler words like "uh," "y'know," and "all right," should not be used.
- The **"Tell 'Em" principle** is effective for structuring a presentation.
- Another useful guideline is the **"KISS 'Em" principle**, which states: Keep It Simple and Straightforward (hence the acronym KISS).
- Body language should be employed.
- The speaker should vary the volume, pitch, voice quality, articulation, and rate while speaking.
- The presentation should terminate with a strong closing.

Reading the Research Report

- **Addresses the Problem** – The problem being addressed should be clearly identified and the relevant background information provided.
- The **research design** should be clearly described in non-technical terms.
- **Execution of the Research Procedures** – The reader should pay special attention to the manner in which the research procedures were executed.
- **Numbers and statistics** reported in tables and graphs should be examined carefully by the reader.

Reading the Research Report

- **Interpretation and Conclusions** – The interpretation of the basic results should be differentiated from the results per se. Any conclusions or recommendations made without a specification of the underlying assumptions or limitations should be treated cautiously by the reader.
- **Generalizability** – It is the responsibility of the researcher to provide evidence regarding the reliability, validity, and generalizability of the findings.
- **Disclosure** – The reader should carefully examine whether the spirit in which the report was written indicates an honest and complete disclosure of the research procedures and results.

Research Follow-up

- **Assisting the Client** – The researcher should answer questions that may arise and help the client to implement the findings.
- **Evaluation of the Research Project** – Every marketing research project provides an opportunity for learning and the researcher should critically evaluate the entire project to obtain new insights and knowledge.

Why repeat a research across countries?

- Develop universal theories
- Investigate the impacts of country-level variables
 - Culture
 - Social practices
 - National wealth
 - Political economy
- Find out international differences for market segmentation or “glocalization” strategies

[489]

Survey Methods: Telephone Interviewing and CATI

- In the United States and Canada, **telephone interviewing** is the dominant mode of questionnaire administration. The same situation exists in some European countries, such as Sweden and the Netherlands.
- In many other European countries, such as Great Britain, Finland, and Portugal, telephone interviewing is not the most popular method.
- In Hong Kong, 96% of households can be contacted by telephone. Yet, given the culture, this is not the most important mode of data collection.
- In developing countries, only a few households have telephones. Telephone directories tend to be incomplete and outdated. In many cultures, face-to-face relationships are predominant. These factors severely limit the use of telephone interviewing.
- Telephone interviews are most useful with relatively upscale consumers who are accustomed to business transactions by phone or consumers who can be reached by phone and can express themselves easily.

Survey Methods: In-Home Personal Interviews

- Due to high cost, the use of **in-home personal interviews** has declined in the United States and Canada, but this is the dominant mode of collecting survey data in many parts of Europe (e.g., Switzerland and Portugal) and the developing world.
- The majority of the surveys are done door-to-door, while some quick sociopolitical polls are carried out on the street using accidental routes.

Survey Methods: Mall Intercept and CAPI

- **Mall intercepts** constitute about 15% of the interviews in Canada and 20% in the United States.
- While mall intercepts are being conducted in some European countries, such as Sweden, they are not popular in Europe or developing countries.
- In contrast, **central location/street interviews** constitute the dominant method of collecting survey data in France and the Netherlands.
- However, some interesting developments with respect to **computer-assisted personal interviewing (CAPI)** are taking place in Europe.

Survey Methods: Mail Interviews

- Because of low cost, **mail interviews** continue to be used in most developed countries where literacy is high and the postal system is well developed.
- Mail interviews constitute 6.2% of the interviews in Canada and 7% in the United States. In countries where the educational level of the population is extremely high (Denmark, Finland, Iceland, Norway, Sweden, and the Netherlands), mail interviews are common.
- In Africa, Asia, and South America, however, the use of mail surveys and mail panels is low because of illiteracy and the large proportion of population living in rural areas.
- Mail surveys are typically more effective in industrial international marketing research, although it is difficult to identify the appropriate respondent within each firm and to personalize the address.

Survey Methods: Mail and Scanner Panels

- **Mail panels** are extensively used in the UK, France, West Germany, and the Netherlands. Mail and diary panels are also available in Finland, Sweden, Italy, Spain, and other European countries.
- Use of **scanner panels** may increase with the advent of new technology. For example, in Germany, two agencies (A. C. Nielsen and GfK-Nurnberg) have installed fully electronic scanner test markets, based on the Behavior Scan model from the United States. Nielsen will use on-the-air television; GfK, cable.

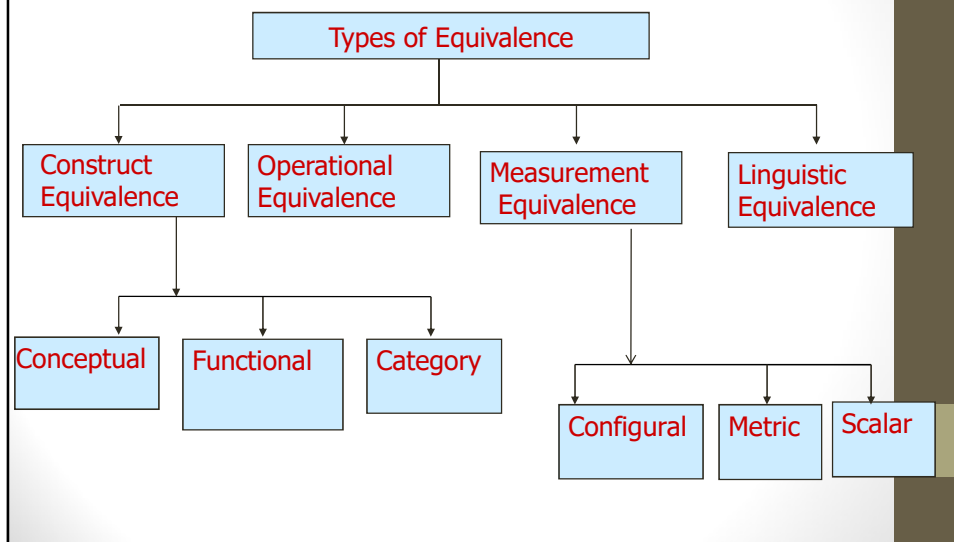
Survey Methods: Electronic Surveys

- In the U.S. and Canada, the use of **e-mail** and the **Internet surveys** is growing not only with business and institutional respondents but also with households.
- The popularity of both e-mail and Internet surveys is also growing overseas. Both these types of surveys are increasingly being used in Western Europe where the access to the Internet is freely available.
- However, in some parts of Eastern Europe and in other developing countries, e-mail access is restricted and Internet availability is even poorer. Hence, these methods are not suitable for surveying the general population in these countries.
- Multinational firms are using both e-mail and the Internet to survey their employees worldwide.
- An important consideration in selecting the methods of administering questionnaires is to ensure equivalence and comparability across countries.

Unique Challenges

- Can all data be simply put together and analyzed?
- Multi-level analyses
- Some equivalence needs to be established beforehand

Scaling and Measurement Equivalence in International Marketing Research



Measurement and Scaling

- **Construct equivalence** concerns whether the marketing constructs have the same meaning and significance in different countries. Construct equivalence is comprised of conceptual equivalence, functional equivalence, and category equivalence.
- **Conceptual equivalence** deals with the interpretation of brands, products, consumer behavior, and marketing effort, e.g., special sales.
- **Functional equivalence** examines whether a given concept or behavior serves the same role or function in different countries, e.g., bicycles.
- **Category equivalence** refers to the category in which stimuli like products, brands, and behaviors are grouped, e.g., principal shopper.

Measurement and Scaling

- **Measurement equivalence** concerns the comparability of responses to particular (sets of) items. It comprises of configural, metric, and scalar equivalence.
- **Configural equivalence** concerns the relationships of measured items to the latent constructs; the patterns of factor loadings should be the same.
- **Metric equivalence** refers to the unit of measurement; the factor loadings should be the same.
- **Scalar equivalence** refers to both the unit of measurement and the constant in the equation between the construct and the items measuring the construct (the intercept).

Measurement and Scaling

- **Operational equivalence** concerns how theoretical constructs are operationalized to make measurements, e.g., leisure.
- **Linguistic equivalence** refers to both the spoken and the written language forms used in scales, questionnaires, and interviewing. The scales and other verbal stimuli should be translated so that they are readily understood by respondents in different countries and have equivalent meaning.

Questionnaire Translation: Back Translation

- In **back translation**, the questionnaire is translated from the base language by a bilingual speaker whose native language is the language into which the questionnaire is being translated.
- This version is then retranslated back into the original language by a bilingual whose native language is the initial or base language.
- Translation errors can then be identified.
- Several repeat translations and back translations may be necessary to develop equivalent questionnaires, and this process can be cumbersome and time consuming.

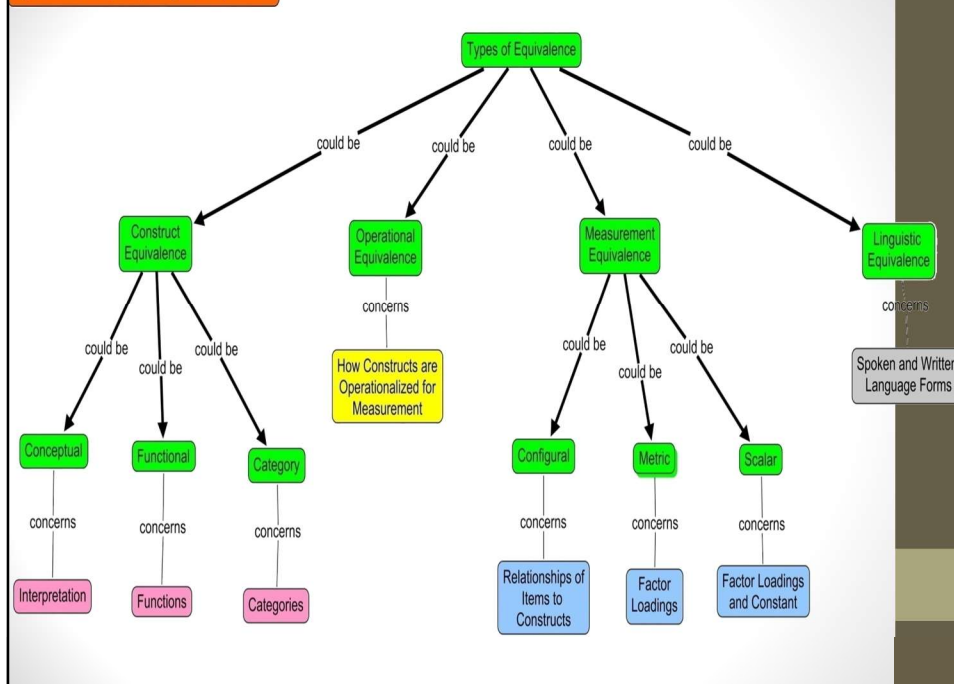
Questionnaire Translation: Parallel Translation

- In **parallel translation**, a committee of translators, each of whom is fluent in at least two of the languages in which the questionnaire will be administered, discusses alternative versions of the questionnaire and makes modifications until consensus is reached.
- In countries where several languages are spoken, the questionnaire should be translated into the language of each respondent subgroup.
- It is important that any nonverbal stimuli (pictures and advertisements) also be translated using similar procedures.

Statistical Software

INTERVIEWER by VOXCO of Montreal (Quebec), Canada (www.voxco.com) provides bilingual interviewing capability for computer-assisted telephone interviewing (CATI). With two keystrokes, operators can switch from a questionnaire written in English to the same questionnaire written in Spanish. INTERVIEWER offered switching capability between most of the major languages as of 2009.

Figure 24.3 Concept Map for Types of Equivalence



RESEARCH ETHICS

(505)

Ethical Issues

- Ethical issues related to the respondents and the general public are of primary concern.
- Disguise can violate the respondents' right to know and result in psychological harm.
- In debriefing sessions, respondents should be informed about the true purpose and given opportunities to ask questions.
- The use of qualitative research results for questionable purposes raises ethical concerns

Ethical Issues

- Deceptive procedures that violate respondents' right to privacy and informed consent should be avoided.
- Video- or audio-taping the respondents without their prior knowledge or consent raises ethical concerns.
- The comfort level of the respondents should be addressed.
- Physical or psychological hazards

The Unabomber

Ted Kaczynski



- https://en.wikipedia.org/wiki/Ted_Kaczynski

Famous Controversial Studies

- Milgram's Obedience Studies
- The Stanford Prison Experiment
- Why are they unethical?

(509)

First, an example brought to us by Radio Lab

- <https://www.wnycstudios.org/podcasts/radiolab/episodes/91721-oops>
- "Be Careful What you Plan for" Beginning at the 5 minute mark.

(510)

Murray Experiment Follow Up

- Review the summary of the APA guidelines for experiments
- Which regulations would disqualify the Murray experiment from IRB approval?
 - Be specific
 - Is there ambiguity/do you need more info?

(511)

Additional examples

- 2 other studies that WERE conducted.
- Consider how these studies could pass IRB regulations
- After each example, there will be a slide with topics to discuss with your partner

(512)

Ethical Dilemma 1 **Middlemist, Knowles, and Matter (1976)**

Middlemist, Knowles, and Matter (1976) conducted a study investigating whether or not invasions of personal space are physiologically, as well as psychologically, arousing. The experiment was run in a men's lavatory in which the investigators closed off one or another urinal. Participants were forced to urinate either in the urinal next to a male confederate of the experimenter or in the urinal one away from the confederate. A second confederate positioned in a toilet stall adjacent to the urinals observed the subjects via periscope and recorded the latency to onset of urination and its duration.

(513)

Ethical Dilemma 1 **Middlemist, Knowles, and Matter (1976)**

- Ethical issues to consider
 - Invasion of privacy
 - Is it ethical?
 - Informed consent
 - Do you need it?
 - Debriefing
 - Should you?



(514)

Ethical Dilemma 2

Harari, Harari, and White (1985)

In a field experiment conducted by Harari, Harari, and White (1985), male participants walking alone or in groups were exposed to a simulated rape. As participants walked along a path to a parking lot, a male and a female confederate of the experimenters acted out a simulated rape. (The male grabbed the female around the waist, put his hand over her mouth, and dragged her into some bushes. The female screamed for help.) Observers stationed at various points recorded the number of participants who offered help. Prior to actual intervention, participants were stopped and told of the true purpose of the experiment. The results showed that 85 percent of the participants walking in groups offered help to the victim, whereas 65 percent of the participants walking alone offered help. Support was found for the idea that individuals in groups are as likely, if not more likely, to help when the victim is clearly in need of help.

(515)

Ethical Dilemma 2

Harari, Harari, and White (1985)

- Ethical Issues to Consider
 - Informed consent
 - Is it ethical not to get it in advance?
 - Impact on participants
 - What might it be?
 - Justifiable? Reversible? Experimenter's obligations?
 - Debriefing
 - How to do it?
 - Unpredictability of participant responses
 - What if one of them had a gun?

(516)

APA Ethical Guidelines

- General Principles:
 - A. Beneficence and nonmaleficence – do no harm
 - B. Fidelity and Responsibility – establish trust, be professional
 - C. Integrity – don't falsify
 - D. Justice – be fair to all participants
 - E. Respect for people's rights and dignity – special care for vulnerable populations

[517]

APA Ethical Guidelines

- Human Relations
 - A. Avoiding Harm
 - B. Informed consent must be clear
 - C. Assent of participants not able to fully consent themselves
 - Privacy and Confidentiality
 - A. Maintain confidentiality
 - B. Explain limits of confidentiality
 - C. Recording permission

[518]

APA Ethical Guidelines

- Research and Publication - contents of consent:
 - A. Purpose of research
 - B. Right to decline/withdraw
 - C. Risks and Benefits
 - D. Incentive
 - E. Contact info

[519]

APA Ethical Guidelines

- Client/patient student and subordinate research participation
 - A. Protect participants from penalties for declining/withdrawing
 - B. If part of a course alternative must be offered
- Dispensing with informed consent
 - A. When there is no expectation of harm AND one of the following
 - B. Anonymous questionnaires, archival material, naturalistic observation
 - C. Evaluation of normal educational practices
 - D. Organizational related factors

[520]

APA Ethical Guidelines

- Offering Inducement
- A. Avoid coercing participation
- Deception in Research
- A. Research must have significant value
- B. No severe pain or emotional distress
- C. Integral to the design (no other way to test this)

Debriefing

- A. Explain and correct misperceptions
- B. Counteract any harm
- C. If withholding is necessary take steps to minimize harm

[521]

Problems with Deception

- Ethical issues
 - Self Esteem
 - Being duped
 - Finding out something negative
 - Failure of real informed consent
- Bias issues
 - Tainting the participants for future studies
 - Deceived participants act differently

[522]

Ethical Issues From Student Projects

- Priming anxiety – example from previous year
- Asking about mental health in general
- Asking about sexuality
- Other issues?

[523]

THE END

[524]