

Review: 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 (e.g., COO should be capable of similar interpretation in different countries)
 - **Discriminant validity** is the extent to which a measure does not correlate with other constructs from which it is supposed to differ (e.g., browsing skills of Centaur consumer are different)
 - **Nomological 法則 validity** is the extent to which the scale correlates in theoretically predicted ways with measures of different but related constructs (e.g., Centaur are risk-prone).

認識論(Epistemology) Measurement Postulates

- (Postulates is an assumption or prerequisites to carrying out some operation.)
- Three postulates of measurement (Coombs 1953):
 1. Either $(a = b)$ or $(a \neq b)$, but not both
 2. If $[(a = b) \text{ and } (b = c)]$, then $(a = c)$
 3. If $[(a > b) \text{ and } (b > c)]$, then $(a > c)$
 1. Classification
 2. Equality in comparison
 3. Transitivity 操作傳遞 (trouble in Ψ such as love, like, a friend of, accept, preference, loyalty, trust, and relationship)

Explanation (解釋)

10-3

Propositional Logic (句子演算) Realism

WIKIPEDIA
The Free Encyclopedia

- Realism 實在論:為希臘哲學家亞里士多德所提出的創見，主要是否定其師柏拉圖提出的真知只存在於觀念世界的說法，主張知識可經由感官經驗而得自於現象世界。知識的由來是與環境的互動而得到。(現實主義, 寫實主義)
- whatever we believe now is only an approximation of reality and that every new observation brings us closer to understanding reality.
- our reality is ontologically independent of our conceptual schemes, perceptions, linguistic practices, beliefs, etc. Realism may be



3

Chapter Ten

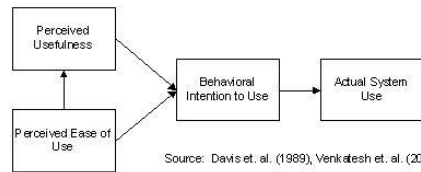
Questionnaire Design





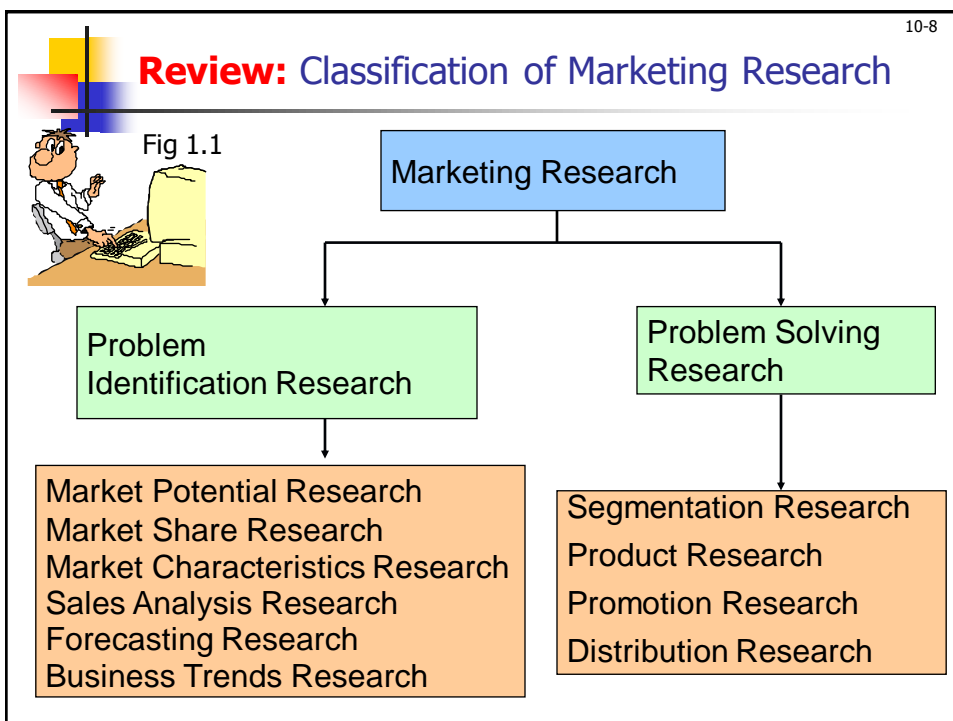
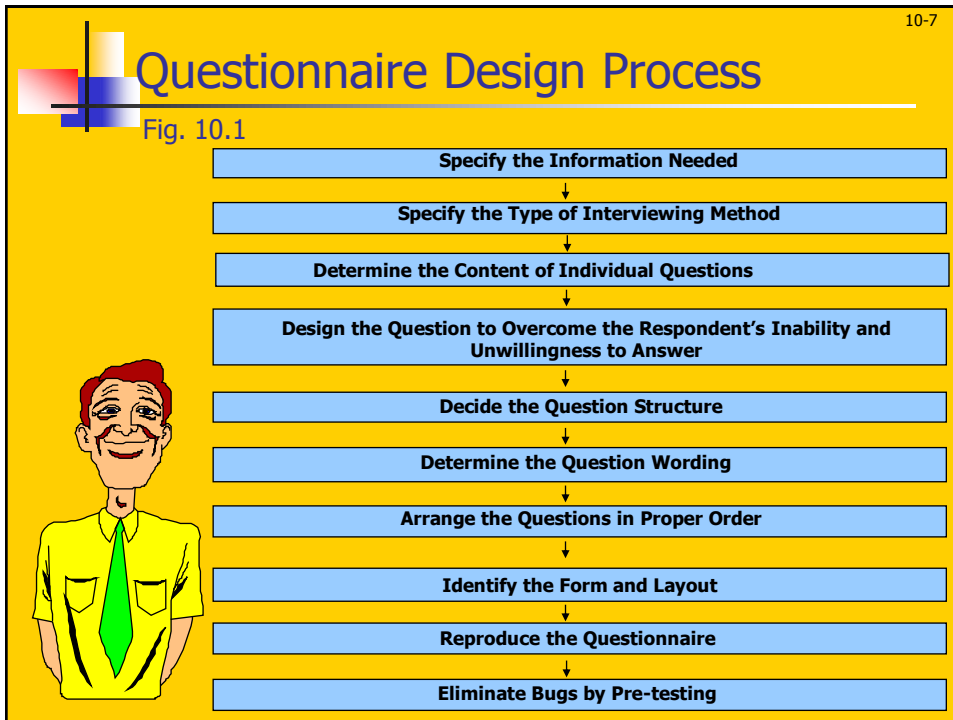
Questionnaire Definition

- 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.





Review: Problem Solving Research¹⁰⁻⁹

(Marketing Management Issues)

Table 1.1

<h4 style="color: blue; text-decoration: underline;">SEGMENTATION RESEARCH</h4> <ul style="list-style-type: none"> Determine the basis of segmentation Establish market potential and responsiveness for various segments Select target markets Create lifestyle profiles: demography, media, and product image characteristics 	 <h4 style="color: blue; text-decoration: underline;">PRODUCT RESEARCH</h4> <ul style="list-style-type: none"> Test concept Determine optimal product design Package tests Product modification Brand positioning and repositioning Test marketing Control score tests
---	--



Review: Problem Solving Research¹⁰⁻¹⁰

(Marketing Management Issues)

Table 1.1 cont.

<p style="color: red; font-size: 1.2em; font-weight: bold;">0.00% APR</p>  <h4 style="color: blue; text-decoration: underline;">PRICING RESEARCH</h4> <ul style="list-style-type: none"> Pricing policies Importance of price in brand selection Product line pricing Price elasticity of demand Initiating and responding to price changes 	<h4 style="color: blue; text-decoration: underline;">PROMOTIONAL RESEARCH</h4> <ul style="list-style-type: none"> Optimal promotional budget Sales promotion relationship Optimal promotional mix Copy decisions Media decisions Creative advertising testing Evaluation of advertising effectiveness Claim substantiation <div style="border: 1px solid black; padding: 10px; display: inline-block; margin-top: 20px;"> \$ SALE </div>
---	--



Review: Problem Solving Research¹⁰⁻¹¹

(Marketing Management Issues)


Table 1.1 cont.

DISTRIBUTION RESEARCH

Determine...

- Types of distribution
- Attitudes of channel members
- Intensity of wholesale & resale coverage
- Channel margins
- Location of retail and wholesale outlets



Review: Concept, construct, and variable¹⁰⁻¹²

(marketing research & consumer behavior issues)

- Marketing Constructs: psychological involvement, brand loyalty, AIDA, nostalgia, gratification, information overload, attribution theory, Hotelling competition, prospect theory
 - Construct: “defined and specified [so] that it can be measured and observed”
- Marketing Variables: sales, advertising, new product adoption rate, retail productivity, awareness, preference, AIDA, satisfaction, clicks, CPM, brand loyalty, etc.
- Measures and Scales: consumer choice and preference, MDS, nostalgia, CSR, virtual reality, competitive advantage, management/financial performance, KSF, KPI, risks, value, “green” marketing, UGC, consumer journey, WOM, Viral marketing, social networks, structural holes



Specify the Information Needed

■ The Introduction

- It must be persuasive
- It must qualify the respondents

■ The Body or Content (variables and constructs)

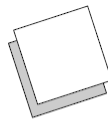
- Facts (management issues)
- Knowledge (prior info on consumer/market behavior)
- Opinions, attitudes, motives, values (research issues)
- Future behavior (spatial and time issues)

■ Basic Data

- 1) Demographics/Socioeconomic Characters
- 2) Psychographics or VALS (value & life style)
- 3) Media habit (O2O2M2S, UGC, WOM)
- 4) Product usage, consumption behavior (H,M,L)

Running with Centaur

Journal of Consumer marketing (2003)



Typology of online shoppers

Ah Keng Kau

Professor, NUS Business School, National University of Singapore, Singapore

Yingchan E. Tang

National Chiao Tung University, Taiwan

Sanjoy Ghose

University of Wisconsin – Milwaukee, Milwaukee, Wisconsin, USA

Keywords *Internet, Purchasing, Market segmentation, Information retrieval*



A. Digital marketing environment (Contextual)

- DME has caused the paradigm shift of marketing 4Ps to 4Ds:
 - Digitalized product, branding, and service;
 - Disintermediation (and re-intermediation) of channel distribution;
 - Diversified marketing communication vehicles (internet, Wi-Fi, m-commerce, GSP, P2P);
 - Direct pricing (full information, online bid)



B. Information Processing (psychological framework)

- information **acquisition skill** is essential to consumer in gathering all the marketing information under DME (Hoffman and Novak 1996).
- There is a degree of **expertise** in browsing the Web. The information acquisition expertise might influence the online shopping behavior (Novak, Hoffman, and Yung 2000)
- Skillful navigating experience reduced likelihood of **experiencing flow** (Johnson, Lohse, and Mandel 1999, Dholakkia and Bagozzi 2001)



C. Telesthesia utility (prior/contingency factor)

- Online **Security**: online payment
- Online marketing support: **warrantee**, product quality
- **Service** support: purchase agency vs. real salesperson
- Online **information search**: Money saving and time saving, usefulness of banner ad



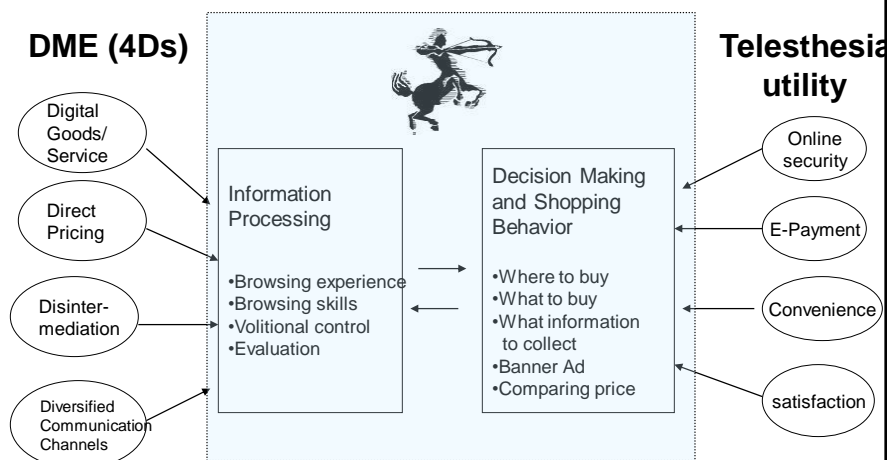
D. e-Shopping Behavior (usage/consequences)

- Sheehan (1999): Gender difference
- Chen and Wells (1999): entertainment, informativeness, and organization
- Donthu and Garcia (1999): demographics, convenience, innovative, impulse, brand and conscious, and attitude
- WWW.ZDNet: Simplifiers, Surfers, Connectors, Bargain Shoppers, Routine Followers, and Sportsers
- Kau, Tang and Ghosh (2003): on-off shoppers, comparative shoppers, traditional shoppers, dual shoppers, e-laggards, and information surfer.

D. e-Shopping Behavior

- Strategic Scenarios: degree of interaction with store vs. price, efficiency, system operator in control (Clemons and Bradley 2001)
 - Five scenarios: Looking for Mr. Goodstore, Sears is My Shepherd, The Electronic Parimutuel Mall, Wegmans online, and Bill's wonderful Adventure
- New Hybrid Consumer: the Centaurs (Wind and Mahajan 2001)

I. Centaur Shopping Behavior Model



Section I

Please indicate how much you agree or disagree with the following statements by clicking on the number that best reflects your opinion.

I. Information search behavior

1. When considering purchase online, I actively evaluate a larger number of brands than when I shop in a traditional retail environment.

Strongly Disagree ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Strongly Agree

2. When considering purchase online, I tend to tradeoff the strengths and weaknesses of a brand before deciding how attractive the brand is.

Strongly Disagree ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Strongly Agree

3. When planning purchase online, I only consider the single most important feature (e.g. best price or best quality) when selecting the brand.

Strongly Disagree ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Strongly Agree

4. When considering purchase online, I will reject a brand if it does not meet my minimum acceptable level on each important product feature.

Strongly Disagree ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Strongly Agree

5. When considering purchase online, compared to traditional retail environment, I find it easier to identify and eliminate brands which are clearly inferior to others.

Strongly Disagree ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Strongly Agree

6. When considering purchase online, I do an overall comparison of different brands before I decide which brand is my most preferred.

Strongly Disagree ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Strongly Agree

Section II

Please answer the following questions.

II. Online buying behavior

1. Have you bought any of the following products/services online during the last 12 months? Please check as many as applicable. Please also indicate the amount you spent (in Sing\$) in your most recent purchase.

<input type="checkbox"/> Computer software	S\$ <input type="text"/>
<input type="checkbox"/> Computer hardware/parts	S\$ <input type="text"/>
<input type="checkbox"/> Clothing and accessory	S\$ <input type="text"/>
<input type="checkbox"/> Tickets for airlines/railway/ship, etc	S\$ <input type="text"/>
<input type="checkbox"/> Tickets for concerts/cinema shows	S\$ <input type="text"/>
<input type="checkbox"/> Stationery	S\$ <input type="text"/>
<input type="checkbox"/> Records/CDs/music	S\$ <input type="text"/>
<input type="checkbox"/> Sporting goods	S\$ <input type="text"/>
<input type="checkbox"/> Books/magazines	S\$ <input type="text"/>
<input type="checkbox"/> Medicine	S\$ <input type="text"/>
<input type="checkbox"/> Electrical/electronic goods	S\$ <input type="text"/>
<input type="checkbox"/> Others, please specify <input type="text"/>	S\$ <input type="text"/>

2. Were you satisfied with the products/ services you bought online?

Very Dissatisfied ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Very Satisfied ☐ Did not buy

3. Do you intend to continue buying online in the next 6 months?

Very Unlikely ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Very Likely

Section III:

The following statements are about general attitudes toward usage of Internet shopping. Please indicate the extent to which you agree or disagree with each statement by clicking on the appropriate options.

III. Telesthesia utility

- 1. Buying online lacks the feel of security**
Strongly Disagree ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Strongly Agree
- 2. You never know if you are getting the intended products when you buy online.**
Strongly Disagree ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Strongly Agree
- 3. Buying online saves time.**
Strongly Disagree ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Strongly Agree
- 4. Buying online saves me money.**
Strongly Disagree ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Strongly Agree
- 5. It is convenient to buy online..**
Strongly Disagree ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Strongly Agree
- 6. Whenever I need to find information, my first choice is to go to the Internet.**
Strongly Disagree ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Strongly Agree
- 7. Web advertisements provide me with useful information for my online shopping.**
Strongly Disagree ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Strongly Agree
- 8. I like to go shopping in traditional shopping malls whenever I have free time.**
Strongly Disagree ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Strongly Agree
- 9. It is difficult to locate products on the Web.**
Strongly Disagree ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Strongly Agree
- 10. I like to buy online because there is no pressure from salespersons.**
Strongly Disagree ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Strongly Agree

Done Internet

Section IV

Please provide some information about yourself by answering each of the following questions. The information gathered will be kept strictly confidential.

IV. Demographics

- 1. Your Gender:** ☐ Male ☐ Female
- 2. Your age range:**
- 3. Your marital status:**
- 4. Your educational level:**
- 5. Your current occupation:**
- 6. Your approximate annual household income:** S\$ For people who are not residing in Singapore, please approximate using the rate conversion: US\$1.00 to S\$1.70
- 7. Your country of residence:**

Please provide us with your mailing address so that we are able to contact you should you win a prize in our lucky draw.

- 8. Mailing address:**
- 9. Email address:**

Done Internet



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. Lord & Taylor	_____
2. Macy's	_____
.	_____
.	_____
10. Wal-Mart	_____



Example on Mail Survey

Best Paper Award

Yu-Ying Yang, Min-Hua Wu and Min-Hui Kao, "Ecologically Conscious Consumer Behaviour and the Adoption of Green Electricity: A Perspective Incorporating Consumer Innovativeness Characteristics", *International Conference on the Development and Application of Big Data and Enterprise Resource Management*, 2018 (ICBDERM), Hualien, Taiwan, 17-18 March, 2018.

Ecologically Conscious Consumer Behavior and the Adoption of Green Electricity: from the Perspective of Innovativeness Characteristics

Ying-Chan Tang*, Min-Hua Wu, and Min-Hui Kao

Institute of Business and Management, National Chiao Tung University, Taiwan. 118, Zhongxiao W. Rd., Taipei 100, Taiwan (ROC)

Abstract

The intensity of public concern about global warming emissions has spiked. One practical solution is to implement "green electricity" such that the power is generated by more environmentally sustainable means. Past literature in the area of ecologically conscious consumer behavior (ECCB) provides the greatest insights on how the consumers can adopt the environmental friendly products and who would pay for



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. Lord & Taylor	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



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?

10-29

- 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?

10-30

- 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 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 (e.g., tax or political party tendency)
- 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 (e.g., Income)



Choosing Question Structure

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

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

- **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

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

- Scales were discussed in detail in Chapters 8 and 9:

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

10-39

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

10-40


- 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)

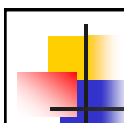


Choosing Question Wording
Avoid Implicit Alternatives

10-41

- 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

10-42

- 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

10-43

“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)



Determining the Order of Questions

10-44

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

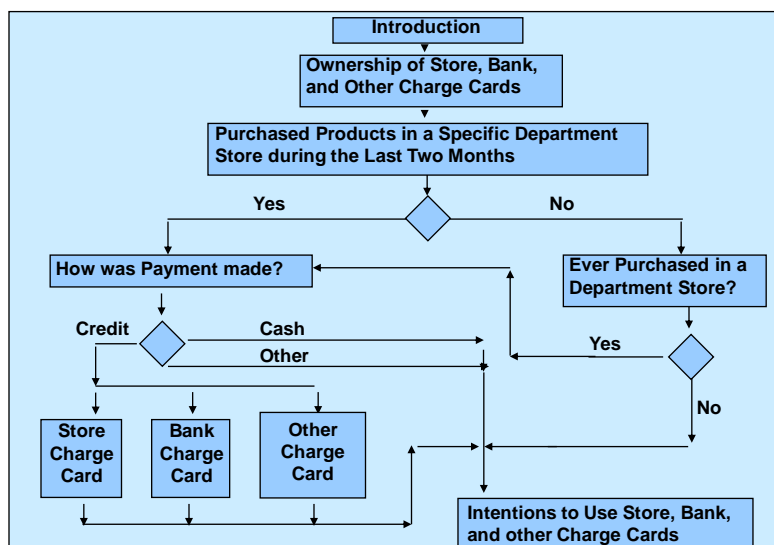
The following guidelines should be followed for branching questions:

- 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

Fig. 10.2





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.



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

Table 10.1

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



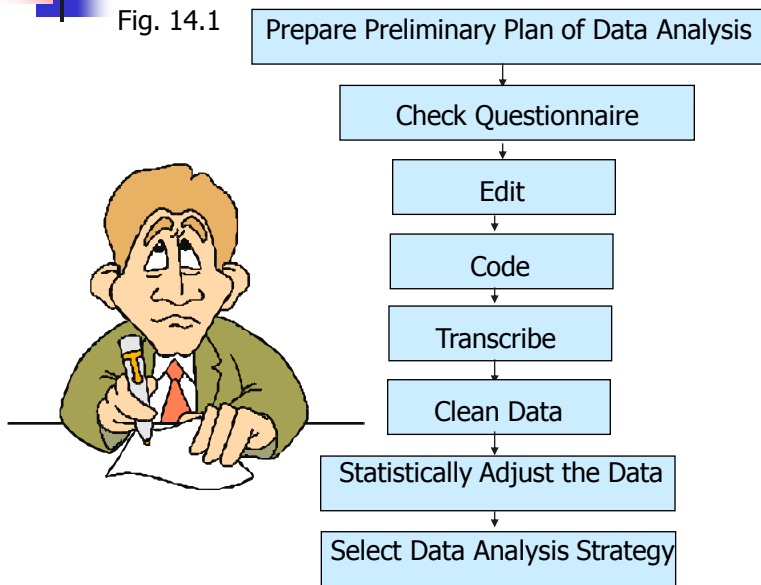
Chapter Fourteen

Data Preparation



Data Preparation Process

Fig. 14.1



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

10-55

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

10-56

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

10-57

Age Group

	Code
Under 15	1
15-24	2
25-34	3
35-44	4
45-54	5
55-64	6
65+	7
Refused	8

Transportation Mode

Bicycle	1
Bus	2
Taxi	3
Subway	4
On foot	5
Private car	6
Company car	7
Motor cycle	8
Small cab/ cart	9
Never been to there	99



Coding

10-58

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:

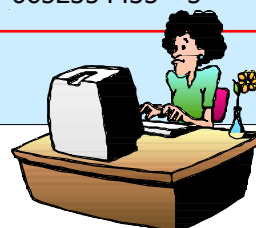
- column number
- record number
- variable number
- variable name
- question number
- instructions for coding



An Illustrative Computer File

Table 14.1

OBS	Fields Column Numbers								
	1-3	4	5-6	7-8	...	26	...	35	77
1	1	001	1	31	01			6544234553	5
2	11	002	1	31	01			5564435433	4
3	21	003	1	31	01			4655243324	4
4	31	004	1	31	01			5463244645	6
5	2701	271	1	31	55			6652354435	5





Data Cleaning Consistency Checks

10-61

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 Consistency Checks

10-62

Purchase of Fashion Clothing, Y

Gender, X

	Yes	No
Male		
Female		



Statistically Adjusting the Data Variable Respecification

10-63

- **Variable respecification** involves the transformation of data to create new variables or modify existing variables (e.g., standardize to z score)
- E.G., the researcher may create new variables that are composites of several other variables. (productivity = output/input)
- 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

10-64

Table 14.2

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

10-65

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:

$$Z_i = (X_i - \bar{X})/s_x$$



Review: Data Transformation and Analysis

10-66

MP3 Data Variables^o

1. **Retailer Type:** Dummy variable for superstore (e.g., Carrefour), 3Cs chain store (e.g., Tsan Kuen), specialty store (service station), and others^o
2. **National Brand** vs. OEM product: ^o
3. **Product Variants:** model ID, package (color), memory size^o
4. **Manufacturer price, Retail price, and gross margin**^o
5. **Dollar gross margin:** net selling price – cost from vendor^o
6.^o

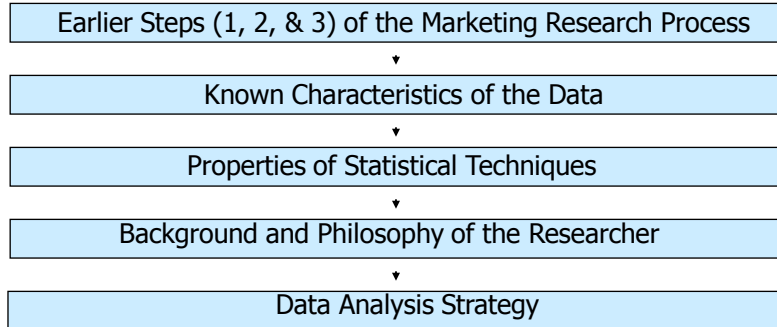
Data Analysis^o

1. **Descriptive statistics:** frequency, %, mean, medium, skewness and kurtosis^o
2. **Cross tabulation:** V1 vs (V5, V6, V7, V9); V10 vs (V5, V6, V9); V15 vs. (V5, V6, V7)^o
3. I.V. = Mean value of national brand whose market share is <5%, <35%, >35%^o
D.V. = V5, V6, V7, V9^o
4. **Regression Model:** e.g., Margin = V1, V9, V10, V11, V15, ^o
5. **Logistic Regression:** DV= V2 or V15 ^o
- 6 **HLM (Random coefficients),** V2 and V15 as the nested variables^o



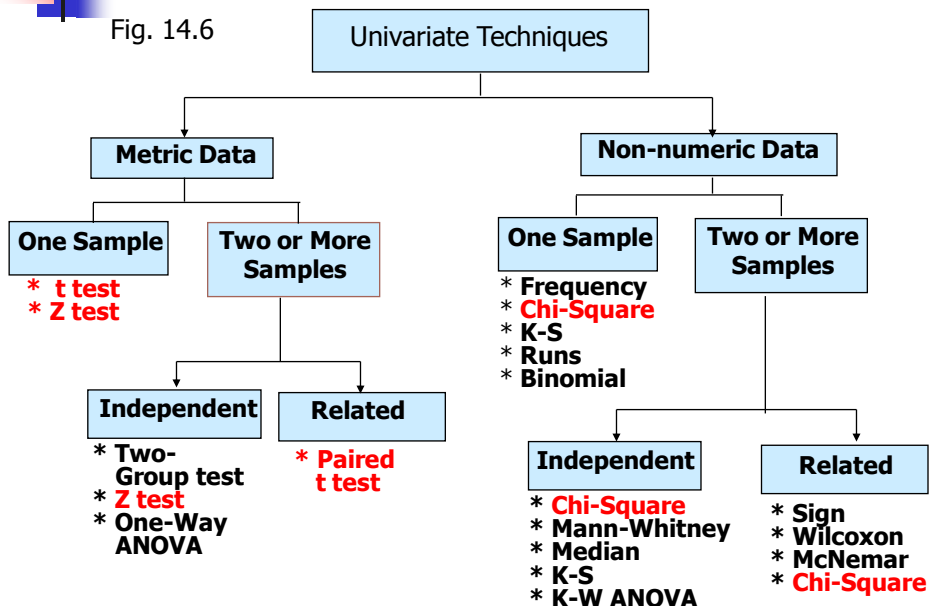
Selecting a Data Analysis Strategy

Fig. 14.5



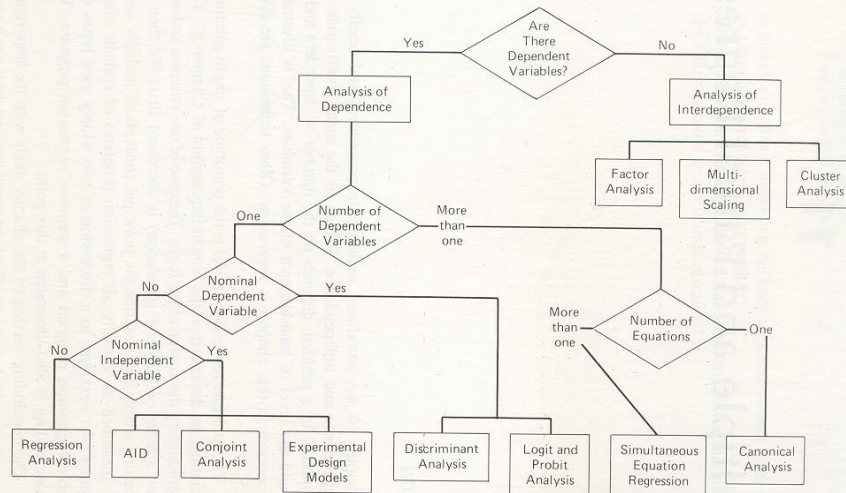
A Classification of Univariate Techniques

Fig. 14.6



A Classification of Multivariate Techniques

Figure 11-1. Multivariate Analysis



Source: D. A. Aaker, "Multivariate Analysis in Marketing," in *Multivariate Analysis in Marketing*, 2nd ed. (Palo Alto, Calif.: Scientific Press, 1981), p. 2.

Looking for info about statistics?

We wrote the book on it.
And you can read it for free!

- Elementary Concepts
- Statistics Glossary
- Basic Statistics
- ANOVA / MANOVA
- Association Rules
- Boosting Trees
- Canonical Analysis
- CHAID Analysis
- C & R Trees
- Classification Trees
- Cluster Analysis
- Correspondence Analysis
- Data Mining Techniques
- Discriminant Analysis
- Distribution Fitting
- Experimental Design
- Factor Analysis
- General Discrim. Analysis
- General Linear Models
- Generalized Additive Mod.
- Generalized Linear Mod.
- General Regression Mod.
- Graphical Techniques

Getting Started with Statistics Concepts

In this introduction, we will briefly discuss those elementary statistical concepts that provide the necessary foundations for more specialized expertise in any area of statistical data analysis. The selected topics illustrate the basic assumptions of most statistical methods and/or have been demonstrated in research to be necessary components of our general understanding of the "quantitative nature" of reality (Nisbett, et al., 1987). We will focus mostly on the functional aspects of the concepts discussed and the presentation will be very short. Further information on each of the concepts can be found in statistical textbooks. Recommended introductory textbooks are: Kachigan (1986), and Runyon and Haber (1976); for a more advanced discussion of elementary theory and assumptions of statistics, see the classic books by Hays (1988), and Kendall and Stuart (1979).

- What are Variables?
- Correlational vs. Experimental Research
- Dependent vs. Independent Variables
- Measurement Scales
- Relations between Variables
- Why Relations between Variables are Important
- Two Basic Features of Every Relation
- Why Significance of a Relation between Variables Depends on the Size of the Sample
- Example: Baby Boys to Baby Girls Ratio
- Why Small Relations can be Proven Significant Only in Large Samples
- Can "No Relation" be a Significant Result?
- How to Measure the Magnitude (Strength) of Relations between Variables
- Common "General Format" of Most Statistical Tests



Overview: FREQ Procedure

The FREQ procedure produces one-way to n -way frequency and contingency (crosstabulation) tables. For two-way tables, PROC FREQ computes tests and measures of association. For n -way tables, PROC FREQ provides stratified analysis by computing statistics across, as well as within, strata.

For one-way frequency tables, PROC FREQ computes goodness-of-fit tests for equal proportions or specified null proportions. For one-way tables, PROC FREQ also provides confidence limits and tests for binomial proportions, including tests for noninferiority and equivalence.

For contingency tables, PROC FREQ can compute various statistics to examine the relationships between two classification variables. For some pairs of variables, you might want to examine the existence or strength of any association between the variables. To determine if an association exists, chi-square tests are computed. To estimate the strength of an association, PROC FREQ computes measures of association that tend to be close to zero when there is no association and close to the maximum (or minimum) value when there is perfect association. The statistics for contingency tables include the following:

Table 35.1 Summer Enrichment Data

Gender	Internship	Enrollment		
		Yes	No	Total
boys	yes	35	29	64
boys	no	14	27	41
girls	yes	32	10	42
girls	no	53	23	76

```
proc freq data=SummerSchool order=data;
  tables Internship*Enrollment / chisq;
  weight Count;
run;
```



Overview: TTEST Procedure

The TTEST procedure performs t tests and computes confidence limits for one sample, paired observations, two independent samples, and the AB/BA crossover design. Two-sided, TOST (two one-sided test) equivalence, and upper and lower one-sided hypotheses are supported for means, mean differences, and mean ratios for either normal or lognormal data.

Table 92.1 summarizes the designs, analysis criteria, hypotheses, and distributional assumptions supported in the TTEST procedure, along with the syntax used to specify them.

Table 92.1 Features Supported in the TTEST Procedure

Feature	Syntax
Design	
One-sample	VAR statement
Paired	PAIRED statement
Two-independent-sample	CLASS statement, VAR statement
AB/BA crossover	VAR / CROSSOVER

```
data time;
  input time @@;
  datalines;
  43 90 84 87 116 95 86 99 93 92
  121 71 66 98 79 102 60 112 105 98
  ;
run;
```

The only variable in the data set, time, is assumed to be normally distributed. The trailing at signs (@@) indicate that there is more than one observation on a line. The following statements invoke PROC TTEST for a one-sample t test:

```
ods graphics on;

proc ttest h0=80 plots(showh0) sides=u alpha=0.1;
  var time;
run;
```



Chapter 3

The UNIVARIATE Procedure

Overview

10-73

The UNIVARIATE procedure provides the following:

- descriptive statistics based on moments (including skewness and kurtosis), quantiles or percentiles (such as the median), frequency tables, and extreme values
- histograms and comparative histograms. Optionally, these can be fitted with probability density curves for various distributions and with kernel density estimates.
- quantile-quantile plots (Q-Q plots) and probability plots. These plots facilitate the comparison of a data distribution with various theoretical distributions.
- goodness-of-fit tests for a variety of distributions including the normal
- the ability to inset summary statistics on plots produced on a graphics device
- the ability to analyze data sets with a frequency variable
- the ability to create output data sets containing summary statistics, histogram intervals, and parameters of fitted curves

```
proc univariate data=homeloan.normaltest;  
  var LoanToValueRatio;  
run;
```

Chapter 52

The NPAR1WAY Procedure

10-74

Overview

The NPAR1WAY procedure performs nonparametric tests for location and scale differences across a one-way classification. PROC NPAR1WAY also provides a standard analysis of variance on the raw data and tests based on the empirical distribution function.

```
proc npar1way data=Gossypol;  
  class Dose;  
  var Gain;  
run;
```

The variable `Dose` is the CLASS variable, and the VAR statement specifies the variable `Gain` is the response variable. The CLASS statement is required, and you must name only one CLASS variable. You may name one or more analysis variables in the VAR statement. If you omit the VAR statement, PROC NPAR1WAY analyzes all numeric variables in the data set except for the CLASS variable, the FREQ variable, and the BY variables.

Since no analysis options are specified in the PROC NPAR1WAY statement, the ANOVA, WILCOXON, MEDIAN, VW, SAVAGE, and EDF options are invoked by default. The following tables show the results of these analyses.