Harvard Business School

9-592-034

October 8, 1991

Market Research

I. Introduction

Marketing research is an integral part of marketing management. Its function is to provide *managers* with information that assists in making decisions. The manager must be aware of the costs and benefits of alternative methods in order to select the best research program and analyze the results in the most useful way. This note presents an overview of the most widely utilized methods, which are quite diverse. For example, consider the brand manager for Quaker Oats' Life Cereal.[1] Seeking ways to increase Life's share of the Ready-to-Eat (RTE) cereal market, her review of the situation showed:

- The impetus to the development of Life had been a *motivation research study* showing that a crisp, strongly nutritional, tan-gold color cereal could appeal to all members of the family.

- After development of the product, Quaker *test marketed* Life in three different textures. The shredded biscuit form won a higher repeat purchase rate and overall share. Advertising weight tests compared a $4.1 million national equivalent budget to a $5.7 million budget. Test market shares were 2.5% and 2.6%, respectively.

- In the *test markets*, Quaker also tested alternative advertising copy: a "Most Useful Protein" campaign and "Fun of Life" campaign. Both generated less-than-hoped-for awareness and brand trial. "Most Useful Protein" generated better repeat purchase.

- Shortly after the national introduction, a *survey* profiled the Life-purchasing family as having more children and higher incomes than average.

- In the first year of national introduction, a *telephone survey* of 2,515 households using RTE cereals gathered data on the awareness, trial, and repeat purchase intention of Life and major competitors. In Year 1, 40% of households were

[1]This description is adapted from the "Quaker Oats Company: Life Cereal," HBS No. 513-157.

Professor Robert J. Dolan prepared this note as the basis for class discussion. It is partially based on two earlier notes, "Marketing Research," HBS No. 585-039 and "Research Methods in Marketing: Survey Research," HBS No. 582-055.

aware of Life. Thirteen percent of those aware purchased Life. In Years 3 and 5 after the introduction, 1,692 households were surveyed. Awareness had increased to 61% by Year 5, and 33% of those aware had purchased Life. By comparison, 97% were aware of Kellogg Corn Flakes and 96% of those aware had purchased Corn Flakes. The survey also determined that nutrition was the primary reason to buy for only 13% of Life's buyers.

- In Year 3, a new ad campaign based on a "growing" theme was tested via an *on-the-air* test. An advertising testing service conducted telephone interviews the day after the ad appeared. Only 5% of consumers remembered seeing the ad, compared to a 25% norm for the product category.

- Later an alternative campaign was tested. The *redemption rate* of a coupon in the ad served as a measure of effectiveness for the print media. To judge effectiveness of the TV ad, a *focus group* was held in one metropolitan area. Unfavorable reaction was found via both tests.

- In Year 5 after the introduction, a new advertising agency had taken over the Life account and conducted its own *survey research* to help in its positioning of Life. Heavy-user families were identified and targeted.

- A *consumer panel study* purchased from the Market Research Corporation of America showed consumers of RTE cereals had no clearly distinguishable characteristics.

- A company *attitudinal study* contrasted hot and RTE cereals. RTEs were seen as more tasty but less nutritious than hot cereals.

- *Industry sources* indicated that Special K had spent $4.1 million in media in the last year. Life's advertising agency was proposing a $3.1 million expenditure for the upcoming year.

The eight major research methods used by Life are shown in **Table A**. They differ in the following respects:

1. ***Primary data vs. Secondary Data***—in primary (1, 2, 3, 4, 5, 7) the data were collected solely to support Life cereal's marketing whereas in secondary (6, 8) the data were collected not specifically for Life. MRCA's Panel #6 is an example of a syndicated service which collects and formats for sale to a number of clients.

2. ***Experimental vs. Non-Experimental***—in the test markets (#2) Quaker manipulated the environment, e.g., by using different advertising spending levels in regions of the country whereas most other methods generated data in the normal course of business conduct.

3. ***Qualitative vs. Quantitative***—the surveys (#3) used large sample sizes and summarized results in a quantitative fashion as compared to the focus group (#7) which utilized only 10 people and made no attempt to quantify the feelings people expressed in their own words.

4. ***Decision Supported***—Life managers used the data to:

   a. *Identify* marketing problems, e.g., low share levels.

b. *Diagnose* marketing problems, e.g., by showing that the low share is due to low awareness and trial levels rather than high trial but low repeat rates.

c. *Evaluate* alternative marketing strategies, e.g., in nutrition or fun positioning.

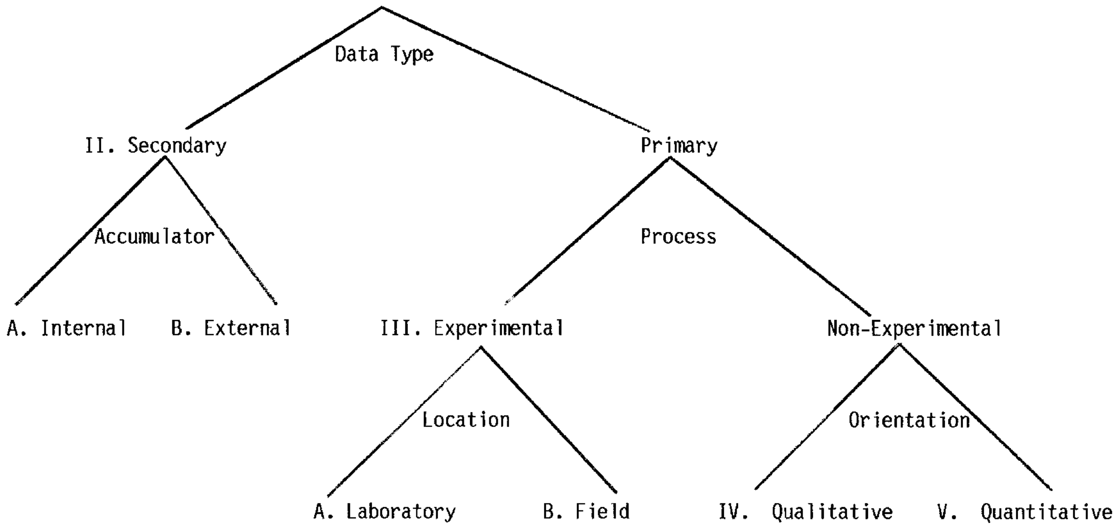
d. *Monitor* subsequent performance and indicate any need for changes in strategy, e.g., when only 13% of buyers say the reason to buy is nutrition.

Table A Market Research Data for Life Cereal

Research Procedure	Data	Decision Assisted
1. Motivation research	Consumer preferences	Product design: Ingredients Color Basic positioning
2. Test market experiment	Share Awareness Trial Repeat rate	Go/no-go national Product form Advertising budget Advertising copy
3. Surveys	Profiles of: Life buyers Heavy RTE users Life and competition: Awareness Trial Repeat Consumers' "reason to buy" Consumer attitudes on RTEs versus hot cereals	Performance monitoring Advertising copy
4. On-the-air test	Advertising recall	Advertising copy
5. Coupon placement test	Redemption rate	Advertising copy
6. Panel data	Demographic profiles	Product positioning
7. Focus group	Reaction to advertisement	Advertising copy
8. Public/industry data	Competitive advertising expenditure	Advertising budget

This note provides an overview of the major research procedures. It also provides references for each major topic which one can consult to develop a deeper understanding of individual areas. **Table B** presents a taxonomy of data collection efforts. The roman numerals in **Table B** correspond to the section of this note in which the particular method is discussed, i.e., we begin by discussing Secondary Data in Section II. Section III covers experiments, both laboratory and field. Sections IV and V both deal with non-experimental situations separating out methods based on the emphasis on qualitative vs. quantitative data.

Table B Data Collection Methods Taxonomy



II. Secondary Data

For secondary data, the current issue is not the major driving force behind the data collection effort. For example, the U.S. Census is a widely used secondary source. As shown in **Table B**, there are two major types of secondary data: that held internal to the firm (e.g., company sales records) and that gathered external to the firm, e.g., by the Census Bureau or a syndicated research supplier such as Nielsen or Information Resources. Secondary data offer potential advantage over primary data in their timely availability and, especially in the case of internal data, their cost.

When an issue arises, the first question should be: can it be resolved with data that have already been collected and are in the company's possession? Accounting records are the most intensively used source of internal, secondary data, e.g., cost of goods sold, unit sales broken down by product, region, or customer, and various expenditures by time period. A second important source is salesperson call reports covering issues such as account potential, account penetration, calls per account, and competitive activity. Finally, there are miscellaneous market studies, databases, and analyses purchased by other parts of the organization. The lack of an efficient mechanism for locating a relevant report outside one's own department can be very costly in terms of time and fees paid.[2]

External data relevant to market decision making are collected by a wide variety of institutions. The principal distinction between these organizations is whether they are firms whose primary business is the compilation and selling of marketing databases or not. The four largest market research firms in the United States (A.C. Nielsen, IMS, Information Resources, and Arbitron) all generate a significant portion of their revenues from the compilation and sale of statistics such as product sales, media exposure, competitive marketing activity and characteristics of buyers. (See the "The Marketing Information Industry," No. 588-027 for details of these firms' products.) Relevant external data can also be obtained from various trade associations, the U.S. government, and international organizations. (A list of sources is provided in "Note on External Sources of Marketing Data," No. 580-107.)

[2]See "The Plugged-In Marketers at General Foods," *Marketing Communications*, March 1984, pp. C9-C13, for a description of a system for companywide use of information.

While secondary data potentially offer some advantage, they must be carefully checked to assess their accuracy and relevance. The manager must understand the data collection methods precisely. While firms such as those mentioned above have very good reputations, some "research reports" available for sale are not reliable. If one does not understand how the data were collected, it is easy to place more reliance on the data than its likely precision warrants. Second, one must check for relevance, especially how "fresh" is the information. For example, is a media habits survey of decaffeinated coffee buyers done in 1989 relevant for media buying in 1992?

III. Experiments

The primary distinguishing feature of experiments is that the researcher manipulates the environment in order to measure the impact. Experiments have been used to assess the effect of many marketing variables, e.g., does:

1. Training in the use of computers for sales call planning increase salesperson effectiveness?

2. A full-page ad have more "drawing power" than a half-page ad?

3. A large advertising budget impact consumers' repeat purchase of the product or only awareness and trial?

4. Having a rainbow package design increase sales over that obtained from a solid red color design?

5. Distributing through factory-owned outlet stores hurt the image of the brand?

6. "Improving" the product via a formula change increase sales if not supported by advertising? If supported by advertising?

7. A price increase decrease unit sales? Sales revenue?

In short, experiments have been used to test changes in every element of the marketing mix.

In the language of experimental design, the variable the researcher manipulates is called the "treatment." The other key concept is the "measurement," the observation and recording of the level of sales, consumer awareness, or whatever the variable of interest is. Typically, one makes both a "before-treatment" and "after-treatment" measurement. For example, in judging the effect of sales force adoption of a new Marketing Sales Productivity System, we might proceed as follows:

1. Measure the unit sales generated for each salesperson during July 1990-June 1991 (call this *MB*, designating *m*easurement *b*efore the treatment).

2. Train the salespeople on Marketing-Sales Productivity Systems on June 28, 29, July 2, 3, 1991 (this is the "treatment").

3. Measure the unit sales generated for July 1991-June 1992 (call this *MA*, designating *m*easurement *a*fter the treatment).

We might then say that the effect of the treatment is MA - MB. There are perils in doing this though. The underlying assumption of such an assessment would be that everything else in the environment is unchanged. What we would really like to compare to MA is not MB, but what sales would have been in July 1990-June 1991 if the system had not been put in place. If we say the effect of the treatment is MA - MB, we are in effect saying that sales in July 1991-June 1992 would have been the same as in July 1990-June 1991 if the system were not implemented. To have a relevant benchmark to compare observed results to, it is usually a good idea to have a "control group" to compare to a "test group." In our example, we could randomly assign salespeople to a test group (to receive the system) or control group (no system). The scheme can be diagrammed as follows:

		July 1990 - June 1991		**July 1991 - June 1992**
		MB_T	X	MA_T
Salesperson	Random			
		MB_c		Ma_c

MB_T = Average sales of test group before system, July 1990-June 1991.

MB_c = Average sales of control group for June 1990-June 1991. Note that MB_T and MB_c should be approximately equal due to random assignment to test control groups.

MA_T = After treatment measurement for test group, July 1991-June 1992.

Ma_c = Comparable period measurement for control group.

Now the effect of the treatment can be more properly estimated as:

$$\underbrace{(MA_T - MB_T)}_{\text{Change in test group performance}} - \underbrace{(MA_c - MB_c)}_{\text{Change in control group performance}}$$

Changes in the economy, the company's product line, competitive actions, etc., all influence the control group as well as the test group. The change in the control group is therefore to be "netted out" of the test group change to arrive at an assessment of the treatment effect.

This brief note cannot discuss all the issues of experimental design, advanced design possibilities, and data analysis procedures. These issues are covered in most marketing research textbooks.[3]

Laboratory and Field Test Markets

Market tests are of two types: laboratory tests and field tests. Laboratory tests are much less costly and time consuming as they simulate the shopping experience of the field. Most well known of these are ASSESSOR (originally developed by Management Decision Systems and now marketed by M/A/R/C) and LTM (Yankelovich, Skelley, and White). The typical procedure is for people to be intercepted at a shopping mall and prescreened for membership in the target group. After measurement of brand preference, consumers are exposed to advertising, and given a chance to purchase the brand under study. After home use, a telephone interview to measure reaction and repeat purchase. Based on these measurements, the model makes a market share or unit sales projection.[4]

---

[3]An excellent elementary discussion is given in D.S. Tull and D.I. Hawkins, *Marketing Research: Measurement and Method*, 2nd ed. (New York: Macmillan Publishing, 1980). D.T. Campbell and J.C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally, 1969), is the seminal piece in the field.

[4]For a full description of such a system and evidence on its accuracy, see G. Urban and G. Katz, "Pretest-Market Models: Validation and Managerial Implications," *Journal of Marketing Research* (August 1983): 221-234.

Field test markets are used to see how the consumer and the trade actually respond when the product (and marketing program for it) are placed on the market. The program anticipated for national launch (or perhaps several candidate programs) is tried out in a portion of the market, as a way of reducing the risk of a marketwide introduction initially. The value of test marketing is a controversial issue. It is widely used, and some maintain that it is the most realistic test which a product can be given. Others argue, however, that the test environment is frequently not representative of the environment of a national introduction because: (1) both the salespeople and the trade give attention to the tested brand which it will not receive on a nationwide basis, and (2) competitors observe the test and distort results by undertaking some atypical behavior, e.g., excessive couponing. This, together with the problems of accurately monitoring test results, has led some firms to seek other methods.

There are two major ways to do a field test market. One is the traditional "matched city pair" method where two or more cities would be used and different marketing programs implemented. Sales from the different market areas would be attributed to the different marketing programs. The second is a "within city" test. This method has been made possible via advances in computer and telecommunications technology. The most well-known of these systems is BehaviorScan offered by Information Resources, Inc. In this system, individual households within a market area can be targeted via split cable TV systems and purchase can be monitored at the individual household level via scanners at all grocery stores in the market. This system has the advantage of better producing the "all else equal" situation assumed in comparison of two different programs.[5]

IV. Non-Experimental/Qualitative

Non-experimental data are generated in the normal course of business. For example, when a new product is introduced supported by the marketing plan management believes best, consumers of various demographic types become aware of the product or don't, develop attitudes toward the product, buy it or don't, and ultimately are satisfied or not. Some of these actions are captured in the secondary data of the firm, e.g., the purchase events. However, the "reason why" is typically not captured in secondary systems and a special effort must be made to obtain that data. The major differentiating characteristic of efforts in this vein is the extent to which they rely on numerically precise measurement. This section discusses qualitative methods and the next quantitative methods.

Qualitative data are characterized by their lack of numerical measurement and statistical analysis. These studies typically involve an in-depth, if somewhat subjective, understanding of the consumer. The two main methods are:

- Individual depth interviews.
- Focus groups.

In the individual depth interview, the interviewer does not have a fixed set of questions to ask the respondent. Rather, the objective is for the respondent to talk freely and in detail about a product or feelings on issues. Interviews can last over an hour. The interviewer's role is a difficult one. He or she has to keep the discussion on the area of interest, but also allow the respondent to take control of the discussion to allow important feelings to surface. This unstructured discussion also requires considerable skill in interpretation. Even given the high cost and skill required, one can see the need for such a technique in some situations. For example, suppose we were interested in finding

[5]See "The New Magicians of Market Research," *Fortune*, July 25, 1983 for a description of the BehaviorScan system.

out why people buy expensive automobiles, such as a Lexus. We can imagine the responses we would get using a structured questionnaire with:

Why did you buy a Lexus? (Check one.)

☐ Good long-term investment.
☐ Road handling.
☐ Comfortable.
☐ Image it creates for me among friends.

The structured form would probably find few people checking off the image box because the respondents feel embarrassed about admitting such a motivation.

In structured interviews, the questionnaire is the dominant factor in the interview and is the same across all respondents. For depth interviews, the interviewer is a dominant factor. Time constraints sometimes make it impossible to use the same interviewer for all respondents. Because of this, the high cost per interview and the difficulty of interpretation, depth interviews are usually limited to exploratory, problem/opportunity definition stages.

The second procedure, focus group interviews, brings together a number of people, typically 8 to 10 individuals, for an open-ended discussion with a moderator. Group composition is controlled by the researcher. For example, focus groups on new telecommunications equipment might involve a number of separate groups such as females/high education, males/blue-collar occupation, couples/both working, couples/women at home, etc.

The moderator of the discussion attempts to have the participants engage in a free exchange of ideas on the subject of interest. Like the depth interview, its advantage over structured questionnaires lies primarily in discussion of sensitive areas. For example, the product manager for a denture adhesive commented that "it's hard to draw out of the consumer the prospect of social embarrassment. You see it come out only in focus groups."[6]

The advantage of bringing a group together rather than just an individual with an interviewer is that the group setting can be less threatening and more natural for the participants. This may stimulate spontaneous discussion and lead to areas which could not be uncovered in a series of individual interviews. Focus groups have many of the problems of depth interviews, viz. the moderator effect and interpretability of results. Consequently, the method is better suited to the generation of ideas rather than systematic analysis of well-structured alternatives. A typical focus group of ten people costs about $2,500 to run and have interpreted. The cost increases significantly with the tightness of the specification on participants, e.g., it is much more expensive to do a session with "plant engineers responsible for organic chemicals" than one with people "who play tennis."[7]

[6]See the "Vicks Toiletry Products Division: Fixodent" case in B.P. Shapiro, R.S. Dolan, and J.A. Quelch, *Marketing Management: Principles, Analysis, and Applications* (Homewood, Ill.: R.D. Irwin, 1985).

[7]See P. H. Berent, "The Depth Interview," *Journal of Advertising Research* (June 1966): 32-9, for details on depth interviews, and H.I. Abelson, "Focus Groups in Focus," *Marketing Communications* (February 1989): 58-61 for discussion of focus groups.

V. Primary/Quantitative

Probably the most commonly used tool in market research is the questionnaire. The design of a survey has four stages.

1. **Problem Statement:** What decision is to be made? What information will assist in making the decision?

2. **Questionnaire Design:** What information do we want to collect in interviews? What interview questions can get us that information from respondents? How should those questions be phrased? How are we going to contact respondents?

3. **Sampling:** Who should our respondents be? How many should we get?

4. **Data Analysis:** How do we tabulate, summarize, and draw inferences from our data?

Stage 1: Problem Statement

You must know the action alternatives to decide if survey research can be useful and, if so, exactly how to proceed. Consider a pharmaceutical firm with increasing unit sales but decreasing profits. That's the problem management wishes to address, but before proceeding with any research the question needs to be broken down. The question might be "Should we increase the price of brand A by 10%?" This clear statement enables us to see that the answer lies in assessing consumer and competitive reaction to a specific change.

Stage 2: Questionnaire Design[8]

Fundamental laws Questionnaire design deals mainly with controlling measurement error. Most of its important points can be imparted by a fundamental law:

Use common sense:

Corollary A: Don't ask a question unless truthful answers to it will provide useful information in making the decision at hand.

Corollary B: If there is more than one way to get a particular piece of information (and there usually is), pick the questions for which respondents are likely to—

a. Know the answer.
b. Be willing to tell you the answer.

This law and its corollaries are pretty simple. Yet, much market research collects facts that help the manager make a right decision only accidentally. Before including any question on a survey, ask yourself, "How will I use the data from this question?" If you can't be any more precise than, "I'll analyze it," it's unlikely the data will be worth anything to you. (There are a few exceptions to

[8]See A.J. Silk, "Questionnaire Design and Development," No. 590-015, for a complete discussion.

Corollary A; for example, you may ask some questions to get respondent involvement or to disguise the purpose or sponsor of the study.)

Question and answer format Each question passing the "information test" should be examined for the burden it places on the respondent: Does he or she have the information you are looking for; will giving a truthful answer embarrass the respondent? How can we deal with "little" differences in question form making "big" differences in response?

Pretest Mentally putting yourself in the respondent's position helps to uncover questionnaire problems, but a pretest of the questionnaire is usual warranted. In a pretest, the questionnaire is administered to a small group of people *like the group to be sampled in the survey*. While filling out the questionnaire, and after they complete it, respondents are asked to explain responses, discuss any ambiguities, and so on. This can reveal unclear or sensitive questions.

Communication mode Finally, there are a number of ways to communicate with questionnaire respondents. These are generally personal interview, telephone, or mail. Many criteria are used in selecting the proper mode. Each mode offers obvious relative advantages. For example, in personal interviews we can show things to respondents, ask and explain complicated questions, and generally hold attention, allowing longer questionnaires. Telephone survey results are obtained quickly. Mail surveys are cheap. Most researchers, however, feel that *well-constructed* and *well-administered* questionnaires yield similar results, regardless of the form of interview.

Stage 3: Sampling

After the questionnaire is designed, pretested, and printed, the question is: whom do we want as respondents? This question breaks down into a number of parts; the first is designation of the target population.

The target population is the group of people to which estimates will be projected. Once the target population is specified, the issue is that of selecting the *sampling frame*, a means of representing the members of the target population. A perfect frame is one in which every member of the population is represented once and only once. It is something of a tautology, but the listing of the *Fortune 500* in *Fortune* magazine is a perfect frame for the population, *Fortune 500* firms. The *Fortune 500* would be an imperfect frame for U.S. business in general, however. Other examples of imperfect frames are a telephone book as the frame for households in a community (because of unlisted numbers and without telephone households) or a voter registration list for individuals over 18 years of age. If we use an imperfect frame, we err in projecting results to the target population (unless we redefine our target population). This infrequently acknowledged error is *frame error*. In many cases, we accept some frame error because the money spent developing a perfect frame could be better spent in other ways, such as in obtaining more respondents.

Sample selection After population and frame selection, we must specify the mechanism for selecting the members of the population to be included. The many ways of selecting a sample can be grouped into two categories.

1. Probability sampling.
2. Nonprobability sampling.

In probability sampling, each unit of the population has a predetermined chance of being included. For example, if in assessing the percentage of first-year MBA students at Harvard Business School owning a graphics printer, we bought a student directory, cut its relevant parts into 800 equal-size slips, each containing the name of one student, put them in a hat, and randomly selected 200 slips, each student would have a 0.25 chance of being selected. The result is a probability sample.

If, instead, we stood in front of the library and intercepted people on their way by until we caught 200 students, we would ultimately have polled the same percentage of the population. However, before we went out, we could not state the probability that any given individual would be included in the sample. Why? Because we don't know the chance that each individual walks by the library, and clearly we'd be wrong if we said everybody had an equal opportunity of being selected. Married students may go out the back door of the classroom building to the parking lot to drive home, some single students head right by the library on their way to the fitness facility, and so forth. This is nonprobability sampling. A common nonprobability sampling procedure is the *convenience sample*, where sampled units are selected not for a representative population, but for ease in getting their response.

The distinction between probability and nonprobability methods is important. The key difference is this: Only probability sampling methods yield raw data from which we can both (1) derive estimates that get closer to the true value in the population as we increase the sample size (we call these consistent estimates), and (2) objectively determine the likely accuracy of our estimates. With nonprobability methods, we cannot objectively measure our accuracy.

This is a pretty strong argument for probability sampling. With nonprobability sampling we get possibly biased estimates, but to make matters worse, we can't even figure out from the data how big the bias is likely to be, or in which direction. The argument for nonprobability methods is if they provide superior control of measurement error. Nonprobability methods are probably most useful in exploratory stages of research; the objectivity of probability methods is usually required in later stages of the research process.

Sample size Having decided whom to have as respondents and how to select the sample, we can determine how many respondents we should have, or what we are entitled to predict given a specified number of respondents. In practice, the second form of the question may be more relevant because in many cases sample size determination is rather ad hoc. Specifically, the sample size is determined by dividing the negotiated budget by the cost of obtaining a respondent. Sample size determination is, however, an economic question that should be analyzed.

Sampling distribution Procedures for sample size determination utilize the notion of sampling distribution. For example, let's define the population as the 800 first-year MBA students at Harvard Business School. Suppose a perfect frame exists for this target population, there is no measurement error and the underlying truth is that half of these individuals have graphics printers and half don't. Now, if we took a random sample of 2 of the 800 individuals, we would observe one of three things, as shown in the following table:

	Sampled Person 2	
Sampled Person 1	Printer	No Printer
Printer	2 Ps	1 P, 1 NP
No Printer	1 P, 1 NP	2 NPs

In terms of printer owners, we would either observe two (top left box), one (top right and bottom left), or none (bottom right). With some simple calculations, we can figure out the chance that each of the above would occur. For example, if the population in reality has 400 "printers" and 400 "non," it's clear that if we sample randomly the odds that the first person we select is "printer" is one half. Now, leaving the first person out, we have 799 people, and the probability that the second individual we pick is "printer" can be determined. If the first sampled person was printer, then 399 of the 799 remaining are "printers." Hence, the probability of the top left box with both "samplees" being printers is ½ x 399/799 = 0.2497. Similarly, if we calculated the other probabilities, we'd find:

Sampled Person 1	Sampled Person 2: Printer	Sampled Person 2: No Printer
Printer	2 Ps (0.2497)	1 P, 1 NP ($^1/_2$ x 400/799 = 0.2503)
No Printer	1 P, 1 NP ($^1/_2$ x 400/799 = 0.2503)	2 NPs (0.2497)

Now, in the traditional method of sampling we assume that the sample information is the whole story; in other words, we have no a priori notions about the extent of printers. Consequently, if we would up in the top left box, our best guess at the printer-owning percentage for first-year MBA students would be 100%. If we were in the top right or bottom left, we'd say 50%. Last, in the bottom right we'd say 0%.

This is a rather laborious way to reach what may be an obvious point: different samples lead to different estimates of the population value. Unless we get all members of our target population as respondents, we inevitably have some risk that the portion in our sample is not perfectly representative. Depending on the sample we draw, we may guess different values for the population as a whole. With no other information available to us, we guess that the value for the target population is the value observed in our sample.

Increased sample size As we increase the size of our sample, we tend to make better estimates of the value in the population. We just saw what happened when we drew samples of two from the target population of 800—400 owning graphics printers and 400 not. Now let's see what happens if we increase our samples to 5 and then to 10.

For samples of 5, our observed sample percentage of graphics printer owners will be 0%, 20%, 40%, 60%, 80%, or 100%. With the aid of the computer, we can do the equivalent of putting the 800 names in a hat, drawing out 5, checking the number of these having programmables, throwing their names back, and drawing another sample of 5. Simulating a total of 50 samples of 5 yielded the observed sample proportions shown on the next page.

The observed or sample proportions tend to bunch up around the true value. Note that because all our samples are of size 5, it is impossible for us to observe the true target population proportion of 50% in any one sample. We get within ±10% of the true value in 30 of the 50 cases (21 times we observed 2 graphics printers and 9 times we observed 3). That seems reasonable for sampling only 5 people, but on occasion we really botch it—there is a 6% chance we would say everybody has graphics printers (we observed 5 in the sample of 5 once) or nobody has graphics printers (we observed none in the sample of 5 twice).

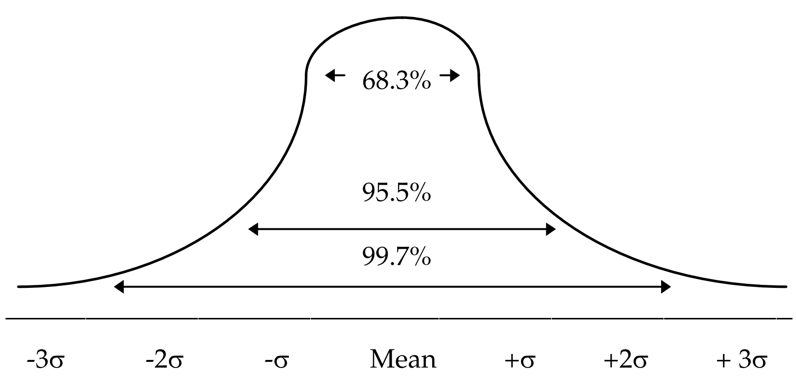
Percent of Printers in Sample of 5	Number of Times Observed
0%	2
20	10
40	21
60	9
80	7
100%	1
Total	50

It seems intuitive that we would "do better" the more students we sampled and, in fact, that's right. Simulating in the same way, but with a sample size of 10 yielded:

Percent of Printers in Sample of 10	Number of Times Observed
0%	0
10	1
20	0
30	7
40	12
50	16
60	6
70	5
80	3
90	0
100%	0
Total	50

We were within 10% on 34 (12 + 16 + 6) of 50 trials for 68% (as opposed to 60% before). Now we never say that everybody has one or nobody has one. The observed proportions from the larger sample are bunched more tightly around the true proportion.

If sample sizes are "big enough," the distribution of their sample means follows the normal distribution. ("Big enough" means about 100 in cases with yes/no, do/don't, or binary data, but as little as 30 with continuous data.) We know the probability that a given observation will fall within a particular range of values of the normal distribution. These ranges are usually counted in units called standard deviations, denoted by σ (sigma). For example, we know that 68.3% of the observations are within one σ of the mean; 95.5% are within 2 σ's; and 99.7% are within 3 σ's. Graphically, we show this as:



Because sample means are approximately normally distributed about the population mean, we can make statements such as the following: "About 95% of sample proportions fall within two standard errors of the population proportion." (We don't want to get bogged down in terminology here, but when the standard deviation we are talking about is the standard deviation of sample means, we call it the standard error.) Thus, if we estimate the population proportion, we can express our confidence in that estimate by constructing intervals around it. For example, the 95% *confidence interval* CI is [L, U] where L = our estimate minus 2 σ's and U = our estimate plus 2 σ's. All we need to know now is how to figure out σ, our counting unit or standard error. For samples of sizes n from large populations and data on binary questions:[9]

$$\textit{Standard error: } \sigma_p = \sqrt{\frac{p(1 - p)}{n}}$$

where, theoretically, p is the true proportion in the population. Of course, the whole point is to find out what p is, so a formula that assumes we already know it isn't too helpful. We used the observed frequency in our sample as an estimate of population proportion, so we do the same here and estimate σ_p by:

$$\hat{\sigma}_p = \sqrt{\frac{(\textit{Observed probability})(1 - \textit{Observed probability})}{n}}$$

Note that n, the sample size, is in the denominator. This means as n gets bigger we expect our standard error to get smaller; in other words, we expect our sample estimates if we drew repeated samples to get tighter around the true mean. This is what we saw in our earlier example; now the reason is mathematically clear.

Prediction of accuracy These results are used primarily to assess the likely accuracy of an estimate based on a given sample size or, conversely, to determine how large a sample is required to obtain a specified level of accuracy. Let's consider an example. An automobile dealer wants to know what proportion of customers consults *Consumer Reports* before shopping for a car. He draws a random sample of 100, and 43 say they consult *Consumer Reports*, 57 say they do not.

Implications:

1. The auto dealer's best guess at the percentage of all customers consulting *CR* is 43%.

2. How good is this best guess?

   We answer by constructing CIs. We know that approximately 95% of the observations in the normal distribution fall within 2 σ's of the mean.

   Based on the sample information, our estimate of σ_p is:

$$\hat{\sigma}_p = \sqrt{\frac{(0.43)\ (0.57)}{100}} = 0.0495$$

   Using this, we are "95% confident" that the true percentage for the population is within 0.0495 x 2 = 0.099 of the sample percentage of 0.43. We are 95% confident the true percentage for the population is in the range 0.331 to 0.529. (This assumes we have not committed any measurement error.)

   If we want to be "99% confident" that our range contains the true mean, we have to expand the range from ±2 σ's to ±3 σ's. Specifically, the 99% range is 0.2815 to 0.5785.

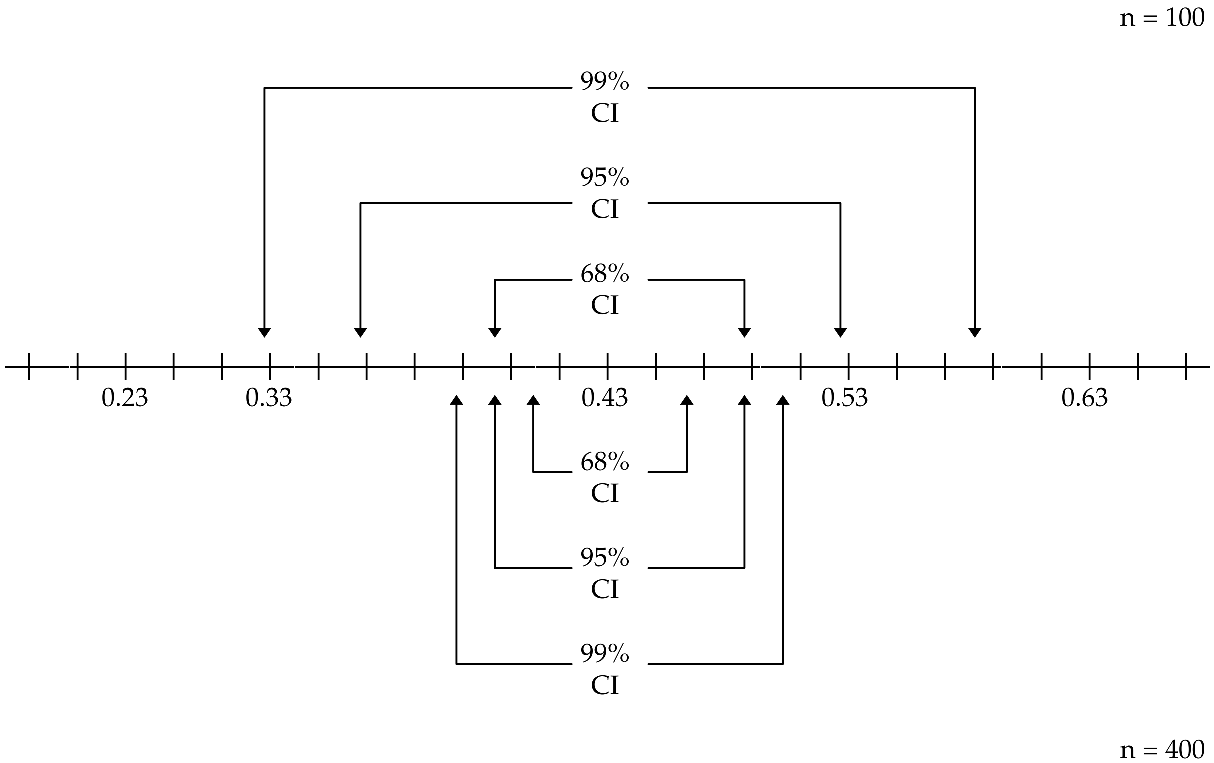
[9]Similar logic applies to continuous-type data with some change in the specifics of computations.

3. Better guesses can be obtained by increasing the sample size.

   If the auto dealer sampled 400 people instead of 100 and found 172 *CR* consulters and 228 nonconsulters, the best guess would again be 43%, but the standard error reduced to:

$$\sqrt{\frac{(0.43)\ (0.57)}{400}} = 0.0275$$

   so our 95% CI contracts to 0.3805 to 0.4795. We can collect these results in a figure showing confidence intervals for different degrees of confidence and sample size. Note that all the CIs for n = 400 are exactly one-half as wide as their counterparts for n = 100. Mathematically, this occurs because the standard error is related to the sample size by a square root rule. Therefore we do not achieve a 50% reduction in the standard error by doubling the sample size; we have to quadruple it.



Sample size and confidence The other way we can use our knowledge of properties of the normal distribution is in determining the sample size required for given levels of precision and confidence. For example, if we want to be accurate within X percentage points at the 95% confidence level, we solve for n in:

$$x = 2\sqrt{\frac{p(1 - p)}{n}}$$

Again, we have p, the true proportion, in our formula, so we have to estimate it. This time we are without the assistance of sample observations. A saving grace, however, is the fact that p(1 - p) can never be bigger than 0.25. So, if we use that value, the n determined provides at least the desired level of precision, more if p is radically different from 0.5. Sample sizes required for 95% levels of confidence are:

CI (in percentage points)	Sample Size Required
±2.0	2,500
±3.0	1,111
±4.0	625
±5.0	400
±7.5	178
±10.0	100

Permissible error There is no general rule on how big an error is permissible. By tradition, many national surveys have sample sizes of 2,000 or so to get 2% to 3% accuracy. However, the reliability required depends very much on the economic stakes of the decision being made.

Stage 4: Data Analysis

While this step is listed last, one should determine how one will analyze the data before they are collected. Depending on the research question, the appropriate data analysis may be:

1. Simple counts, e.g., the number of people in the survey who never hear of a product.

2. Graphical displays.

3. Assessment of association between two variables.

4. Assessment of the impact a set of independent variables has on a "dependent variable.

The important point is that the analysis be conducted in a way to reveal true relationships rather than biased to support some preconceived idea.

VI. Conclusion

There is a wide variety of market research techniques available. Each is capable of contributing to a certain set of issues. Each is also capable of leading the manager astray if not executed properly and interpreted in light of its limitation.

If market research is to be useful, the manager must take an active role in:

1. Initiating the research effort, i.e., a managerial decision to be made must drive the data collection and analysis.

2. Working with the research specialists to define the problem and determine how information will be used.

3. Reviewing the proposed research design as to potential benefit versus cost.

4. Ensuring that the data are analyzed in the most useful way and any uncertainty about the validity of the data is made clear.

5. Selecting the format for the presentation of the research results to other members of the management team.

While there is a great deal of debate on the best research procedure to follow for any given problem, there is a general agreement among managers that the quality of research information has a major impact on the quality of marketing decisions. Consequently, it is a part of the marketing process which the effective marketing manager must understand and direct.