



3 1293 00675 7680

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE
APR 1 1992		
APR 29 1992		
JUN 04 1992		

MSU Is An Affirmative Action/Equal Opportunity Institution

Marketing research
demand for studies
therefore, imperative
advantages of the
study.

One of the objectives
consumer preferences
be used to evaluate
Every design but one
least one session
city Consumer Preference
the basis for determining
cedure. No one can
data analysis can
stead, it is recognized
cedure which best
within the sponsor's
each testing procedure
one should recognize

ABSTRACT

AN EVALUATION OF THE METHODS USED IN DESIGNING AND ANALYZING CONSUMER PREFERENCE STUDIES

by Raymond Albert Marquardt

Marketing research workers are being confronted with an increasing demand for studies designed to measure consumer preferences. It is, therefore, imperative that they become aware of the advantages and disadvantages of the methodical alternatives available for use in such a study.

One of the objectives of this study was to describe and evaluate consumer preference panel designs and types of data analysis which can be used to evaluate consumer preferences for all types of products. Every design but one was used to guide the presentation of products to at least one session of the Michigan State University-Wayne State University Consumer Preference Panel. These empirical observations provided the basis for determining the advantages and disadvantages of each procedure. No one of the more than fifteen statistical designs or methods of data analysis can be recommended for use in all testing situations. Instead, it is recognized that the marketing researcher should use the procedure which best satisfies the project objectives, but which remains within the sponsoring organization's budget, time, and talent restraints. Each testing procedure has certain advantages and disadvantages which one should recognize when he designs a consumer preference study.

A second objective of the laboratory and was defined as a panel to evaluate the product of trained or untrained panel is that a relative product differences between testers do not reflect the trained panel are allowed convenience from household to household cause many products. A advantage is that the exist when the product. The mass consumers are reached by conducting mass consumer interview. The fact that a large number of its main disadvantages and the growth. The consumer can provide

A second objective was to describe and evaluate the contribution of the laboratory and mass consumer preference panels. The laboratory panel was defined as a panel whose members congregate at a central location to evaluate the products being tested. Its membership consists of either trained or untrained testers. The chief advantage of the trained laboratory panel is that a relatively small number of panelists can be used to detect product differences. Its chief limitation is that responses of trained testers do not reflect consumers' preferences. The advantages of the untrained panel are: (1) it insures a uniformity in samples tested, (2) it allows convenience in testing because the samples are not moved from household to household, and (3) the cost per product tested is small, because many products can be tested during one panel session. Its disadvantage is that the tests are not conducted under the same conditions as exist when the product is consumed.

The mass consumer panel was defined as any panel in which the members are reached at their residence or office. The two methods of conducting mass consumer panels are by mail questionnaire and by personal interview. The chief advantage of the mail questionnaire technique is that a large number of respondents can be reached at a relatively low cost. Its main disadvantages are that only a few questions can be asked in one mailing and that the conditions under which samples are tested are not known. The chief advantage of the personal interview technique is that it can provide detailed information on consumers' attitudes toward the

product being test

interviewing is do

tioning. Another

results.

Many of the p

panel may be over

that the panel refl

which the product

should be uniform

tested. Results o

identified by symb

order. When a ra

maximum to be con

product being tested. One of its disadvantages is that, unless telephone interviewing is done, it is more expensive to conduct than is mail questioning. Another limitation is that interviewer attitudes may bias the results.

Many of the problems one should consider in the organization of a panel may be overcome by proper experimental design, which can insure that the panel reflects the characteristics of the market population in which the product will be sold. The samples presented to the panel should be uniform in size, temperature, and any other variables not being tested. Results obtained in this study indicate that samples should be identified by symbols, such as (), *, &, and %, which suggest no implied order. When a rank order design is used, four items are the recommended maximum to be compared in one tasting series.

AN EVALUATION OF THE METHODS
USED IN DESIGNING AND ANALYZING
CONSUMER PREFERENCE STUDIES

By

Raymond Albert Marquardt

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics

1964

The author e
sisted in the de

mitted to: Dr.

Dr. W. J. E. C

me and guidan

The financ

the Food Scien

Research Foun

cited.

Finally,

and encourag

Full res.

author.

ACKNOWLEDGMENTS

The author expresses his sincere appreciation to all those who assisted in the development of this thesis. The author is particularly indebted to: Dr. H. E. Larzelere, Dr. W. S. Grieg, Dr. L. V. Manderscheid, Dr. W. J. E. Crissy, Dr. V. E. Smith, and Dr. J. H. Stapleton for the time and guidance each offered.

The financial assistance provided by the Agricultural Economics and the Food Science Departments at Michigan State University and the Sugar Research Foundation, 52 Wall Street, New York City was deeply appreciated.

Finally, the author wishes to thank his wife, Alberta, for her patience and encouragement during the course of the thesis preparation.

Full responsibility for errors in this thesis should be placed on the author.

ACKNOWLEDGMENT

LIST OF TABLES .

LIST OF FIGURES

LIST OF APPENDICES

Chapter

I. INTRODUCTION

The Imp

The

Th

Th

The

The

Br

The

Us

Objecti

Organiz

II. TYPES OF
OF EACH

Laborat

Par

Use

Use

Use

Use

Use

Use

Mass

Par

Use

Me

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	ii
LIST OF TABLES	viii
LIST OF FIGURES	x
LIST OF APPENDICES	xi

Chapter

I.	INTRODUCTION	1
	The Importance of Consumer Preference Testing	1
	The Importance of Product Development	2
	The Process of Product Development	4
	The Need for Pre-Market Testing	6
	The Need for an Evaluation of Present Product Grades	9
	The Need for an Evaluation of the Effect of Product Branding.	11
	The Need for Establishing Techniques which can be Used to Evaluate Consumers' Preferences	12
	Objectives	12
	Organization of Thesis.	13
II.	TYPES OF PREFERENCE PANELS AND THE CONTRIBUTION OF EACH TYPE OF PANEL	15
	Laboratory Panels	15
	Panel Membership	16
	Uses of Laboratory Panels	17
	Quality Control	17
	Evaluating New or Improved Products	18
	Evaluation of Branding Upon Consumer Preferences	19
	Advantages of Laboratory Panels	19
	Disadvantages of Laboratory Panels	21
	Mass Consumer Panels	24
	Panel Membership	25
	Uses of Mass Consumer Panels	26
	Methods of Conducting Mass Consumer Panels	27
	Mail Questionnaire Method	27

III. PROBLE

The Pr

Panel

The

The

The

Simult

Prepar

Size

Tem

For

Dis

Sam

Pos

Ord

Nun

Con

Exa

IV. ONE SAM

Hedoni

Exa

Adv

On

Dis

of

Readin

Exa

Adv

Ev

Dis

Ev

V. TWO SA

Paired

Te

Chapter	Page
Causes of Non-response	27
How to Obtain High Returns from Mail Surveys.	28
Advantages of the Mail Questionnaire Method .	31
Disadvantages of the Mail Questionnaire Method	32
Personal Interview Method	34
Advantages of Personal Interview Method . . .	34
Disadvantages of Personal Interview Method .	34
III. PROBLEMS TO CONSIDER IN PANEL ORGANIZATION	36
The Problem of an Adequate and Accurate Sample of Panelists	36
The Problem of Representativeness	36
The Problem of Non-Cooperators	43
The Problem of Adequate Panel Size	45
Simultaneous or Single Stimulus Sample Presentation . .	47
Preparation and Serving of Samples	51
Size of Samples.	51
Temperature of Samples.	51
Form in Which Samples Are Served	52
Discussion at Judging Sessions	52
Sample Coding	52
Position on Cards.	53
Order of Presentation of Samples	54
Number of Samples and Replications	56
Consistency of Panel Preferences	57
Example of Panel Organization	58
IV. ONE SAMPLE TESTS	61
Hedonic Scale	61
Example of the Use of Hedonic Scale.	63
Advantage of Hedonic Scale in the Evaluation of One Sample	63
Disadvantage of Hedonic Scale in the Evaluation of One Sample	64
Readiness to Buy Index	65
Example of the Use of the Readiness to Buy Index . .	67
Advantages of Readiness to Buy Index in the Evaluation of One Sample	71
Disadvantages of Readiness to Buy Index in the Evaluation of One Sample	71
V. TWO SAMPLE TESTS	72
Paired Comparison Tests.	73
Test Procedure	73
Use of Sign Test on Ranks.	74

Triang
Tes
Exa
Adv
Dis

VI. THREE A

Series
Tes
Adv
Dis
Chi So
Tes
Exa
Adv
Dis
Coeffi
Analy
Tes

Exa
Adv
Dis
Individ
Test i
Tes
Exa
Adv
Dis

VI. FIVE OR

Balance
Analys
Test
Exa
La

Chapter	Page
Example of the Use of Sign Test	78
Advantage of Sign Test	79
Disadvantages of Sign Test	79
Use of Wilcoxon Test on Successive Interval Scales	80
Example of the Use of the Wilcoxon Test . . .	83
Advantages of the Wilcoxon Test.	84
Disadvantage of the Wilcoxon Test	86
Triangle Test.	87
Test Procedure	87
Example of Triangle Test	91
Advantages of Triangle Test	93
Disadvantages of Triangle Test.	94
VI. THREE AND FOUR SAMPLE TESTS	96
Series of Paired Comparison Tests.	96
Test Procedure	96
Advantages of Series of Paired Comparison Tests . .	97
Disadvantages of Series of Paired Comparison Tests .	99
Chi Square Test	99
Test Procedure	99
Example of Chi Square Test.	102
Advantages of Chi Square Test	104
Disadvantages of Chi Square Test	105
Coefficient of Concordance (Friedman Two Way Analysis of Variance)	106
Test Procedure	107
Tied Rankings	107
Testing the Significance of W	108
Example of the W and Friedman Test	111
Advantage of the W and Friedman Test	112
Disadvantage of the W and Friedman Test	113
Individual Paired Tests to be Used if Over-All Test is Significant	113
Test Procedure	114
Example of Paired Test	115
Advantage of Paired Test	117
Disadvantage of Paired Test	118
VII. FIVE OR MORE SAMPLE TESTS	119
Balanced Incomplete Block Design Analyzed by Analysis of Variance	121
Test Procedure	121
Example of an Analysis of a 4 by 4 Balanced Lattice Experiment	124

Adv
Dis
Fracti
Tes
Adv
Dis
Multip
Tak
Sch
R
A
I

VIII. TESTS TO
MORE F

Use of
Measu
Test
Exan
Adva
Disa
Use of
the Rel
Test
Adva
Disa
Series
Identid
Test
Adva
Disa

IX. DISCUSS

Optimu
Testing
Interpr
Validiti
Conclu
New
Eva
Co

Chapter		Page
	Advantage of Balanced Lattice Designs	128
	Disadvantages of Balanced Lattice Designs	129
	Fractional Factorial Design	129
	Test Procedure Illustrated on an Example.	129
	Advantage of Fractional Factorial Designs	137
	Disadvantages of Fractional Factorial Designs.	137
	Multiple Comparison Tests	138
	Tukey's Simultaneous Comparisons.	138
	Test Procedure	138
	Example of Tukey's Simultaneous Comparisons	139
	Advantage of Tukey's Method	141
	Disadvantage of Tukey's Method	141
	Scheffe's Method of Multiple Comparisons.	142
	Test Procedure	142
	Example of Scheffe's Method	142
	Advantages of Scheffe's Method.	143
	Disadvantages of Scheffe's Method	144
VIII.	TESTS TO DETERMINE WHETHER A PANELIST WILL PAY MORE FOR HIS PREFERRED PRODUCT.	145
	Use of the VonNeumann-Morgenstern Model in Measuring the Relative Utility of Two Products.	145
	Test Methodology and Assumptions	145
	Example of VonNeumann-Morgenstern Method	148
	Advantage of VonNeumann-Morgenstern Method	159
	Disadvantage of VonNeumann-Morgenstern Method	161
	Use of the Demand-Price Concept in Measuring the Relative Utilities of Two Products	162
	Test Procedure Illustrated on an Example.	162
	Advantage of Demand-Price Method	170
	Disadvantage of Demand-Price Method.	171
	Series of Rankings Involving Different Prices on Identical Products.	171
	Test Procedure Illustrated on an Example.	171
	Advantage of Series of Rankings	174
	Disadvantage of Series of Rankings	175
IX.	DISCUSSION	176
	Optimum Degree of Testing	176
	Testing Sequence	179
	Interpretation of Results	182
	Validity of Panel Results in Predicting Sales	184
	Conclusions	187
	New Product Development Problems	187
	Evaluation of Product Popularity Against Competition Problems	188

Chapter

Eva
Pro
Co
Qu
Qu

APPENDICES. . .

HELIOGRAPHY. . .

Chapter	Page
Evaluation of the Effect of Branding, Advertising, Promotion or Packaging Problems	189
Cost Reduction Problems	190
Quality Improvement Problems	191
Quality Control Problems	192
Variation in Source of Ingredients	192
Routine Quality Control	192
Storage Studies	192
APPENDICES.	195
BIBLIOGRAPHY.	208

Table

- I. Differences
Nine
Comp
of Ho
- II. Probabil
- III. Results
- IV. The Obs
Ranki
- V. Scale Sc
- VI. Rankings
- VII. An Examp
- VIII. Analysis
to Det
Prepar
of Sal
- IX. Factors a
- X. Formulas
- XI. Sample Pa
- XII. Typical Q
- XIII. Preferenc
in the
- XIV. Factorial
Obtai
- XV. Illustrati
Averag

LIST OF TABLES

Table	Page
I. Difference in Palatability Ratings Obtained From a Nine-Point Scale of Pork Loin Roasts, Fresh as Compared with Fresh-Frozen, East Lansing Panel of Housewives, April 1957. Session I.	85
II. Probability in Triangular Preference Tests.	90
III. Results of Triangular Panel Tests of Ham	92
IV. The Observed and Expected Frequency of Panelists Ranking Four Kinds of Sausage with a 1, 2, 3, or 4 . . .	104
V. Scale Scores Obtained From Panelists	110
VI. Rankings Obtained From Panelists' Scale Scores	110
VII. An Example of the Paired Chi Square Test	116
VIII. Analysis of Variance of a 4 by 4 Balanced Lattice Design to Determine Consumer Preferences Among Hams Prepared with Four Levels of Sugar and Four Levels of Salt	127
IX. Factors and Levels of Factors in Experiment	130
X. Formulas to be Tested in Experiment	131
XI. Sample Pairs Presented to Panelists in Experiment	132
XII. Typical Questionnaire Data Used in the Experiment. . . .	133
XIII. Preference Scores and Acceptability Ratings Obtained in the Experiment.	134
XIV. Factorial Contrasts for Selected Consumer Test Results Obtained in the Experiment.	136
XV. Illustration of Some Significant Factor Effects Using Average Scores Obtained in the Experiment	136

Table

XVI. Example
Consum

XVII. Consumer
at App

XVIII. Example
Prefer

XIX. Calculat
Brand

XX. Number o
Value

XXI. Amount E
They

XXII. Amount o
Forfe

XXIII. Number o
as Fir

XXIV. Number o
Large
Sizes
Fifth
29¢ to

XXV. Scheffe A

Table		Page
XVI.	Example of Tukey's Simultaneous Comparisons on Consumer Preferences for 16 Types of Ham	140
XVII.	Consumer Preferences Among Sugar Differences in Ham at Approximately the 5% Level of Significance	141
XVIII.	Example of Scheffe's Multiple Comparisons on Consumer Preference for 16 Types of Ham	143
XIX.	Calculation of the Minimum Utility Placed in the Known Brand Turkey	154
XX.	Number of Panelists and their Indirect Evaluation of the Value of the Well-Known Brand Turkey in Utils	155
XXI.	Amount Extra the Panel Members Indirectly Indicated They Would Pay for a Well-Known Brand Turkey	158
XXII.	Amount of Money Panel Members Indicated They Would Forfeit to Receive a Known Brand Turkey	167
XXIII.	Number of Panelists Ranking Different Sizes of Eggs as First, Second, Third, Fourth, and Fifth Choice.	173
XXIV.	Number of Panelists Ranking Larger (Jumbo and Extra Large) Sizes of Eggs and Smaller (Medium and Small) Sizes of Eggs as First, Second, Third, Fourth and Fifth Choice when the Larger Sized Eggs are Priced 29¢ to 35¢ Higher Per Dozen than the Smaller Eggs	173
XXV.	Scheffe Analysis of Variance Procedure	205

Figure

- I. Quest
Pre
- II. Quest
to R
- III. Paneli
- IV. A Perfe
Proo
- V. Questio
of T
- VI. Approxi
in C
- VII. Questio
Price
- VIII. Approxin
Branc
- IX. Demand
of Te
- X. Testing S
Sever

LIST OF FIGURES

Figure		Page
I.	Questionnaire Illustrating the Range of Consumer's Predispositions Toward a Product	66
II.	Questionnaire Illustrating the Rise of the Readiness to Buy Index	68
III.	Panelists' Evaluation of Sterile Milk	70
IV.	A Perfectly Consistent Panelists' Rankings of Four Products in 12 Paired Comparison Tests	98
V.	Questionnaire Used in Measuring Relative Utilities of Two Brands of Turkey	149
VI.	Approximate Demand Curve for a Known Brand Turkey in Comparison with an Unknown Brand Turkey	160
VII.	Questionnaire Presented to Panelists in a Demand- Price Study	164
VIII.	Approximate Demand Curves for Well Known and Unknown Brand Turkeys as Derived from Regression Analysis . .	169
IX.	Demand and Cost Curves for Different Degrees of Testing	179
X.	Testing Sequence in Consumer Preference Testing of Several Products	182

Appendix

A Sum of
Diff

B Analysis
Squa

C Example
That

D Analysis
Bala
Prefe

E The Sch

F Regress

LIST OF APPENDICES

Appendix		Page
A	Sum of the Ranks for Sausage Containing Four Different Levels of Sugar	196
B	Analysis of Variance Table for Calculation of Sums of Squares from Balanced Lattice Design	197
C	Examples of Other Balanced Incomplete Block Designs That Can Be Used to Evaluate Five or More Products. .	198
D	Analysis of Variance Computations for the 4 by 4 Balanced Lattice Experiment on Consumer Preferences for Ham	203
E	The Scheffe Analysis of Variance	204
F	Regression Data for UBT and KBT in Demand-Price Study .	207

Marketing re
with an incre
sumer behav
increasing n
study, are u
of study and
cate the need
though disag
indicates a g
methods. It
adaptability

The main jus
such studies
conditions an
action. These
environment
possible gain

Consumer p

It is particularly

product testing a

of product grade

important role in

Other uses

tests are:

¹Work Gro
erence," A Rep
College, July 1
of Agriculture (1
pp. 94-5.

²Ibid., p.

CHAPTER I

INTRODUCTION

The Importance of Consumer Preference Testing

Marketing research workers in recent years have been confronted with an increasing demand for studies designed to measure consumer behavior and preference. In response to this demand an increasing number of projects, varying widely as to method of study, are undertaken each year. The wide variety in methods of study and the controversy among researchers over them indicate the need for careful evaluation of alternative methods. Although disagreement is sometimes more apparent than real, it indicates a great interest in the development of more satisfactory methods. It also points to the need for careful appraisal of the adaptability and limitations of the various research methods.¹

The main justification for consumer preference studies is that such studies may indicate present or future behavior whenever conditions are such as to permit a transfer from desire to positive action. These conditions may originate within the consumers' environment or may be provided by the merchandiser who foresees possible gains accruing to his business.²

Consumer preference testing of products is important in many areas. It is particularly important in the area of product development. However, product testing also plays a large role in the evaluation or establishment of product grades. Then too, consumer preference testing can play an important role in the evaluation of the effect of product branding.

Other uses for the information obtained from consumer preference tests are:

¹Work Group I, "Techniques of Studying Consumer Behavior and Preference," A Report of the Marketing Research Workshop, Michigan State College, July 13 to 21, 1951, prepared by the United States Department of Agriculture (Washington: Agriculture Research Administration, 1951), pp. 94-5.

²Ibid., p. 95.

1. to improve
2. to keep
a product with
which are used
sales organization
ingredients
consumer products
tive advertising
direct effort
how much of
herent quality
effort, thus
selling necessity
confidence to
superiority,
in a futile effort
tics are at fault
with the possibility
with cost of production
both practical

Recent business

trigs, chemical

oriented, as illustrated

growth companies

sales and profits

approximately 50 per cent

since World War II

Thousands of

annually since

industries approach

and expenditures

1 Donald R.
Testing," Food

2 Ralph W.
ening: Perspectives
ally, (Homes)

1. to improve the products in accordance with market preferences, 2. to keep abreast of changing preferences, 3. to add "frills" to a product which consumers want and to eliminate and avoid those which are unessential, 4. to supply selling arguments for the sales organization to use, 5. to determine how far economy in ingredients may be practiced while maintaining or increasing the consumer preference for the product, 6. to find the most effective advertising appeals regarding the product's qualities, 7. to direct effort toward special groups of buyers, 8. to determine how much of the trade preference for a particular product is inherent quality and how much is due to past and present selling effort, thus obtaining a clue to the amount of advertising and selling necessary for promotion of a new product, 9. to give new confidence to the selling organization through proof of product superiority, 10. to avoid lavish selling and advertising expense in a futile effort to promote sales when the product's characteristics are at fault, 11. to aid in pricing a product in accordance with the possibilities of preference for it, rather than in accordance with cost of production, and 12. to aid in devising packages having both practical and aesthetic appeals.¹

The Importance of Product Development

Recent business history indicates that "growth industries" such as drugs, chemicals, electronics, etc. have been heavily "new product" oriented, as illustrated by research and development expenditures. Most growth companies in all industries attribute a large percentage of their sales and profitability to their new products. It is common to have approximately 50 percent of sales in products that are new to the company since World War II.²

Thousands of new products have been introduced into the market annually since the end of World War II. The growth rate of many industries appears closely correlated with the amount of effort and expenditure put into new-product activity. Trade journals

¹ Donald R. G. Cowan, "Developing and Improving Foods by Consumer Testing," *Food Industries*, Vol. 13, (March 1941). p. 44.

² Ralph W. Jones, "Management of New Products," *Managerial Marketing: Perspectives and Viewpoints*, "ed.," William Lazer and Eugene J. Kelly, (Homewood, Illinois: Richard D. Irwin, Inc., 1962), p. 444.

continue to be
But the rate of
in many circles
this, certainly
success part
electrical ap

The new ma

tion to manufact

managers now m

This means taking

departments.² M

satisfy current o

Since World

competition. Th

the marketing mix

there has also be

product quality.

majority of indus

competition."³

tion is likely to

both the ability

in quality of pro

¹ Hector Laz
Hill Book Compan

² Ibid., p.

³ W. Smith
Agricultural Mar
Chapt. VIII. (To

⁴ Ibid., p.

continue to be filled with news and descriptions of new products. But the rate of success of such products is relatively so low that in many circles the failures are described as alarming... Despite this, certain industries and companies report a high degree of success particularly in pharmaceuticals, foods, chemicals, and electrical appliances.¹

The new marketing concept reflects the modern company's determination to manufacture what the customer wants. Therefore the marketing managers now make provisions for product planning as a marketing tool. This means taking product planning out of the production or engineering departments.² Now each firm produces products which are expected to satisfy current or expected consumer wants.

Since World War II there has been an increased trend toward non price competition. This has meant increased competition in other elements of the marketing mix. Because one of these elements is product quality, there has also been increased emphasis placed upon improving present product quality. In fact, Greig states that "quality competition among the majority of industrial firms may be more highly competitive than price competition."³ It appears that the emphasis placed upon quality competition is likely to increase because "as education and income increases, both the ability to discern between qualities and the wants for variation in quality of products increases."⁴

¹Hector Lazo and Arnold Corbin, Management in Marketing (McGraw-Hill Book Company, Inc., New York, 1961), p. 191.

²Ibid., p. 24.

³W. Smith Greig, "Quality Competition and Product Development," Agricultural Market Analysis: Development, Performance, Process. Chapt. VIII. (To be published.)

⁴Ibid., p. 5.

Then too, re

investment which

to keep selling m

design project wh

which may make

Another reas

that in many inst

posed with the ri

costs, while mai

time, and volume

The Process of Pr

Jones lists s

from the setting c

company's previo

importance it att

consists of searc

this stage the pro

ined. The seco

analysis to deter

IFICATIONS, invol

¹ Gerald C.

² Kirk McCrea

³ Jones, op.

Then too, research and good design is now considered as a one-time investment which can be amortized over a period of years. It is important to keep selling management on the benefits to be derived from a one-time design project which may cost one-tenth of the advertising program, but which may make the advertising program much more effective.¹

Another reason product development is becoming more important is that in many instances, both the supplier and the distributor are being faced with the rising costs. One method for them to meet these rising costs, while maintaining profits at the same level, is by increasing volume, and volume can be increased by handling new products.²

The Process of Product Development

Jones lists six stages which comprise the complete process involved from the setting of company objectives to ultimate product success.³ The company's previous commitments of time, talent, and money influence the importance it attaches to each stage. The initial stage of EXPLORATION consists of searching for product ideas to meet company objectives. In this stage the product fields of primary interest to the company are determined. The second stage consists of SCREENING or performing a quick analysis to determine which ideas are pertinent. The next stage, SPECIFICATIONS, involves the expansion of the idea, through creative analysis,

¹Gerald C. Johnson, "Effective Marketing Begins on the Design Board," Journal of Marketing, (July 1948), p. 33.

²Kirk McCreary (unpublished manuscript), p. 4.

³Jones, op. cit., p. 449.

into the concrete

its program.

The fourth

paper into a pro

will pertain mai

use of experimen

which were made

last stage, COM

full-scale produ

tion and resource

The primary

analyze data obt

ate new and ext

with stage, TEST

Currently, th

ation. However

research directed

first phase invol

tained are usual

ment. They are

panel can yield

acteristics such

This type of tes

¹ McCreary,

into the concrete business recommendation of the product's features and its program.

The fourth stage, DEVELOPMENT, consists of turning the idea-on-paper into a product-in-hand, producible and demonstrable. This thesis will pertain mainly to the fifth stage, TESTING. This stage involves the use of experiments necessary for verification of earlier business "judgments" which were made with the life expectation of the product in mind. The last stage, COMMERCIALIZATION, consists of launching the product in full-scale production and sale, thereby committing the company's reputation and resources.

The primary purpose of this investigation is to design, collect and analyze data obtained from consumer preference panel tests used to evaluate new and existing products. This corresponds to the work done in the fifth stage, TESTING, of Jones' process of product development.

Currently, there is no standard procedure in this type of product evaluation. However, there are usually three general phases in conducting research directed at the pre-market evaluation of new products.¹ The first phase involves the use of laboratory panels because results so obtained are usually obtained in the earlier stages of new product development. They are useful in guiding the development work. The laboratory panel can yield results which can be applied to most of the sensory characteristics such as the taste, smell, color, and tenderness of the product. This type of testing will be presented in the first section of Chapter II.

¹McCreary, op. cit., p. 7.

The next

homes of typical

actory panel.

of consumers who

testing can provide

product. It can

sumer preferences

in Chapter II.

The last phase

accomplished by

sales results are

in limited test and

ational detail on

The Need for Pre

Each firm attempts

right. The product

entiation strategy

The product differentiation

firm attempts to achieve

product by advertising

is merchandising

closely conform

attempt to obtain

The next phase consists of testing the product in actual use in the homes of typical consumers or in the large scale untrained consumer laboratory panel. In this phase, the product is presented to a selected group of consumers who use the product and then evaluate it. This type of testing can provide information which can help improve and perfect the product. It can also give a reasonably reliable indication as to the consumer preference for the product. This type of testing will be discussed in Chapter II.

The last phase is the actual testing of consumer acceptance. This is accomplished by using market or retail sales tests. In this phase, actual sales results are obtained on the new product when it is put on the market in limited test areas. It is beyond the scope of this thesis to go into additional detail on this type of market testing.

The Need for Pre-Market Testing

Each firm attempts to shift the demand curve for its products to the right. The producer accomplishes this by following either product differentiation strategy, market segmentation strategy or a combination of these. The product differentiation strategy is a promotional strategy in which the firm attempts to alter the consumer's attitudes and beliefs concerning the product by advertising and promotion. The market segmentation strategy is merchandising strategy which attempts to adapt product quality to more closely conform to consumer preferences. In either strategy the firm must attempt to obtain reliable information on consumers' tastes, wants, and

habits. To obtain

product quality

wants.¹

The success

not any easy task

survives to make

we are victims

quality, poor tim

either on a poor

ment program.

The food ind

tinuous developm

grocery departme

1950 to 1960 in 1

came from new pr

Another chara

food manufacturer

this competition

¹ Greig, op.

² Albert M. F.
Managerial Mark
and Eugene J. Ke
p. 444.

³ "The Super
1959, p. 19.

⁴ "The Big C
Annual Grocery

habits. To obtain such information, the firm must vary the product and product quality experimentally to test new hypotheses concerning consumer wants.¹

The success creation, manufacturing, and sale of new products is not any easy task. Only one of ten new products which reach the market survives to make a profit for its company. The products that do not survive are victims of high price, lack of demand, poor distribution, poor quality, poor timing or lack of reliability, most of which can be blamed either on a poor analysis of the market or on an inadequate product development program.²

The food industry particularly appears to be characterized by a continuous development of new products. The number of items carried in the grocery department in the average supermarket has increased from 2200 in 1950 to 4900 in 1958.³ Nearly 70 percent of the 1958 supermarket sales came from new products which were not on the market ten years ago.⁴

Another characteristic of the food industry is the competition among food manufacturers for shelf space in the supermarket. The keenness of this competition is illustrated in a report by Super Valu Stores. In 1958,

¹ Greig, op. cit., p. 10.

² Albert M. Rockwood, "Planning a Product-Development Program," Managerial Marketing: Perspectives and Viewpoints, "ed.," William Lazer and Eugene J. Kelley, (Homewood, Illinois: Richard D. Irwin, Inc., 1962), p. 444.

³ "The Supermarket Industry Speaks," Chicago, Supermarket Institute, 1959, p. 19.

⁴ "The Big Challenge in Food Marketing," This Week Magazine, Eighth Biennial Grocery Study, New York, 1959, p. 9.

only 1 out of
Of every eight
them.¹

In view of
even stronger
the buying com
the market pot
interested in
not. Both the
that the new pr
retail store.²

The neces
product is reent
sidered. This V
tion dollars to p
that this new pr
large companies
reputation in the
serious effect on

¹ "Supermar
Grocery Study ,

² McCreary

³ "The Big C

⁴ Ibid. , p.

only 1 out of 4 of the 5000 items presented to Super Valu were accepted. Of every eight items accepted, Super Valu subsequently dropped five of them.¹

In view of this keen competition, and with the prospect of it getting even stronger, the food manufacturer must be in a position to present to the buying committee of the food distributor factual information concerning the market potentials for a new product. This procurement committee is interested in knowing whether the consumer readily accepts the new product. Both the manufacturer and the distributor have to be fairly certain that the new product will sell before it will ever get on the shelf in a retail store.²

The necessity of determining beforehand the market potential of a new product is reenforced when the cost of developing a new product is considered. This Week Magazine estimates that it costs more than one million dollars to put a new product on the National Market.³ The chances that this new product will be a success are only 1 in 10.⁴ While some large companies could absorb such a loss, the resulting damage to their reputation in the distribution trade and with the consumers could have serious effect on the sales of their established line of products.

¹"Supermarketing, U.S.A.," This Week Magazine, Eighth Biennial Grocery Study, New York, 1957, p. 9.

²McCreary, op. cit., p. 5.

³"The Big Challenge in Food Marketing," op. cit., p. 29.

⁴Ibid., p. 9.

Thus, the
as it enables
for his propos
properly condu
will demand.
that might be s

Th

The fundam
means for cons
efficiently. Th
producer, consu
grades they pref
effectiveness of th
producers, and o

If the comm
is felt directly.
personal inspect
wholesalers are

The problem
facilities. Some

1. There is

1 Rockwood

Thus, the importance of a good market survey cannot be overestimated as it enables the manufacturers to lay out a complete set of specifications for his proposed product: its size, life, weight, function and cost. A properly conducted survey can indicate what the prospective customers will demand. It can also guide the establishment of the quality of product that might be sold for a given price.¹

The Need for an Evaluation of Present Product Grades

The fundamental economic justification of grades is that they afford a means for consumers to register their preferences more accurately and more efficiently. Thus, if the grading system is carried all the way back to the producer, consumers are better able to encourage the production of the grades they prefer. Therefore, it is the consumers who determine the effectiveness of the grades which are sometimes set up by technologists, producers, and dealers.

If the commodity is not changed in form, the influence of the consumer is felt directly. Even if the consumer buys the product on the basis of personal inspection and not on grade, the grades used by shippers and wholesalers are indirectly related to what the consumer wants.

The problem of choosing correct grade standards involves several difficulties. Some of the major difficulties are listed below.

1. There is no general agreement as to the number of attributes having

¹Rockwood, op. cit., p. 439.

economic ind

average of a

the relevant

2. Once

be attached to

influence? H

have in relatio

products.

3. Where

pend upon the

certain qualitie

available.

Consumers

grades and so a

by encouraging t

readily recognize

differentials in l

conform more clo

provided that st

process, and me

desires. Thus,

to see if a suffi

for a higher grad

economic influence upon most agricultural products. A grade is a weighted average of a number of such attributes. Thus, there is a need to identify the relevant attributes for each product.

2. Once these factors have been identified, a relevant weight must be attached to each attribute, e.g., which factors have the most economic influence? How much more (or less) economic influence does one factor have in relation to the others? These weights are not now known for most products.

3. Where the boundaries between grades should be placed will depend upon the degree to which the various users will pay premiums for certain qualities rather than substitute adjacent qualities within the ranges available.

Consumers frequently do not understand the presently established grades and so are not willing to pay price differentials for these grades. By encouraging the establishment of grades that consumers can more readily recognize, consumers will thereby be more likely to pay price differentials in line with the various grades. If the price differentials conform more closely with grade differences, economic incentives will be provided that stimulate producers and marketing agencies to produce, process, and merchandise commodities in line with consumers economic desires. Thus, there is a need for testing of current or proposed grades to see if a sufficiently large group of consumers will pay a premium say, for a higher grade in contrast with a lower grade.

4. There

description of

is a need to f

such as inche

The relati

mine consumer

nation about th

The N

The value

teen overlooked

deemphasized i

of farm products

some agricultural

Makens, in his

for turkeys conc.

to pay a premium

other agricultural

from a product of

of branding upon

the particular pr

tion program.

1
J. C. Mack
for frozen Whole
State University

4. There is a difficulty in translating consumer's preferences into a description of the product in objective and measurable terms. Thus, there is a need to formulate grade standards in terms of definite measurements such as inches, pounds, a certain number on a color scale, etc.

The relatively small amount of research that has been done to determine consumers' preferences presents a need for definite quantitative information about the details of consumer preferences.

The Need for an Evaluation of the Effect of Product Branding

The value of brands to the marketing of agricultural products has often been overlooked. One of the reasons that product differentiation has been deemphasized in most agricultural commodities is that effective branding of farm products themselves is quite difficult. However, in the case of some agricultural products brands do affect consumer purchase decisions. Makens, in his study on the influence of brands upon consumer preferences for turkeys concluded that 20 to 25 percent of the consumers were willing to pay a premium for a well known brand.¹ In all probability there are other agricultural products which would benefit in the way of added receipts from a product differentiation program. Without an evaluation of the effect of branding upon consumers' preferences, it will never be known whether the particular product would or would not benefit from a product differentiation program.

¹ J. C. Makens, "The Influence of Brands Upon Consumer Preference for Frozen Whole Turkey," (unpublished Ph.D. dissertation, Michigan State University, 1963).

The lack
described above
Hanson who re
at Michigan S

Currently
in numerous jo
has been no pr
the publication

The princip
summer preferenc
used to evaluate
nally food produ
and analyses ha
Others are estab
the field of cons
have been used

1 Helen Lou
Quality, A Report
College, July 1
of Agriculture (V
pp. 73-79.

Need for Establishing Techniques which can be used to Evaluate Consumers' Preferences Toward Products

The lack of established techniques for carrying out tests of the type described above has been noted by many people. Among them is Helen L. Hanson who revealed this belief at the Marketing Research Workshop held at Michigan State University in 1951.¹

Currently articles describing some consumer preference tests appear in numerous journals, notes, etc., but to the writer's knowledge there has been no previous attempt to describe the designs and the analyses in one publication.

Objectives

The principal objective of this thesis is to describe and evaluate consumer preference panel designs and types of data analysis which can be used to evaluate consumer preferences for all types of products, especially food products. Some of these consumer preference panel designs and analyses have originated with the writer and/or his associates. Others are established statistical techniques which had not been used in the field of consumer preference testing. Still other types of analysis have been used and were established by other researchers in this field.

¹ Helen Louise Hanson, Sensory Difference Tests in the Evaluation of Quality, A Report of the Marketing Research Workshop, Michigan State College, July 13 to 21, 1951, Prepared by the United States Department of Agriculture (Washington: Agricultural Research Administration, 1951), pp. 73-79.

A sec

tion of the
of the cons
opinion and
preferences,

The third
item one enc

The proce
evaluating the
This is done in
the problems to
Chapters IV thro
limitations of th
methods of analy
generalization purpo
Chapter IV in wh
contains an eval
presentation of
discussion of mo

1 James D.
Consumer Purch
State University

A second objective is to describe and evaluate the potential contribution of the laboratory and mass consumer preference panels. The purpose of the consumer preference panel, sometimes referred to as the product opinion and attitude panel, is to determine consumers' reactions, attitudes, preferences, and opinions concerning specific products.¹

The third objective is to discuss and evaluate the importance of problems one encounters in the organization of a consumer preference panel.

Organization of Thesis

The procedure followed in this study is that of first describing and evaluating the contribution of the various types of preference panels. This is done in Chapter II. Chapter III is devoted to an examination of the problems to consider in the organization of any preference panel. Chapters IV through VIII consist of an evaluation of the advantages and limitations of the various techniques of presentation and the various methods of analyzing the data obtained from such presentations. For organization purposes, the thesis is divided into the following chapters: Chapter IV in which one sample tests are discussed, Chapter V which contains an evaluation of two sample tests, Chapter VI which contains a presentation of three and four sample tests, and Chapter VII which is the discussion of more than four sample tests.

¹James D. Shaffer, "Methodological Bases for the Operation of a Consumer Purchase Panel," (unpublished Ph. D. dissertation, Michigan State University, 1952), p. 37.

determine if a p

It consists of a

must consider in

The studies

analysis were c

to fellow worker

observations pro

advantages of ea

Chapter VIII contains an evaluation of tests which can be used to determine if a panelist will pay more for his preferred product. Chapter IX consists of a discussion of some of the general problems which one must consider in consumer preference testing.

The studies which illustrate the use of each design and method of analysis were conducted by the writer unless a special reference is given to fellow workers who cooperated in a particular project. These empirical observations provided the basis for determining the advantages and disadvantages of each procedure.

There are
laboratory pa
laboratory pa
non to apprai
consumer pan
at his own res
mail questionn

The gener
native forms of
the different sa
similar establis
prices. An effe
sible. The cor
the samples, t
samples, and

CHAPTER II

TYPES OF PREFERENCE PANELS AND THE CONTRIBUTION OF EACH TYPE OF PANEL

There are essentially two types of consumer preference panels; the laboratory panel and the mass consumer panel. Preference testing in the laboratory panel consists of the panelists congregating at a central location to appraise the products being tested. Preference testing in mass consumer panels consists of researchers attempting to reach each panelist at his own residence or office. This may be done either by the use of a mail questionnaire or by a personal interview.

Laboratory Panels

The general procedure used in laboratory panels is to present alternative forms of the product to the panelists. The panelists may compare the different samples with each other, to a predetermined standard, to a similar established product or to their own personal preferences and prejudices. An effort is made to control as many of the variables as is possible. The controls include: lighting, uniform quality and preparation of the samples, the absence of outside distractions, temperature of the samples, and specified procedures.

There are two groups of judges, the trained and the untrained.

The trained judges use a standard testing procedure under consideration for testing. The method involves using pure chemical standards and measuring changes in flavor and odor over a period of time, reliability which is maintained at different times.

Thus, the trained judges (e.g., 1 to 10), these judges are trained according to fixed standards have been used for standards have been used.

The untrained judges are evaluating samples and preferences and members usually

¹ McCreary
² Washington
Industries, Vol.

Panel Membership

There are two types of panel membership; namely, the trained member and the untrained member.

The trained panel is composed of members who have been trained in testing procedures, who are familiar with the products and characteristics under consideration and who have had extensive experience in sensory testing. The members of the trained panel may be given threshold tests using pure chemicals to determine their acuteness of sensitivity to small changes in flavors. Their ability may also be tested for the degree of reliability which they demonstrate when evaluating the same samples at different times.¹

Thus, the trained panel involves the use of a small number of judges (e.g., 1 to 10), each expert in his knowledge of the particular product. These judges are usually carefully instructed to score the products according to fixed rules, which are called standards of perfection. These standards have previously been agreed upon, and only foods for which standards have been established can be scored.²

The untrained panel member has had less experience in testing and evaluating samples. His evaluation of samples is based upon his personal preferences and prejudices. Because of their lack of experience, untrained members usually have less sensitivity to small changes in flavors. This

¹McCreary, op. cit., p. 10.

²Washington Platt, "How Will Consumers Rate Your Product?" Food Industries, Vol. 9 (1937), p. 8.

type of pa

rather than

The un

ranging in

lack of expe

cording to c

must be few

standard of p

can be teste

previously de

The type

ratory panels

new or improve

sumer preferen

Quality Control

The trained

of quality contr

grading. 3

1. McCreary

2. Platt, or

3. McCreary

type of panel is designed to determine what the "typical" consumer prefers, rather than to detect minute differences between samples.¹

The untrained consumer panel involves many consumers with panels ranging in size from 25 to several thousand members. Because of their lack of experience consumers cannot be expected to score products according to complex fixed rules. Thus the rules presented to consumers must be few and very simple. The individual preferences provide the standard of perfection for each consumer. Preferences for any product can be tested in this type of panel, whereas only those products with a previously determined standard may be tested in a trained panel.²

Uses of Laboratory Panels

The type of panel used is mainly determined by its purpose. Laboratory panels can be used for (1) quality control, (2) the evaluation of new or improved products, and (3) the evaluation of branding upon consumer preferences.

Quality Control

The trained panel is usually used for work being conducted in the area of quality control. This type of work includes ranking, scoring, and grading.³

¹McCreary, op. cit., p. 9.

²Platt, op. cit., p. 8.

³McCreary, op. cit., p. 11.

When a s
with a predete
ard of perfect
jurity of disc
different from

Scoring o
been develop
to measure hi
arket testing

Grading i
re-determine
adequately ref

Evaluating New

Laboratory
testing. In thi
conjunction wit
technologist, c
in developing th
be prepared for
tent informatio
wants and desi

¹Platt, op

²McCrear

When a system of scoring is used, the panelists compare the samples with a predetermined standard of perfection. It is assumed that the standard of perfection set by scoring is that which is preferred by a great majority of discriminating consumers, but the standard may be considerably different from what the consumer prefers.¹

Scoring can be used only for those products for which standards have been developed. Thus, the technique gives the manufacturer a "yardstick" to measure his product by, but it has limited use, especially in the pre-market testing of new products.

Grading is similar to scoring in that the samples are compared to a pre-determined standard. However, these standards sometimes fail to adequately reflect the preferences of the consumer.

Evaluating New or Improved Products

Laboratory panels are frequently used as a method of pre-market testing. In this instance, the panels are conducted along with, and in conjunction with, the development of new products. The services of the technologist, chemist, engineer, or home economist are usually utilized in developing the product to a point where several alternative forms can be prepared for further testing. Laboratory panels may be used to supplement information gathered from preliminary investigations of consumers' wants and desires in guiding product development work.²

¹Platt, op. cit., p. 7.

²McCreary, op. cit., p. 10.

Both the
underlying a
"typical" co
representati
evaluates th
untrained pa
Evaluation o

Another
consumer pra
used in this
is representa
is designed s
consumers re

There are
to the laborato
control is obta
This control ov
of test conditi
studies. This
Thus, the res
between the c

Both the trained and the untrained panel member may be used. The underlying assumption is that the untrained panel member represents the "typical" consumer. However, the difficulty in getting a panel which is representative of the entire market is obvious. The trained member still evaluates the product on the basis of some predetermined standard. The untrained panelists base their judgment on their personal experience.

Evaluation of Branding Upon Consumer Preferences

Another use of laboratory panels is the evaluation of branding upon consumer preferences for various products. Only the untrained panel is used in this type of work. Again the objective is to obtain a panel which is representative of the market for the product being tested. The testing is designed so that it can be determined if a sufficiently large group of consumers reflect a preference among the different brands.

Advantages of Laboratory Panels

There are numerous advantages which have been mentioned in relation to the laboratory panel. The first such advantage is that a great deal of control is obtained over the conditions under which the tests are conducted. This control over other environmental conditions insures greater uniformity of test conditions and samples than is possible in mass consumer panel studies. This reduces the number of variables entering into the tests. Thus, the results obtained are more closely a reflection of the differences between the characteristics of the various samples themselves, rather

than a co

Conv

ducted in

ment and

most effio

personnel'

search met

evaluating

Anothe

to conduct.

product is r

session.

Still and

which paneli

different test

Finally i

ber are likely

person. Thus

accurate refle

test results.

¹ McCre

² Ibid.,

³ Max E.
Preferences,
Marketing Resea

⁴ McCre

than a combination of sample differences and external variables.¹

Convenience is the second advantage. The tests are usually conducted in one location, consequently there is no need for moving equipment and samples from place to place. That is, through scheduling, the most efficient use can be made of both the panel members' and the test personnel's time.² Thus, it appears that the controlled experimental research method is simple, fast, and one of the most direct methods of evaluating preferences toward a product.³

Another advantage is that laboratory panels are relatively inexpensive to conduct. Even for the larger consumer preference panels, the cost per product is relatively small because several products can be tested at each session.

Still another advantage is that the researcher can quickly determine which panelists are consistent in their rankings of identical products in different test series.

Finally it has been suggested that both the trained and untrained member are likely to be more attentive and conscientious than the "average" person. Thus, the results from the untrained panel are probably a more accurate reflection of their preferences, adding to the accuracy of the test results.⁴

¹McCreary, op. cit., p. 22.

²Ibid., p. 23.

³Max E. Brunk, "Discussion of Research on Consumer Behavior and Preferences," Market Demand and Product Quality, A Report of the Marketing Research Workshop, July 1951, Michigan State University.

⁴McCreary, op. cit., p. 22.

There are
The first such
operation in
testing bias
trained memb
consumer. T
panels may f
and education
members in la
in the Michig
erence panel,
group. While
consumers in
well families

Another d
cause difficult
develop, e.g.
uniform, and
testing period

¹Smith G
Dehydrated M
Vol. 5, No. 2

Disadvantages of Laboratory Panels

There are also several important disadvantages of the laboratory panel. The first such disadvantage is that it is difficult to achieve complete co-operation in the initially chosen sample, thus sample bias and subsequent testing bias may occur. That is, it seems valid to assume that the untrained members have more interest in testing products than the "normal" consumer. This would tend to bias the panel results. Thus, laboratory panels may fail to adequately reflect the distribution of age, sex, income, and education of the consumers in a given market. Careful selection of members in large panels may help to eliminate this defect. For example, in the Michigan State University-Wayne State University consumer preference panel, members were selected who fell into a particular selected group. While the panel cannot be considered representative of all the consumers in the Detroit area, it is believed that it does represent fairly well families with the characteristics of the group selected.¹

Another disadvantage is that maintaining experimental control may cause difficulties in some situations. A whole range of problems may develop, e.g., securing uniform samples, keeping preparation methods uniform, and maintaining sample quality over time when the length of the testing period is fairly long.

¹Smith Grieg and H. Larzelere, "Consumer Taste Preference Among Dehydrated Mashed Potato Products," National Potato Council News, Vol. 5, No. 2, (1954) p. 4.

T

ly far

which a

product

greater i

ence for e

ferences h

those expr

very point

the mass co

though certa

that no diffe

it does indic

appear to be.

One of th

tainty encoura

one setting.

from confusion

parative judgm

experiment de

¹ P. G. M.

Laboratory Typ

Vol. 9, (

² McGear

The conditions under which laboratory panels are carried out are usually far from natural. Samples are tested in small amounts under surroundings which are different from what they would be when the consumer uses the product in his home. However, the more natural the conditions, the greater is the number of variables which enter into the expressed preference for each sample. One might suspect that there would be some differences between preferences expressed under laboratory conditions and those expressed under home conditions. One study conducted to test this very point found that there was a general agreement in preferences between the mass consumer panel and the untrained member laboratory panel, although certain discrepancies did occur.¹ While this study is not a proof that no difference exists between the results of these two types of panels, it does indicate that the difference may be less than it may intuitively appear to be.²

One of the limitations of food flavor evaluation methods is the uncertainty encouraged in the comparative judging of more than five samples at one setting. The uncertainty arises from reduced acuteness, as well as from confusion or fatigue associated with making a large number of comparative judgments at one time. However, the usual practice is that an experiment designed to test several variables results in as many as ten or

¹P. G. Miller, J. H. Nair, and A. J. Harriman, "A Household and Laboratory Type of Panel for Testing Consumer Preference," Food Technology, Vol. 9, (1955) p. 449.

²McCreary, op. cit., p. 25.

twelve di

ples. In

manner tha

It has

icient as t

that continu

duces the re

members ma

would tend t

to the effect

in his work w

nenced judge

the same. It

professional i

Ortengren

tested through

answers. This

to give their o

some skill inv

¹ P. G. M

² George
Vol. 22, (195

³ John Ort
Marketing

twelve distinct samples with direct comparisons needed between all samples. In such a situation, the experiment should be designed in such a manner that no more than five samples are compared in any one session.

It has been pointed out that panel members tend to become more proficient as they become more experienced. Some researchers have found that continuous use of a panel develops a professional attitude which reduces the representativeness of the panel membership. Also, experienced members may acquire more acute senses of taste, smell and sight which would tend to make their judgment somewhat atypical.¹ It is not clear as to the effect experience actually does have on the results. Garnatz found in his work with the tenderness of meat that when matched pairs of experienced judges and inexperienced tasters are used, the results were about the same. It was concluded that evidently the judges did not become too professional in testing.²

Ortengren believes that "another indication of learning bias is manifested through a tendency to receive fewer "don't know" and "no preference" answers. This tendency appears when panel members are repeatedly asked to give their opinion about the same product, particularly when there is some skill involved in the answering."³

¹P. G. Miller, J. H. Nair, and A. J. Harriman, op. cit., p. 449.

²George Garnatz, "But What Do the Consumers Say?" Food Industries, Vol. 22, (1950) p. 1336.

³John Ortengren, "When Don't Research Panels Wear Out?" Journal of Marketing (April 1957), p. 442.

Another

from such

Thus, some

used to int

used when

results obt

opinions.

if not accou

conclusions

perimental d

the research

He must also

or control the

Finally t

sistent in the

yield biased r

During such a

placed on the

Mass co

of a large gro

¹ McCres

Another disadvantage of laboratory panels is that the results obtained from such panels are usually expressed in relative, not absolute, terms. Thus, some of the assumptions underlying many of the statistical analyses used to interpret panel results are not entirely met. Caution should be used when interpreting the results from laboratory panels. Sometimes the results obtained are based only on a relatively small number of respondents' opinions. There are also many forms of biases and testing errors which, if not accounted for, will influence the results and may lead to incorrect conclusions. A number of these errors and biases can be removed by experimental design and through appropriate statistical techniques. Thus, the researcher must be aware of the nature of these errors and biases. He must also be aware of the techniques which can be used to eliminate or control these errors or biases.¹

Finally tests used to determine if the majority of panelists are consistent in their rankings of identical products in different series may yield biased results if the tests are conducted in one short panel session. During such a session, panelists may be able to recall the rankings placed on the products in the previous series.

Mass Consumer Panels

Mass consumer panels are used to place the product into the hands of a large group of consumers. The objective is to obtain the consumers'

¹McCreary, op. cit., p. 26-27.

preference for the
panel essentially
the laboratory pro
comparison of the
of the test prod
with similar pro
this information
of consumers (2
answers may be
to the consumer

The mass o
members. In o
as they are use

There are
established or p
a small amount
panel is to us
replacement p
r may have b
alist who dro

preference for the product or products being tested. The mass consumer panel essentially provides the same type of information as is provided by the laboratory panel. The mass consumer panel yields information on the comparison of the test product with competing products and on the qualities of the test product that are its advantages or disadvantages in competing with similar products. When mass consumer panels are utilized to provide this information, it should be obtained from tests covering a large number of consumers (200 or more) using the product in their homes. Thus, the answers may be a reasonably accurate estimation of the market's response to the consumers preference toward the product being tested.¹

Panel Membership

The mass consumer panels usually consist of inexperienced panel members. In other words, mass consumer panels are untrained type panels, as they are used to determine what the "typical" consumer prefers.

There are essentially two types of mass consumer panels. In the established or permanent mass consumer panel the members are usually paid a small amount for their participation. The policy followed in this type of panel is to use the same panelists to test different products. A controlled replacement policy is followed to replace the panelist who may have moved or may have become a noncooperator. This consists of replacing the panelist who dropped out with a panelist who possesses the same characteristics.

¹McCreary, op. cit., p. 29.

The c

In this typ

period. A

Mass

that the un

may be util

cause the e

professiona

standards.

standards ad

consumer pa

When th

products and

to the panel n

writing or ora

market. In th

the consumer

of turkey he

etc. "This m

for the estab

¹Shaffer

²Ibid., :

The other type of panel is the ad hoc or one-time mass consumer panel. In this type of panel, panelists are selected to do testing during one time period. A new panel is selected for each new testing session.

Uses of Mass Consumer Panels

Mass consumer panels may be used to provide the same information that the untrained laboratory panel provides. The trained laboratory panels may be utilized to a greater extent in quality control work. This is so because the experts do much of the work in this area. This means a few professional tasters do the majority of the testing against predetermined standards. However, untrained panels may be used to determine if grade standards adequately reflect the preference of the consumer. Thus, mass consumer panels may be used indirectly in the area of quality control.

When the mass consumer panel is used to evaluate preferences toward products and services already on the market the product need not be sent to the panel member.¹ Instead the panelist may be quizzed, either in writing or orally, concerning products which he may have purchased on the market. In this type of questioning most of the emphasis is placed on what the consumer does. In other words, the panelist may be asked what brand of turkey he buys or why he buys one brand in preference to other brands, etc. "This method is used extensively to determine consumer preferences for the established brands."²

¹Shaffer, op.cit., p. 39.

²Ibid., p. 40.

The two

are the mail

Mail Question

The proo

sample of co

preferences f

may then be

instructions,

along with th

then return th

tions in this p

portunity for c

if the questio

needed in orde

wording of the

Causes of No

Many the

to answer ma

researchers a

if some recip

tion of inter

¹ McCre

Methods of Conducting Mass Consumer Panels

The two methods which are used to conduct mass consumer panel tests are the mail questionnaire method and the personal interview method.

Mail Questionnaire Method

The procedure used in the mail questionnaire method is that a random sample of consumers is selected. If the test objective is to determine preferences for several products, samples of the products being compared may then be mailed to the selected households. A letter of introduction, instructions, a questionnaire and a return addressed envelope are sent along with the samples. The panelists test the sample as directed and then return the completed questionnaire through the mail. Several variations in this procedure may occur. Because there is not face-to-face opportunity for clarification, simple but clear wording of the directions and of the questionnaire is important in the mail-in-method. Caution is also needed in order not to force biased responses through the structure and wording of the questionnaire.¹

Causes of Non-response

Many theories have been evolved to explain the failure of some people to answer mail inquiries. Among the most common reasons proposed by researchers are the lack of interest in the subject, the unresponsive mood of some recipients, lack of understanding of the questions asked, suspicion of intentions, and the mental chore involved. However, Robinson and

¹McCreary, op. cit., p. 39.

Agis-

respo

being

being p

importan

How to C

Robi

planning

the produc

people mos

job. He sh

answer. Th

questions to

turned over t

suggestions.

The seco

tive, easy to

simple and at

understood. I

through the us

¹
R. A. R
able, " Journal

²
Ibid.

Agisim conducted a study which indicated that the major causes of non-responsiveness are: either mislaying or forgetting about the questionnaire, being absent from home at the time the questionnaire was received, or being preoccupied with other matters.¹ Such findings lend weight to the importance of a follow-up mailing with an additional questionnaire enclosed.

How to Obtain High Returns from Mail Surveys

Robinson and Agisim have formulated six rules that are helpful in planning mail surveys.² The first rule involves thoroughly understanding the product of field being researched. The researcher should talk to the people most concerned or familiar with the problems and background of the job. He should discard questions that respondents are not competent to answer. The next step is to prepare a preliminary questionnaire of the questions to be asked. The preliminary questionnaire should then be turned over to experts in the field for their reactions, criticisms, and suggestions.

The second rule consists of composing a questionnaire that is attractive, easy to fill out, and of interest to all receiving it. This means using simple and at the same time interesting questions, briefly stated and easily understood. If the questionnaire is long it can be made to appear brief through the use of attractive type, good layout, and plenty of open space.

¹R. A. Robinson and Philip Agisim, "Making Mail Surveys More Reliable," Journal of Marketing (April 1951), pp. 415-424.

²Ibid.

Dressi

page ma

nionnaire

experime

the scale

paper stood

color, with

The th.

respondents

average con

are difficult

replies, what

too long, etc.

Rule num

sure, or induc

Agisim claim t

returns of 5 to

warn that any

as a magnet t

quarter-dollar

¹Ibid.

²R. P. E.
University Pr

Dressing up the questionnaire increases returns.¹ An attractively printed page made up on the suitable stock and color selection gives the questionnaire more eye appeal. On the subject of colored paper stock, an experiment conducted by Eastwood indicated that dark colors rank low in the scale of effectiveness while light colors rate high. Yellow-colored paper stock was found to bring a higher percentage return than any other color, with pink running second.²

The third rule is that of pretesting the questionnaire to guard against respondents' misunderstanding. This pretesting should be done among average consumers. It is in this phase that one finds out which questions are difficult to answer, which ones will produce biased or ambiguous replies, what terms can be misinterpreted, whether the questionnaire is too long, etc.

Rule number four consists of determining the kind of premium, enclosure, or inducement which will pull the maximum response. Robinson and Agisim claim that a well-chosen premium can mean the difference between returns of 5 to 10 percent and responses of 70 to 80 percent. They do warn that any premium which attracts a certain class of respondents acts as a magnet to draw a distorted sample. They recommend using the U.S. quarter-dollar as an inducement for increasing returns, and indicate that

¹Ibid.

²R. P. Eastwood, "Sales Control By Quantitative Methods," (Columbia University Press, 1940) pp. 288-293.

this prac

parently

value of

enclosure

tion to the

nique of e

turns are s

required.

The fif

go along wi

letter, so i

thought. To

in warm, hu

Rule six

up reminders

questionnaire

respondent to

stressing the

survey. If po

closed becau

Robinson

to secure from

R. A. R

this practice has met with considerable success. The quarter-dollar apparently produces about as high a return as a dollar bill. It is not the value of the enclosure, but the psychology of it that counts. However, enclosures of stamps do not increase the rate of return. A common objection to the use of premiums is that it is too expensive. Actually the technique of enclosing a quarter dollar may cost less per return, because returns are so much greater that only a fraction of the normal mailing is required.

The fifth rule involves the development of the right kind of letter to go along with the inquiry. Returns can be increased by the right kind of letter, so it is important to give this phase of the operation considerable thought. To make up for cold paper and ink, the letter should be couched in warm, human, friendly, and very appreciative language.

Rule six consists of following through the original mailing with follow-up reminders. These reminders are sent several days after the original questionnaires have been mailed out. The follow-up letter may remind the respondent to return the questionnaire in case it has slipped his memory, stressing the importance of the addressee's answer to the success of the survey. If possible an additional copy of the questionnaire should be enclosed because the first copy may have been mislaid.

Robinson and Agisim believe that with modern methods it is possible to secure from 70 to 93 percent returns on mail surveys.¹ They believe

¹R. A. Robinson and Philip Agisim, op. cit., pp. 415-424.

that the

be re-ex

findings

between

inherent v

cluded the

is given to

effect on t

Advantages

The ma

large number

comparison

searcher wh

second adva

in a short pe

The mail

the products a

test allows th

non, eye-app

The mail

on the basis o

of homemaker

panel technic

that the old question of biased sample from surveys made by mail needs to be re-examined in the light of the new techniques discussed above. Their findings indicate that with high level returns the differences that existed between respondents and non-respondents were due to chance and not to inherent variations in characteristics of the two sub samples. They concluded that when returns reach the neighborhood of 80 percent, reliability is given to the findings because non-respondents would have little, if any, effect on the total.

Advantages of the Mail Questionnaire Method

The mail method offers the researcher the opportunity of reaching a large number of respondents at a relatively low cost. Its low cost, in comparison with the personal interview method, should appeal to the researcher who is striving for optimum allocation of the research dollar. A second advantage is that the method is very simple and can be conducted in a short period of time.

The mail questionnaire method also allows the consumer to evaluate the products after they are prepared and used. The mechanism of such a test allows the panelist's preference to be influenced by ease of preparation, eye-appeal of the package, keeping qualities, etc.

The mail method can yield accurate data if the panels are "controlled" on the basis of geographic census division, population distribution, age of homemaker, and total family income. O'Dell concludes that the mail panel technique merits consideration if an estimation of trend or direction,

rather than

the project

Disadvanta

A series

occurs in s

respondent.

those who r

those who d

rate may be

this case th

rate ranges b

A further

asked in the

dig beneath t

preference.³

There app

of the informa

it was found t

¹William
of Marketing

²Trienah
Economics, V

³McCrea

rather than an estimation of a parameter, is all that is required to fulfill the projects objectives.¹

Disadvantages of the Mail Questionnaire Method

A serious disadvantage of this method is the low return rate which occurs in some studies. It is believed that the failure of some of the respondents to cooperate introduces a source of bias. It is claimed that those who return the questionnaire have different characteristics than do those who do not answer.² It should be pointed out that the response rate may be much higher if an established or permanent panel is used. In this case the low return problem is virtually eliminated as the response rate ranges between 80 and 90 percent.

A further disadvantage is that only a few simple questions can be asked in the mail method. This does not allow the researcher to really dig beneath the surface to find out the reasons behind the respondent's preference.³

There appears to be some difference of opinion as to the reliability of the information gained through the use of the mail method. In one study it was found that mail returns did have a high degree of representativeness

¹William F. O'Dell, "Personal Interviews or Mail Panels?" Journal of Marketing, Vol. 26, (October 1962) pp. 34-39.

²Trienah Meyers, "Predicting Market Acceptance," Journal of Farm Economics, Vol. 37 (1955), pp. 1387-94.

³McCreary, op. cit., p. 40.

of the entire
a statistical
acteristics b

A "sequ
panelists can
question. Th
the panelist'
distract the r

If the ho
"universe," t
when the indi

Another o
of testing the
this time lap

Then too
of any two p
acceptance o
know the cor

¹A. Gre
nonnaires in

²J. D. 3
The Journal

³Raymo
Acceptance,

⁴Ibid.

of the entire sample.¹ On the other hand, a study by Shaffer indicated a statistically significant difference in the estimates of population characteristics between the mail and personal interview methods.²

A "sequence bias" may be a problem in a mail panel survey, because panelists can read through a questionnaire before answering the first question. The objective of the questioning may become evident and bias the panelist's response. Disguising the subject of the questionnaire may distract the respondent but it also complicates the questionnaire.

If the housewives or other individual family members are the defined "universe," the mail panel method can elicit a moderation of opinion when the individual respondent consults other members of the family.

Another disadvantage is that the time elapsed since the experience of testing the product always covers a wide range and estimates made at this time lapse are faulty.³

Then too, the additional confusion arising from the sequence of use of any two products makes a statistical treatment of their comparative acceptance difficult.⁴ To add to the confusion, the researcher doesn't know the conditions under which the panelist may have tested the samples.

¹A. Greenberg and M. N. Manfield, "On the Reliability of Mail Questionnaires in Product Tests," The Journal of Marketing, Vol. 21 (1956), pp. 342-345.

²J. D. Shaffer, "Estimating Population Characteristics by Mail Survey," The Journal of Farm Economics, Vol. 41 (1959), p. 836.

³Raymond Franzen and Darwin Teilhet, "A Method for Measuring Product Acceptance," The Journal of Marketing, Vol. V (October 1940), pp. 156-161.

⁴Ibid.

In this

The interview

sample and

viewer must

variations of

face contact

Advantages

The per

to be "general

obtain more

response than

Disadvantages

The pers

schedule than

able cost advantage

study.² How

per completed

Another

is always the

into the results

¹ Rose M.
Quality of C

² J. D.

Personal Interview Method

In this type of panel an interviewer contacts each of the panelists. The interviewer may conduct the entire test himself, or he may leave the sample and questionnaires with the respondent. In this case the interviewer must return later to pick up the questionnaire. There are several variations of this method but all have a common element, namely face-to-face contact between the panelist and the interviewer.

Advantages of Personal Interview Method

The personal interview method is considered by several researchers to be "generally preferable to the mailing technique."¹ It is possible to obtain more detailed information and to obtain a higher percentage of response than by using the mail-in method.

Disadvantages of the Personal Interview Method

The personal interview method is usually more expensive per completed schedule than the mail method. A study by Shaffer indicated a considerable cost advantage in favor of the mail method for one particular type of study.² However, if the interviews are conducted by telephone the cost per completed schedule may be reduced sharply.

Another disadvantage of the personal interview method is that there is always the possibility of bias through interviewer attitudes entering into the results. This bias can be partially eliminated by careful interviewer

¹Rose Marie Valdes and E. G. Roessler, "Consumer Survey on Dessert Quality of Canned Apricots," Food Technology, Vol. 10 (1956), p. 481.

²J. D. Shaffer, op. cit., p. 836.

training and

results due

interpretati

The ex

conducted

was found

13.5 perce

greatest so

bias, origin

Severa

for selectin

biases acro

tend to sele

as having re

is that in an

from the inte

Another

questions on

during the in

¹McCrea

²J. F. M
tion Study, "
Agricultural

³Roy E.
Interviewer
(1953), pp. 2

⁴O'Dell

training and strict procedures. Other errors can also be included in the results due to inaccurate recording of answers by interviewers or by misinterpretation of answers by interviewers.¹

The extent to which an answer may occur is illustrated by a study conducted by Metz. In this study designed to detect answer error, it was found that the total milk purchased as reported by household was 13.5 percent greater than the amounts shown on dealer's records. The greatest source of bias appeared to be the person interviewed. This bias, originating with the interviewer, was found to be far less.²

Several researchers suggest that interviewers use non-random bases for selecting sample households, and that there is consistency in these biases across interviewers.³ Their evidence indicates that interviewers tend to select households which they "liked" and which they perceived as having relatively high incomes. A clear-cut implication of their study is that in any survey, the selection of households needs to be removed from the interviewer's control.

Another limitation of the personal interview method is that personal questions on ego-challenging subjects may evoke noncommittal answers during the interview.⁴

¹McCreary, op. cit., p. 42.

²J. F. Metz, Jr., "Accuracy of Response Obtained in a Milk Consumption Study," Methods of Research Marketing, Paper No. 5, Department of Agricultural Economics, Cornell University (1956).

³Roy E. Carter, Jr., Verling C. Troidahl, and R. Smith Schuneman, "Interviewer Bias in Selecting Households," Journal of Marketing (April 1963), pp. 27-34.

⁴O'Dell, op. cit., p. 39.

The

Panel.

population

properly, t

of that whi

similar man

It is e

population

describe the

for selectin

eralizations

to extrapolat

the definition

The sam

tistical popu

panelists, o

lar type of p

On the other

CHAPTER III

PROBLEMS TO CONSIDER IN PANEL ORGANIZATION

The Problem of an Adequate and Accurate Sample of Panelists

Panels are drawn to provide otherwise unknown information about the populations of which they are samples. If such research is conducted properly, the panel provides information which is representative or typical of that which would be found if the whole population were studied in a similar manner.

It is especially important for one to define carefully the statistical population he wants to study. This should be done in precise terms which describe the population in detail. The researcher must then devise a plan for selecting elements from the population. This design must allow generalizations about the population being studied. One must be careful not to extrapolate his findings to be representative of groups not included in the definition of the population.

The Problem of Representativeness

The sample of panelists selected must be representative of the statistical population which is being evaluated. In selecting a sample of panelists, one might confine his attention to present users of the particular type of product. One might attempt to survey potential users as well. On the other hand, a true cross-section of the population might be chosen.

One might

answer dif

selections

how usele

ago in det

that time v

the number

throughout

It app

selected s

for the proo

should be s

economic g

age group.³

the character

will be, sol

In order

usually nece

drawn by sim

to be studie

¹ John H.
pp. 131-136

² Platt,

³ Nair,

One might well ask, "Just where shall the dividing line be drawn?" The answer differs for different products. Here it is pertinent to observe that selections confined to users only, are apt to be misleading. Consider how useless such a panel would have been to researchers several years ago in determining preference for a new dry soup mix. Actual users at that time were comparatively few and confined to limited areas. Today the number of users is much larger, and these users are distributed throughout the world.¹

It appears that the number and distribution of the consumers to be selected should bear a definite known relationship to the potential market for the products being tested.² Ideally, the composition of the panel should be such that it gives a proportional sampling by: family size, economic group, city size, sex, nationality groups, education level, and age group.³ In short, what is desired is a panel that accurately reflects the characteristics of the market population to which the product is, or will be, sold.

In order to select a panel which meets the above requirements, it is usually necessary to present a preliminary questionnaire to a sub-sample drawn by simple random or systematic sampling techniques from the sample to be studied.

¹John H. Nair, "Mass Taste Panels," Food Technology, Vol. 3 (1949), pp. 131-136.

²Platt, op. cit., pp. 7-11.

³Nair, op. cit., p. 134.

Simple

entire stat

ples has a

method of

has been s

unit will b

ple drawn

System

pling. Eac

To select a

at random f

at random a

selected is

For example

sequent unit

unit determin

is that it is

can be used

(e.g., an urb

When us

list from whi

¹William
New York and

²Ibid.,

Simple random sampling is a method of choosing n units out of the entire statistical population (N) such that every one of all possible samples has an equal chance of being chosen.¹ In the most commonly used method of simple random sampling without replacement, the unit which has been selected is not replaced in the population from which the next unit will be selected. Thus, no unit can appear more than once in a sample drawn by this method.

Systematic sampling is somewhat different from simple random sampling. Each of the units in the population may be numbered from 1 to N . To select a panel with a sample size of n equal $\frac{N}{k}$, a number is selected at random from $1 \dots k$. The sample then consists of the number selected at random and every subsequent k th number. Thus, if the random number selected is r , then the sample consists of $r, r + k, r + 2k, r + 3k$, etc. For example, if k is 25 and if the first unit drawn is number 13, the subsequent units are numbers 38, 63, 88, etc. The selection of the first unit determines the entire sample.² The chief advantage of this method is that it is easy to draw a sample by this method. Systematic sampling can be used where the ordering of the population is essentially random, (e.g., an urban telephone directory may be an example).

When using systematic sampling it is crucial to recognize that the list from which the panelists' names are drawn should be random with

¹William G. Cochran, Sampling Techniques, John Wiley & Sons, Inc., New York and London (1953), p. 11.

²Ibid., p. 160.

respect to

tory is no

random wi

of the pop

A res

appropriat

telephone

phone dire

These grou

not afford

have unlist

dwellings v

The selecti

represent th

age children

The pre

individuals v

the definition

with the Kro

the highest a

rate that the

wealthy lack

fill out ques

¹C. L.

Vol. 13 (May

respect to the characteristics investigated. In general, a telephone directory is not a random list of consumers; however, it may be reasonably random with respect to food preferences. It also provides a current listing of the population.

A researcher must also recognize that a telephone directory is the appropriate list to use in selecting a panel only if the universe is that of telephone subscribers. When a researcher selects a sample from a telephone directory, he eliminates three groups of people from the frame. These groups are: (1) the extremely poor people who believe that they cannot afford a telephone, (2) the extremely wealthy or famous people who have unlisted phone numbers, and (3) some single people who live in dwellings which contain only one phone for a large number of residents. The selection of a sample from a telephone directory also tends to over-represent the middle and upper class families, especially those with teenage children who have more than one telephone in their home.

The preliminary questionnaire makes it possible to screen out those individuals who do not possess the characteristics desired according to the definition of the population. For example, as a result of his experience with the Kroger Food Foundation, Arnold believes that for taste testing, the highest and lowest income groups may be omitted. His results indicate that these people all have nearly normal taste reactions, but that the wealthy lack a genuine interest, and many of the poor lack the ability to fill out questionnaires intelligently.¹ Cowan, manager of Commercial

¹C. L. Arnold, "Do Consumers Have Good Taste?" Food Industries, Vol. 13 (March 1941), p. 45.

Research C

customed t

He believe

while the

While the

panelists

sults do n

for all typ

Resea

greatly. H

icantly low

age range a

probably no

Both se

due to a sin

¹ Cowan

² J. W.

stuffs by Lab
Technology

³ K. Ben

Laboratories

⁴ D. Kn

Perspective
pp. 207-216

⁵ Elsie

Differences
U.S.D.A.,

Research Co., believes that the wealthy classes are apt to be more accustomed to perceiving delicate and varied flavors, aromas and colors. He believes further that poor people are likely to be at the opposite extreme, while the middle income consumers may provide more dependable results.¹ While these results indicate that panels containing only middle income panelists can be used in studying preference for most products, the results do not indicate that such panels can be used to study preferences for all types of products.

Researchers' preferences as to desirable age ranges for panelists vary greatly. Hopkins believes that persons under 30 years of age have significantly lower taste sensitivity.² Bengtsson and Helm consider the optimum age range as 30 to 40 years.³ Knowles and Johnson conclude that there is probably no correlation between ability to identify primary tastes and age.⁴

Both sexes are usually used in panels. "Taste deficiency is primarily due to a single recessive gene, not sex-linked nor sex influenced."⁵

¹ Cowan, op. cit., p. 42.

² J. W. Hopkins, "Precision of Assessment of Palatability of Food-stuffs by Laboratory Panels," Canadian Journal of Research, Section F, Technology 24 (1946), pp. 203-214.

³ K. Bengtsson and E. Helm, "Principles of Taste Testing," Wallerstein Laboratories Communication, Vol. 9 (1946), pp. 171-180.

⁴ D. Knowles and P. E. Johnson, "A Study of the Sensitiveness of Perspective Food Judges to the Primary Tastes," Food Research, Vol. 6, pp. 207-216.

⁵ Elsie H. Dawson and Betsy L. Harris, Sensory Methods For Measuring Differences in Food Quality, Agriculture Information Bulletin No. 34, U.S.D.A., Washington, D. C. (1951), p. 13.

Snyder reports

the children can

One research
sensitivity.²

in identifying

Thus, one
stifying prefer

In determining
consumer
sary to est
be as large
a general
contrasting
tified unde

One way
tion to the nu
select the fir
allocation.
the preliminary
dividing the
 $N_1, N_2, \dots, N_k$

¹ L. H.
(1931), pp.

² Hopki

³ K. G.
International

⁴ Cowa

Snyder reports that when neither parent can taste the compound, none of the children can.¹

One researcher reports that females have significantly lower taste sensitivity.² However, another investigator found that men tend to excel in identifying solutions by taste, while women excel in identifying odor.³

Thus, one can see that valuable information can be derived by classifying preferences according to the characteristics of the panelists.

In determining these conflicting tendencies on the part of different consumer groups, it is important that the number of people necessary to establish a difference in preference on account of age must be as large for each age group as would be necessary to establish a general preference. As a consequence, the establishment of contrasting group preferences is more expensive but may be justified under particular circumstances.⁴

One way to insure that each consumer group is represented in proportion to the number of people in that group in the entire population is to select the final sample by stratified random sampling with proportional allocation. This selection may be based upon the information obtained in the preliminary questionnaire. Stratified sampling is the technique of dividing the population of N units into subpopulations (called strata) of $N_1, N_2, \dots, N_L$ units. These strata are non-overlapping, and together they

¹L. H. Snyder, "Inherited Taste Deficiency," Food Science, Vol. 74 (1931), pp. 151-152.

²Hopkins, op. cit., pp. 203-214.

³K. G. Weckel, "Put Your Taste Buds to Work on Your Flavor Problems," International Association Milk Dealer Bulletin 34, pp. 136-140.

⁴Cowan, op. cit., p. 44.

comprise the wh
determined, a s
between the str
stratum. The s
Stratified random
in each stratum

Stratified r
of the fact that
The stratified r
introduced by n

When the p
MSU-WSU pane
This policy con
come a non-coo
characteristics
a preliminary p
permanent pane
and (3) replaci
panelist from th
panelist who is

¹ Cochran,
² James D.
Consumer Purch
State University

comprise the whole of the entire population. When the strata have been determined, a sample of a size which bears a proportional relationship between the stratum size and the size of other strata is drawn from each stratum. The selections are made independently in each different strata. Stratified random sampling is the method of taking a single random sample in each stratum.¹

Stratified random sampling has advantages over other methods because of the fact that not all of the initial contacts will become sample members. The stratified random sampling method tends to compensate for the bias introduced by non-cooperation.²

When the panel is an established or permanent panel, such as the MSU-WSU panel, the policy of controlled replacement may be followed. This policy consists of replacing a panelist who may have moved or become a non-cooperator with another panelist who possesses the same characteristics. Thus, the over-all plan would consist of: (1) selecting a preliminary panel by random or systematic sampling, (2) selecting the permanent panel by stratified random sampling with proportional allocation, and (3) replacing the panelists who drop out by drawing randomly a new panelist from the panelists possessing the same characteristics as the panelist who is being replaced.

¹Cochran, op. cit., pp. 65-110.

²James D. Shaffer, "Methodological Bases for the Operation of a Consumer Purchase Panel," (unpublished Ph.D. dissertation, Michigan State University (1952), p. 194.

While it is
it is not always
sample. Prelim
are probably ju
for many produ
population.¹ T
somewhat less
that the more r
be.

Some indiv
or unwilling to
the responding
cooperators' pr
the answer to t
"no a priori rec
traits that prev
evaluate the qu
preference test

¹ Hunter, 'Marketing Rese
² Ibid., p.

While it is desirable to have as representative a sample as possible, it is not always economically feasible to achieve a purely representative sample. Preliminary surveys and screening are an added expense which are probably justifiable only to a certain point. An added problem is that for many products it is difficult to identify the exact relevant consumer population.¹ Thus, the sample being used in a particular test may be somewhat less than representative. However, it should be remembered, that the more representative the sample, the more valid the results should be.

The Problem of Non-cooperators

Some individuals, when contacted for panel membership, are unable or unwilling to cooperate. This raises the question of how representative the responding panel members are of the total population. Since non-cooperators' preferences are not known, there is no actual way to find the answer to this question. However, it has been stated that there are "no a priori reasons for believing that the circumstances or personality traits that prevent a person from participating are related to his ability to evaluate the quality characteristics of a product."² In one mass consumer preference test conducted by the U.S.D.A., researchers found that there

¹Hunter, "Consumer Preference for a 6 to 1 Applejuice Concentrate," Marketing Research Report No. 343 (Washington: U.S.D.A. AMS), 1959.

²Ibid., p. 16.

were differ

non-partici

been under-

ever, it cou

had on the r

Shaifer

consumer pur

more coopera

cooperators s

groups within

sample of coo

the middle-cla

Although S

of household ci

this did not cau

considered for t

the total sample

Although Sh

for a purchase p

erence panel, f

¹ Ibid.

² Shaifer, c

³ Ibid., p.

were differences in the background characteristics of the participators and non-participators.¹ In this study it appeared that certain groups may have been under-represented because of their unwillingness to cooperate. However, it could not be shown what effects, if any, this under-representation had on the results of the survey.

Shaffer found that in the selection of panel members for the M.S.U. consumer purchase panel, households with certain characteristics were more cooperative than other households.² He concluded that a sample of cooperators selected in an unrestricted manner would not represent all groups within the population in the proper proportion. Thus, an unrestricted sample of cooperators could be expected to be strongly biased in favor of the middle-class household.

Although Shaffer noted a considerable difference in the distribution of household characteristics between the cooperators and non-cooperators, this did not cause any great bias in the average consumption of the products considered for the sample of cooperators as compared with the average for the total sample.³

Although Shaffer's findings were obtained while selecting panelists for a purchase panel, they also apply to the selection of a consumer preference panel, for the initial process of selection is the same in both cases.

¹Ibid. ,

²Shaffer, op. cit., p. 181.

³Ibid. , p. 182.

TI

panel i

evaluate

after the

panel. S

same type

panel mem

the panel.

A suffici

of prejudice c

to produce sta

acteristics. P

are very notice

it is necessary

fore, "the numb

proximated by u

able superiority

permissible erro

¹ Shaffer, C

² Cowan, C

The researcher must be aware of still another source of bias if the panel is an established or permanent panel which is regularly used to evaluate products. This bias is a result of panelists who fail to cooperate after they have indicated that they would be permanent members of the panel. Shaffer, in his study on consumer purchase panels, found that the same types of families which prove to be the problem in recruiting for panel membership are the same types which are most likely to resign from the panel.¹

Problem of Adequate Panel Size

A sufficient number of panelists should be used to level out the peaks of prejudice or valleys of opinion. The exact size of the panel necessary to produce stable results depends on the contrasts in the product's characteristics. Reactions will be more reliable if the product's characteristics are very noticeable. When the product's superiority is less pronounced, it is necessary to base conclusions on a great many more tests. Therefore, "the number of testers required for acceptable results may be approximated by using the results of preliminary tests to estimate the probable superiority of one product over another, and by deciding upon the permissible error in advance."²

¹Shaffer, op. cit., p. 177.

²Cowan, op. cit., p. 43.

For exam
wish to know
(or vice versa
sure that his
may also ass
sample drawn

Thus, P is
Since p is ass
 $(\bar{p} \pm 1.96 S_p)$,
may write 1.96
that if the rese
for A and B (e.
specifications
their values are
Hence $n = 384$,
requirements.
product B is ind
70 percent of th
 $\bar{p} = .7$ and $(1 - \bar{p})$
In this case, n

¹ The paired
² Cochran,
³ Ibid., p.

For example, a researcher in charge of a paired comparison test may wish to know the percentage of people preferring product A over product B (or vice versa).¹ He may specify that he is satisfied if he is 95 percent sure that his estimate ± 5 percent contains the true parameter value. He may also assume that the population is more than 20 times as large as the sample drawn, and that the sample percentage, p , is normally distributed.²

Thus, P is to lie in the range $(p \pm .05)$ except for a 1 in 20 chance. Since p is assumed normally distributed about P , P will lie in the range $(p \pm 1.96 S_p)$, apart from a 1 in 20 chance.³ Now $S_p = \sqrt{\frac{p(1-p)}{n}}$; thus he may write $1.96 \sqrt{\frac{p(1-p)}{n}} = .05$ or $n = \frac{3.84 p(1-p)}{.0025}$. One may now see that if the researcher believes there is no difference in the preferences for A and B (e.g. $p=.5$), a larger sample would be required to meet the specifications cited above. In this case, $p=.5$ and $1-p=.5$, hence their values are substituted in the formula to give $n = \frac{3.84 (.5) (.5)}{.0025}$. Hence $n=384$, and a sample size of 384 is needed to meet the specified requirements. However, if the probable superiority of product A over product B is indicated by previous tests, the researcher may believe that 70 percent of the people prefer A over B. Consequently, the values $p=.7$ and $(1-p) = .3$ are substituted in the formula giving $n = \frac{3.84 (.7) (.3)}{.0025}$. In this case, $n = 322$.

¹The paired comparison test procedure is discussed in Chapter IV.

²Cochran, op. cit., p. 17.

³Ibid., p. 51.

The research
size will be, b
He may be con
will give the m
lined above. R
may differ from
satisfy the spe
involves the dr
estimators of P
structed from th
narrow enough t
is terminated.
tions, addition
constructed fro
terminated if th
ifications. If t
specified requi
val is shortene

Sim

The time i
should be cons
used to evalua
product must b

The researcher does not usually know how large the required sample size will be, because he does not know the value of the true parameter P . He may be conservative and estimate p to be .5, because this estimate will give the maximum required panel size to meet the specifications outlined above. However, this method may be costly, for the true parameter may differ from .5. If this were the case, a smaller panel size would satisfy the specified requirements. The method of sequential sampling involves the drawing of a small sample to obtain p and S_p , which are estimators of P and σ_p respectively. A confidence interval is then constructed from the estimators of p and σ_p . If the confidence interval is narrow enough to meet the specified requirements, the sampling process is terminated. If the confidence interval is too wide to meet the specifications, additional observations are made. A confidence interval is again constructed from the estimates of P and σ_p . The sampling process is terminated if the confidence interval is narrow enough to meet the specifications. If the confidence interval is still too wide to satisfy the specified requirements, sampling is continued until the confidence interval is shortened enough to meet these requirements.

Simultaneous or Single Stimulus Sample Presentation

The time interval between judging samples is another problem which should be considered when organizing an untrained panel which is to be used to evaluate products. This is a more serious problem when the product must be tasted. If the products are to be evaluated on the basis

of their appe
a product fro
involved is v
comparison,
(single stimu

If the pro
single stimulu
rates the rea

There are t
select a m
erences for
single stim
for consum
home at a g
product--th
erence judg
in the meth
taste prefer
tion.

The second
method of s
of the food
It has been
that three it
design when
sign, four it
five or more
for a method

¹Robert M.
search," Paper
Washington,

²Forest E.
of Single Stimu
Psychology, V

of their appearance, the problem is less serious as the consumer selects a product from many competing products at the retail store. The question involved is whether samples should be presented simultaneously for direct comparison, or presented individually with a time interval between tests (single stimulus).

If the products must be tasted, several researchers contend that the single stimulus method is more realistic, for it "more closely approximates the real life situation in which opinions are formed."¹

There are two fundamental considerations that would lead one to select a method of single stimulus approach in research on preferences for variations of a food product. First, the method of single stimulus is a duplication of the situation that is typical for consumers. Seldom does the consumer have available in his home at a given time several variations of a particular food product--the situation that would be conducive to making preference judgments based upon the direct comparisons involved in the method of comparative judgments. Realistic research on taste preference should attempt to utilize the actual home situation.

The second factor that will force the experimental design into a method of single stimulus model is the number of the variations of the food product being tested, since adaptation is a variable. It has been our experience, in both laboratory and home situations, that three items is the maximum number for a paired comparison design when testing occurs in one session; with a rank order design, four items is the maximum. When the number of items is five or more, comparative judgment models should be abandoned for a method of single stimulus design.²

¹Robert M. Walsh, "What Is and What Is Not Good Consumer Research," Paper presented at the Twelfth Reciprocal Meat Conference (Washington, U.S.D.A., A.M.S., 1959), p. 2.

²Forest E. Clements, James A. Bayton, and Hugh P. Bell, "Method of Single Stimulus Determinations of Taste Preference," Journal of Applied Psychology, Vol. 38, No. 6, 1954, pp. 446-451.

Advocat

prove upon t

presentation

uated, a con

ucts must be

to estimate th

high-flavored

is therefore g

or six times.

pair, for the s

product with t

Franzen a

consumer is s

and the contro

over a sufficie

ciality of askin

procedure acco

bite or two is

a whole serving

when the foods

or milk.² In j

¹ Franzen

² Washing
Industries (Ma

Advocates of the "extended use" panel claim it is an attempt to improve upon these disadvantages of the simultaneous method of sample presentation. They believe that, in order for a product to be fully evaluated, a consumer use test of long duration and repeated use of the products must be run to allow the novelty effect of the product to wear off and to estimate the fatigue effect. This is especially important in the case of high-flavored products and new concepts. The untrained consumer panel is therefore given the same pair of products to evaluate as often as five or six times. The panelist may not be aware that he is judging the same pair, for the samples are usually coded differently. A swing away from a product with time or education of taste toward a product is then detectable.

Franzen and Teilhet advocate the use of a procedure in which the consumer is supplied with sufficient quantities of both the test product and the control product to permit their use under normal conditions and over a sufficient period to establish preference.¹ This avoids the artificiality of asking for "expert" reactions to hypothetical situations. Their procedure accounts for the fact that sometimes a food is preferred when a bite or two is taken, whereas one would quickly tire of it if he had to eat a whole serving. This problem appears in a still more pronounced form when the foods being tested are eaten every day, such as bread, coffee, or milk.² In judging these foods, the question as to whether one will tire

¹Franzen and Teilhet, op. cit.

²Washington Platt, "Why Consumer-Preference Tests?" Food Industries (March 1941), pp. 40-41.

of them on con

where this situ

sample present

Other rese

presentations t

various sample

not compare th

he compares th

samples combi

Another di

that the untrain

the samples the

Schwartz a

effect the time

tester's respon

pair was presen

of single stimu

7, and 10 days

obtained under

samples was le

to diminish.

1 "Consum.
Grades," Mark

2 M. Schw
tra Paired Com

of them on constant repetition is of the utmost importance. In instances where this situation is likely to occur, the single stimulus method of sample presentation should be utilized.

Other researchers have pointed out that the series of single stimulus presentations tends to form the standard or norm in terms of which the various samples are rated.¹ In this case, the untrained panelist does not compare the current sample with former individual samples; instead, he compares the current sample to a single impression of several former samples combined.

Another difficulty with the single stimulus method of presentation is that the untrained panelists may not have been able to remember accurately the samples they have tested prior to judging the current sample.

Schwartz and Pratt have conducted an experiment to determine what effect the time interval between tasting samples had on the untrained tester's responses.² Two pairs of soups were used in the tests. Each pair was presented in a series of simultaneous tests, and then in a series of single stimulus tests, with the interval between samples being 1, 3, 7, and 10 days. Their findings were that the strongest preferences were obtained under the simultaneous presentation. As the interval between samples was lengthened, there appeared to be a tendency for preferences to diminish.

¹"Consumer Preferences for Frozen Peas in Relation to Standards for Grades," Marketing Research Report No. 280 (U. S. D. A. , AMS, 1958), p. 8.

²M. Schwartz and C. H. Pratt, "Simultaneous vs. Successive Presentation in a Paired Comparison Situation," Food Research, Vol. 21 (1956), pp. 103-108.

The size
food products
enough to rea
A sufficient a
quantity.² Ho
product being

When tast
is the optimum
the samples at
is normally ser
body temperatur
C., and taste i

¹W. F. Do
Food Technolog

²I. McCart
Mrs. America a
Inc., Oakland,

³E. C. Cr
ustrated.

⁴E. Helm a
statorys Comm

Preparation and Serving of Samples

Size of Samples

The size of sample to present panelists is most serious in testing of food products. Dove states that samples for food testing must be large enough to reach all taste organs, but not so large as to cause fatigue.¹ A sufficient amount of sample for two or three bites is usually a normal quantity.² However, the exact size of sample depends upon the particular product being tested.

Temperature of Samples

When taste tests are conducted, one encounters the problem: what is the optimum sample temperature? Many researchers recommend serving the samples at a uniform temperature, and one at which the specific food is normally served. Others believe that most products are best tested at body temperature.³ The gustatory nerves cease to function at 50 degrees C., and taste is also strongly reduced below 15 degrees.⁴

¹W. F. Dove, "Food Acceptability--Its Determination and Evaluation," Food Technology, Vol. 1 (1947), pp. 39-50.

²I. McCarthy, "How Can We Make Sure of Pleasing the Tastes of Mrs. America and Her Family?" (Hallmark Testing Service, Safeway Stores, Inc., Oakland, California) 6 pages.

³E. C. Crocker, Flavor, New York and London (1945), 172 pages illustrated.

⁴E. Helm and B. Trolle, Selection of a Taste Panel, Wallerstein Laboratorys Communication, Vol. 9 (1946), pp. 181-184.

All fo
food is cu
Hallmark t
plain, and

The pu
at the time
tend to bias
mendation m

It is ge
should be ke
the judging s
the panelist.
responses on

One of th
our consumer

1 Dawson

2 McCarti

3 Dawson

Form in Which Samples Are Served

All food samples should be served in the form in which that particular food is customarily eaten, at normal temperature, and at normal strength.¹ Hallmark testing service recommends serving the samples of food first plain, and then in the form in which they are usually consumed.²

Discussions at Judging Sessions

The purpose of the tests should not be disclosed to untrained panelists at the time of testing. If such information is known, the panelists may tend to bias their answers in favor of the purpose of the test. This recommendation may be reversed if the professional panel is used.

It is generally agreed that talking by panelists during the judging should be kept to a minimum, but that discussions are permissible after the judging session has elapsed.³ This plan helps maintain the interest of the panelist. It also gives respondents the opportunity to give independent responses on the products being evaluated.

Sample Coding

One of the precautions a researcher should be aware of when carrying out consumer preference tests is that the samples must be so labeled that

¹Dawson and Harris, op. cit., p. 25.

²McCarthy, op. cit.

³Dawson and Harris, op. cit., p. 29.

their comp

identity of

withheld fr

should be

or packagin

should be

The code s

A test

any of four

tested were

is no differe

A portion

panelists pre

products on th

card which ha

is done for ea

evaluated the

(top, second t

symbol.

The concl

ference among

their composition is not known to those who are doing the testing. The identity of the manufacturer whose product is being evaluated should be withheld from the panel members. All the samples presented to the panel should be placed in identical containers unless the test involves branding or packaging labels as variables which influence preferences. The samples should be identified by symbols rather than by numbers, letters, or name. The code should be such that no implied order is suggested.

A test was conducted to determine if panelists had a preference for any of four symbols used at the M.S.U.-W.S.U. panel. The four symbols tested were (), *, &, and %. The conclusion of this study was that there is no difference among the preferences toward the four symbols.

Position on Cards

A portion of the same experiment was devoted to determining whether panelists prefer any position on the card upon which they are ranking products on the basis of appearance. The usual procedure is to use a card which has a short line with the identifying symbol beside it. This is done for each sample being tested. Hence, in the experiment which evaluated the preferences for the four symbols there were four such lines (top, second top, third top, and bottom) each designated by an appropriate symbol.

The conclusion of this portion of the study was that there is no difference among the preferences toward the four positions on the card.

In the
of each p
the single
the strong
entation,
a contribu

Other
is a tender
by the sing
there is dis
searchers u
should be a
this source
of the panel
of the panel

Thus, the
effect exerts
practical sign
determining te

¹ Schwart

² Nair, Op

³ Ibid., p

Order of Presentation of Samples

In the Schwartz and Pratt study, the preference for the second sample of each pair presented became stronger as the time interval increased in the single stimulus series of presentations. This position effect became the strongest at an interval of seven days. With the simultaneous presentation, they concluded that the order of presentation was eliminated as a contributing variable.¹

Other researchers report different results. Nair believes that there is a tendency to prefer the sample tasted first, whether foods are presented by the single stimulus method or by the simultaneous method.² While there is disagreement as to the exact nature of this position effect, researchers using either the single stimulus or the simultaneous techniques should be aware that it does exist, and should take steps to eliminate this source of bias. Nair has overcome this difficulty by asking one-half of the panel to sample one of a pair first and by requesting the other half of the panel to taste the other sample first.³

Thus, there is much disagreement as to the way in which the position effect exerts itself. In fact, it may be that the position effect has no practical significance. A study conducted by Dedolph and O'Rourke in determining taste preferences for 12 varieties of walnuts indicates that

¹Schwartz and Pratt, op. cit., p. 107.

²Nair, op. cit., p. 133.

³Ibid., pp. 131-136.

the order of
effect upon

In a st
combination
Grieg concl
counted for
in the analy
(which is th
is required t
and when th
ever, the pa
from the orde
jacent to the

Some cri
checked agai
Kroger Food F
stimulus meth
from the stron
flavor of anot

¹ Interview

² See Cha

³ Arnold,

⁴ Platt, C

the order of sample presentation and tasting does not have a significant effect upon consumer preferences for walnuts.¹

In a study to determine preferences for ham containing 16 different combinations of salt and sugar, Marquardt, Pearson, Larzelere, and Grieg concluded that the order of sample presentation could not have accounted for more than one tenth of one percent of the total sums of squares in the analysis of variance test.² The calculated F for the column effect (which is the order of sample presentation effect) was .01. An F of 2.75 is required to show significances when there are three degrees of freedom and when the alpha level has previously been set at the .05 level. However, the panelists may have tasted the samples in an order which differed from the order of sample presentation as all four samples were placed adjacent to their appropriate symbols on one round plate.

Some critics question tests where the strong taste of one sample is checked against a milder flavored item. Arnold's experience with the Kroger Food Foundation seems to disprove this.³ Platt uses a single stimulus method of sample presentation to overcome the difficulty arising from the strong flavor of one sample predominating over the more delicate flavor of another.⁴

¹Interview with Richard Dedolph, October 28, 1963.

²See Chapter VII.

³Arnold, op. cit., p. 45.

⁴Platt, op. cit., p. 50.

As h

three ite

the simul

order des

tasted at

the basis

series is

It is

session, f

may be obt

panel sess

the number

The num

increased if

series of pro

tests.

In a stud

three series of

¹ Clemen

² Anonym.
pp. 314-315.

³ Bengtss

Number of Samples and Replications

As has been reported previously, several researchers believe that three items is the maximum number for a paired comparison design when the simultaneous testing method is used in a taste test. When a rank order design is used, four items is the recommended maximum to be tasted at any one sitting.¹ If the samples are to be evaluated only on the basis of appearance, the maximum number to be tested in any one series is much larger.

It is possible to duplicate taste tests too many times in one panel session, for the senses become dulled rather rapidly. Wrong impressions may be obtained if one attempts to taste too many samples during one panel session.² The stronger the taste and odor of a product, the smaller the number of samples an individual can taste before he must rest.³

The number of samples of one product which can be tasted can be increased if the panel is arranged so that the taster will be ranking other series of products, requiring only visual inspection, between his tasting tests.

In a study conducted by Dedolph and O'Rourke, each panelist ranked three series of walnuts. Each series contained four walnuts, so the

¹Clements, Bayton, and Bell, op. cit., pp. 446-451.

²Anonymous, "Flavor in Foods," Food Industries, Vol. 9 (1937), pp. 314-315.

³Bengtsson and Helm, op. cit., pp. 171-180.

panelis

guided

if fatigu

In this

the 12 s

The

panel se

allowed

"It is

results fro

the compon

new groups

from time to

in two produ

although cor

Of course, w

needed in ord

in consumer

¹ See Ch

² Cowan

panelists each tasted 12 samples. The presentation of samples was guided by a 12 by 12 latin square design making it possible to determine if fatigue contributed a significant amount to the total sum of squares. In this study, it was concluded that the column (order in which each of the 12 samples was tasted) sums of squares was not significant.¹

Therefore, it appears that numerous samples may be tested at one panel session if appropriate designs are used, and if the panelists are allowed to rest their taste buds between the series of taste tests.

Consistency of Panel Preferences

"It is always worthwhile to pay attention to the consistency of the results from successive tests as they are made because they comprise the components of the final results. Repeated tests should be made with new groups until the accumulated results are stable."² Thus, consistency from time to time is important. For example, the repetition of comparisons in two products in different tests should lead to the same conclusions, although consumer groups may be different in the repeated experiments. Of course, when months or years intervene between tests, caution is needed in order to distinguish inconsistency from changes taking place in consumer preferences.

¹See Chapter VII.

²Cowan, op. cit., p. 43.

Some

vide infor

out the te

five times

of walnuts

sub-divide

ments by s

preference

study the p

cause the t

An exam

State-Wayne

by sending o

were chosen

tematic samp

turned via the

Approximately

and returned.

the people wh

were in the m

1
Interview

Some experimental designs can be utilized in a manner that will provide information about the consistency of panelist's preferences throughout the testing period. Dedolph has used a 12 by 12 latin square replicated five times to determine consumers preferences for 12 different varieties of walnuts. As a portion of this analysis, the total sum of squares was sub-divided in eight different categories. One of these categories, treatments by square, is in fact, a consistency test which determines if the preferences toward the 12 varieties differed between the squares. In his study the panelist's preferences were consistent over the squares, because the treatments by square category was not significant.¹

Example of Panel Organization

An example of the way a consumer panel is organized is the Michigan State-Wayne State University consumer preference panel which was selected by sending out questionnaires to 6,500 Detroit households. The names were chosen from listings in the Detroit telephone directory by the systematic sampling method. Five percent of these questionnaires were returned via the Detroit post office due to insufficient or incorrect address. Approximately 20 percent of the 6,500 questionnaires sent were completed and returned. The M. S. U. -W. S. U. consumer panel was selected from the people who returned their questionnaires. The panelists selected were in the middle income group earning four to ten thousand dollars per

¹ Interview with Richard Dedolph, October 28, 1963.

year, and

elists se

people w

and who w

phone to s

acteristics

population

Since

panel has m

Detroit, Mich

group, but s

some instan

main, the su

but do not ha

able on each

Members

meeting. Imm

ers are given

ranked and th

group is brief

in a large roo

pends to

his ranking c

has ranked a

year, and in the middle age group of 30 to 45 years of age. All the panelists selected had received 12 to 13 years of formal education. Those people who belonged to these groups, who had returned questionnaires, and who were willing to take part in the panel, were contacted by telephone to schedule their visit. An effort is made to maintain these characteristics of this panel, and it is assumed to be typical of a consumer population with these characteristics in the Detroit area.

Since its initiation in 1956, a subsample of 125-175 panelists of the panel has met four to five times a year at Wayne State University in Detroit, Michigan. This subsample is mainly selected from the original group, but several new panelists are also selected for each panel. In some instances the panel consists of both a man and his wife, but in the main, the subsample of the panel comprises men or women who are married but do not have their spouses with them. Detailed information is available on each panelist.

Members rank several different types of products at each panel meeting. Immediately prior to each panel session, groups of 10-20 members are given instructions concerning the different series of items to be ranked and the forms on which to rank each series of products. A new group is briefed every half hour. The products are displayed on tables in a large room, and from ten to twenty panelists at a time proceed independently to rank the samples. When an individual member completes his ranking of the products, his forms are checked to make sure that he has ranked all of the products within each series.

they

ordin

for the

written

The members are not told of the purpose of any specific test. Usually they are told only that there are samples of the product to be ranked according to their preference. They are not asked to explain the reasons for their preferences. The respondents can, however, make voluntary written comments.

One
tudes tow
useful for
on the mar
sumers' ne
how ready

When t
values with
sample. Th
various degr
adoption of r
continuum of
scale, and l
tween. The
indicates th

PANEL DESIGNS AND ANALYSIS SECTION

CHAPTER IV

ONE SAMPLE TESTS

One sample tests are sometimes used to evaluate consumers' attitudes toward new or present products. The one sample test is particularly useful for testing new products for which there is no comparable product on the market. Another use of the one sample test is to determine consumers' needs for a product, now on the market. That is, to determine how ready the respondents are to buy the product.

Hedonic Scale

When the hedonic scale is used, it is necessary to have a system of values with which the panel members can express their evaluation of the sample. This system may be composed of numbers or letters which denote various degrees of like or dislike. However, the hedonic scale is an adoption of rating scales, whereby the respondent rates the sample on a continuum of like and dislike. Dislike extremely is at one end of the scale, and like extremely at the other end, with several degrees in between. The following is an example of a nine-point hedonic scale which indicates the range of consumers' attitudes:

_____ Like Extremely

The response
his evaluation

The response
analyzed by
product. The
hedonic scale
assigned to
down the scale
method are for
value in evaluation

Another
responses for
array of the
Both methods
consumers' attitudes

- _____ Like Very Much
- _____ Like Moderately
- _____ Like Slightly
- _____ Neither Like Nor Dislike
- _____ Dislike Slightly
- _____ Dislike Moderately
- _____ Dislike Very Much
- _____ Dislike Extremely

The respondent merely checks the degree which he feels best describes his evaluation of the sample under consideration.

The results from a single product hedonic rating test are usually analyzed by calculating the arithmetic mean of the responses for the product. This is done by assigning numerical values to the points on the hedonic scale. For example, on a 9-point scale, a value of 9 may be assigned to the highest category, 8 to the next best category and so on down the scale. Although the validity of the results obtained by this method are frequently questioned, the results apparently are of some value in evaluating and describing attitudes toward the product.

Another method of analysis consists of computing the percentages of responses falling into various categories on the scale. Alternatively an array of the responses may be made and the median or mid-choice found. Both methods result in a very descriptive and general analysis of consumers' attitudes toward the product.

Crea

Homemak

A five-po

the soup.

Each panel

nearly indic

Eightee

another twe

housewives

their reactio

Disliked the

the soup Ver

Adve

The hec

test to cons

¹ Resea
executed th

Example of the Use of Hedonic Scale

Cream of tomato soup was tasted and ranked by 97 housewives at a Homemakers Panel conducted at Michigan State University in July of 1963.¹ A five-point hedonic scale was used to obtain each panelist's reaction to the soup. The categories listed were:

- _____ Like Very Much
- _____ Like
- _____ Neither Like Nor Dislike
- _____ Dislike
- _____ Dislike Very Much

Each panelist was instructed to place a check by the category which most nearly indicated her attitude toward the soup.

Eighteen panelists indicated that they liked the soup Very Much, and another twenty-four panelists said they Liked the soup. Twenty-five housewives checked the Neither Like Nor Dislike category to describe their reaction to the taste of the soup. Twenty-one panelists said they Disliked the soup. Only eight housewives indicated that they Disliked the soup Very Much.

Advantage of Hedonic Scale in the Evaluation of One Sample

The hedonic scale method of evaluating one sample is a very simple test to construct, explain to panelists, and to analyze. It gives a

¹ Researchers from the MSU Food Science Department planned and executed this study.

descriptiv

tested.

Althou

face value

ifications.

in a positio

another.

Disac

There i

from the hec

people cons

preference.

less (on a ni

such a rating

Evidence sho

less) may res

cessful.¹ H

as such, bec

physiological

to the type o

¹ D. R.
Food Engineer

descriptive evaluation of consumers' attitudes toward the product being tested.

Although the ratings obtained from the hedonic scale are not taken at face value, they do provide a guide as to direction of advantageous modifications. Monadic hedonic scale ratings will often put the researcher in a position to assess several experimental variants relative to one another.

Disadvantage of Hedonic Scale in the Evaluation of One Sample

There is always a danger involved in interpreting the data obtained from the hedonic scale method. Peryam and Girardot indicate that some people consider the values derived from the scale as fixed indicators of preference. In other words, they may consider a mean value of five or less (on a nine-point scale) as an indication that each product having such a rating has a low degree of preference and is therefore unacceptable. Evidence shows that in some cases a fairly low hedonic value (five or less) may result from a series of tests, but the product may still be successful.¹ Hedonic values are not absolute, and should not be interpreted as such, because ratings vary with such factors as the psychological and physiological state of the panelists. Rankings may also vary according to the type of consumer group tested. Thus, it is undesirable to attempt

¹D. R. Peryam and N. S. Girardot, "Advanced Taste Test Methods," Food Engineering, Vol. 24 (1952), p. 61.

to establish a fixed standard of interpretation of the hedonic scale ratings obtained when only one product is being tested.

Readiness to Buy Index

The readiness to buy index is based upon the belief that consumers are not equally predisposed to buy any given product. The range of their predispositions is thought of as a psychological continuum, running from an intention to buy the item immediately to a firm intention of never buying the item. Between these two extremes, the continuum runs through a positive zone ending with intention not accompanied by clear-cut buying plans. It then runs through a neutral area in which disposition might swing either way. Finally, the continuum runs through an area in which attitudes are negative but not firmly set.¹

A consumer's position on this continuum can be measured by having the respondent select the statement which most accurately reflects his intentions toward the product being tested. The panelist then expresses the results on an attitude scale similar to the one shown in Figure I. This particular scale is designed to measure the readiness to buy specific products which are likely to be purchased frequently. Other versions measure readiness to buy larger, less frequently purchased items.

¹William D. Wells, "Measuring Readiness to Buy," Harvard Business Review (July-August 1961), pp. 81-87.

<u>Buying Continuum</u>	<u>Attitude Scale</u>
I. Firm and immediate intent to buy this specific product.	___ I am going to buy some right away. ___ I am going to buy some soon.
II. Positive intention without definite buying plans.	___ I am certain I will buy some sometime. ___ I will buy some sometime.
III. Neutrality: Might buy, might not buy.	___ I may buy some sometime. ___ I might buy some sometime, but I doubt it.
IV. Inclined not to buy the product, but not definite about it.	___ I don't think I'm interested in buying any.. ___ I probably will never buy any.
V. Firm intention not to buy the product.	___ I know I'm not interested in buying any. ___ If somebody gave me some, I would give it away, just to get rid of it.

Figure I. Questionnaire Illustrating the Range of Consumer's Predispositions Toward a Product

Measures of readiness to buy are closely linked to hedonic scales which show how well a product is liked. However, measures of readiness to buy differ from hedonic scales by combining a consumer's regard for the product with an assessment of its purchase probability in his overall buying plans. As a result, the readiness-to-buy measure may more nearly approach purchasing behavior, for it taps intention which is the mental forerunner of action. Readiness-to-buy tests focus on attitude differences toward specific products. Thus, such tests take special note of how far from buying the unready consumer is. Wells believes that readiness-to-buy measures should be viewed both as predictors and as

market data worthy of independent considerations.¹ Although such measures cannot replace sales figures as evidence of the final outcome, they can be used to give an advance sign of the success or failure of a marketing plan. These measures also can be used to "trouble shoot" when sales are not responding as expected. In this way, the elements of the marketing mix can be adjusted or shaped to meet the specific market.

The results obtained from readiness-to-buy questionnaires may be analyzed by calculating the arithmetic mean of the responses for the product. This is done by assigning numerical values to the points on the Attitude Scale. Another method of analysis is to express as percentages the number of responses falling into the various categories on the Attitude Scale. Lastly, an array of responses may be made, and the median may be obtained from this array.

In the one sample case, all three methods of analysis result in a very descriptive and general analysis of consumers' intent to purchase the product. As in the analysis of preferences from the hedonic scale design, the values derived from the Attitude Scale should not be taken to be fixed indicators of preference or intent. For this reason it may be better to test the consumer on a comparison of the product being tested and the most comparable product being marketed successfully.

Example of the Use of the Readiness to Buy Index

Sterile milk was tasted and evaluated by 109 consumers at the MSU-

¹Wells, op. cit.

WSU consumer preference panel.¹ A five-point attitude scale was used to obtain each panelists' reaction toward the milk. Each panelist was given a copy of the questionnaire appearing in Figure II.

MICHIGAN STATE UNIVERSITY

Consumer's Opinion of Milk Quality

This milk was sterilized so there would be no flavor change in packaging it in the sterile condition. If this milk is packaged in regular milk cartons in the sterile condition, it will keep 4 weeks in a refrigerator or at room temperature--If available now:

☐ I would buy some right away.

☐ I would certainly buy it sometime.

☐ I may buy it sometime.

☐ I don't think I am interested in buying any.

☐ I am positive I would never buy any.

Name _____ Comments _____

Figure II. Questionnaire Illustrating the Use of the
Readiness to Buy Index

Each panelist was then instructed to place a check beside the category which most nearly indicated his attitude toward the milk.

Thirty-six panelists indicated they would buy some right away.

Another forty panelists placed a check by the "I would certainly buy it"

¹ Researchers from the MSU Dairy Department presented the samples to the panelists.

sometime" category, and twenty-two panelists said they may buy it sometime. Eight panelists indicated that they didn't think they were interested in buying any. Only three panelists said they were positive that they would never buy any of the milk.

In a study of this type, it may be helpful to visualize the market for sterile milk as a series of concentric circles as shown in Figure III. The center circle and the others closest to it represent the classes of consumers most favorably disposed toward sterile milk; these people may make an effort to purchase this particular product. As one goes further from the center, he finds consumer's less attracted to sterile milk and more interested in other beverages. In the extreme outside circles, one finds consumers who have a violent dislike for the sterile milk, and could not be persuaded to buy it if almost any alternative were available.

This concept, sometimes called the "onion" concept of demand, invites the market researcher to define the characteristics of potential customers who desire the product strongly.¹ He may also determine the characteristics of consumers who are so negatively disposed that it does not pay to try to sell to them.

The researcher may use the chi-square test of independence to determine if education, income or other variables influence a consumer's evaluations toward the product.

¹Alfred R. Oxenfeldt, Pricing for Marketing Executives (Wadsworth Publishing Company, San Francisco, 1961), p. 24.

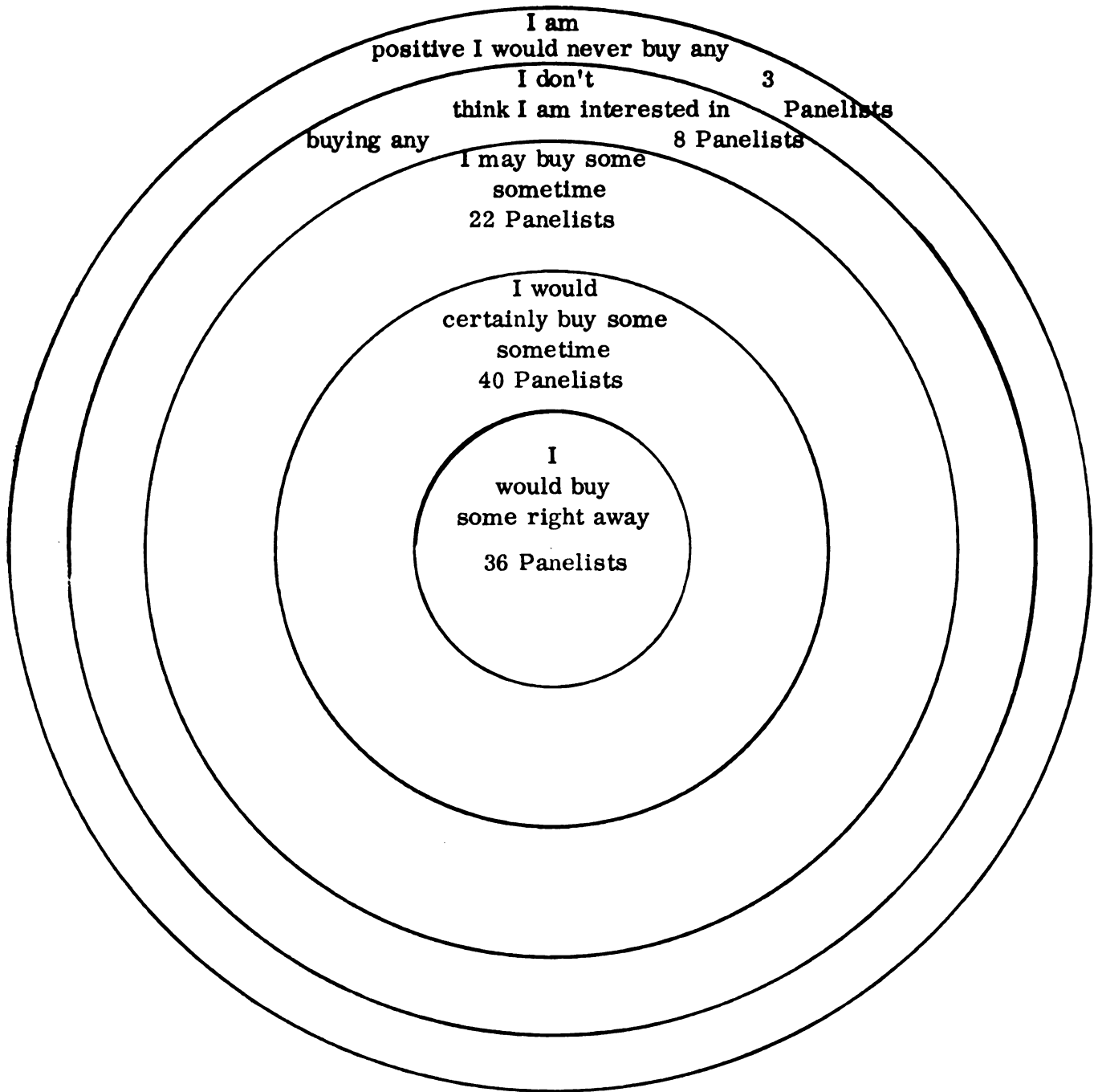


Figure III. Panelist's Evaluation of Sterile Milk

Advantage of Readiness-to-Buy Index in the Evaluation of One Sample

Measures of readiness to buy combine a panelist's regard for the product with an assessment of its purchase probability in the panelist's over-all buying plans. Thus, the readiness to buy index may more nearly approach purchasing behavior as it focuses upon intention to buy, while the hedonic scale method does not include the intention to buy dimension. The readiness-to-buy scale eliminates the problem of separating panelists who buy the product from panelists who have no intention of ever buying the product.

Disadvantage of Readiness-to-Buy Index in the Evaluation of One Sample

The readiness-to-buy index when used on only one sample yields only a descriptive evaluation of the panelist's attitude toward the product being tested. The values obtained from the attitude scale are not absolute, and should not be interpreted to be fixed indicators of consumer buying intent.

CHAPTER V

TWO SAMPLE TESTS

The technique of testing the consumer on a comparison of two products rather than on an inspection of one product is recommended by Arnold.¹ By utilizing this method, a researcher can show whether the product being tested earns a superiority of preference in relation to other products included in the experiment. In this way, the researcher can have more confidence in his findings than if he relies on abstract scaler judgment, as in the case of one product tests described in Chapter IV. Investigators, including the Kroger Foundation, resort to single product tests only when there is no product on the market that is comparable to the product being tested.²

Thus, for most research purposes, the findings on established products must clarify the degree to which other products compete, what possibilities of substitution exist, and how a contemplated change will modify the use of the product. In the case of a new product, the evidence must show its place among similar products and the potential competitive stress of each of them.

¹Arnold, op. cit., p. 46.

²Ibid.

Paired Comparison Tests

Test Procedure

A paired comparison test is a method whereby the panelist judges paired samples by comparing one sample with the remaining sample. The panelist then selects the sample he prefers the most. In some studies, panelists may be asked to place a check by the symbol which identifies the sample he likes best. Another method is to ask the panelist to place a 1 by the symbol which identifies the sample he prefers and to place a 2 by the other sample.

Both the hedonic scale and the readiness-to-buy index methods may also be used to evaluate consumers' preferences toward two products. In either method, the panelist is requested to check the category which most nearly represents his opinion of the product he is evaluating. The panelist is then requested to check the category which most nearly represents his opinion of the other product being tested. In other words he evaluates both products on the same scale which appears once for each product.

One difficulty in interpreting a two product ranking test is how to differentiate between definite choices and mere guesses on the part of individual testers. In the presentation of two samples to a group of panelists, this is not a real difficulty, since the researcher is dealing with groups rather than separate individuals. It is obvious that if all the members of the panel voted by guess, on the basis of pure chance, the votes for the two products would be approximately equal within the limits of the

probable

have no

the resu

error all

observed

then suc

inherent

each othe

be the cle

smaller w

Use of Sig

One n

sample pre

binomial te

products to

indicate th

knows that

the proport

Analys

and minus

which can

panelists c

the panelis

hypothesis

probable error due to the size of the sample. In other words, if the testers have no definite opinion about the merits of the products they are testing, the results of the test will show significant differences only beyond the error allowed by the alpha level. On the other hand, if differences are observed greater than may be accounted for by reasonable probability, then such differences may be supposed to be due to some factor or factors inherent in the products tested. The nearer two test products approach each other in acceptability, and the more alike they are, the fewer will be the clear-cut choices of the individual members of the group, and the smaller will be the observed difference between the two products.

Use of Sign Test on Ranks

One method of analyzing the results obtained from paired comparison sample presentations is the sign test. This test sometimes called the binomial test or McNemar's Test, may be used when there are only two products to be assessed. In the paired comparison presentation, panelists indicate their preference for product A or product B. Therefore, if one knows that if the proportion of panelists preferring product A is P , then the proportion of panelists preferring product B is $1-P$.

Analyses of data with sign test requires the researcher to use plus and minus signs rather than quantitative measures. The null hypothesis, which can be tested by the sign test, states that the probability of the panelists choosing product A over product B is equal to the probability of the panelists preferring product B over product A. In other words, the null hypothesis states that the probability of the panelists choosing either but

not both
that the
not equa
preferenc
tailed tes

Whe
tion of th
In this an
ranks pro
sign to ea
who rank p
analysis f
chance alo
who prefer
prefer prod
true one wo
rankings to
then reject
This is ind
ability ass
and minus

¹ Sidne
(McGraw-H
² Ibid.

not both products is one-half. The alternative hypothesis tested may be that the probability of the panelists choosing A over B, or vice versa is not equal to one half. This is a two-tailed test. If the direction of the preference were predicted in the alternative hypothesis, the test is a one-tailed test.

When one uses the sign test, his attention is focused on the direction of the differences between each panelist's ranking of the products. In this analysis the researcher assigns a plus sign to each panelist who ranks product A over product B. Similarly the researcher assigns a minus sign to each panelist who ranks product B over product A. All panelists who rank products A and B with the same number are dropped from the analysis for the sign test; thus, the N is correspondingly reduced.¹ If chance alone were operative, one would expect the number of panelists who prefer product A over product B to equal the number of panelists who prefer product B over product A. In other words, if the null hypothesis is true one would expect about half of the signs assigned to the panelists rankings to be positive and half to be negative. The null hypothesis is then rejected if the dichotomy is significantly different from a 50-50 split. This is indicated if too few of either sign occur.² The one-tailed probability associated with the occurrence of a particular number of plus signs and minus signs can be determined by reference to the binomial distribution

¹Sidney Siegel, Nonparametric Statistics for the Behavioral Sciences (McGraw-Hill Book Company, Inc., New York, 1956), p. 71.

²Ibid., p. 68.

with $P=1$

over the

The

tions one

tion. In

under the

The

have bee

assigned

$$\binom{N}{x} = \frac{N!}{x! (N-x)!}$$

The

even mor

x is the c

assigned

A table w

ity yielde

null hypo

In s

other is

1. 1b

2. 1b

3. 1b

4. 1b

5. 1b

with $P=1-P=1/2$ where N =the number of panelists who prefer one product over the other.¹

The binomial distribution is the sampling distribution of the proportions one might observe in random samples drawn from a two-class population. In other words, it can give the various values which might occur under the null hypothesis, that $P_A=P_B=1/2$.

The one-tailed probability of obtaining x panelists whose rankings have been assigned a plus and $N-x$ panelists whose rankings have been assigned a minus is given by $p(x)=\binom{N}{x}1/2^x(1-1/2)^{N-x}$. In this formula

$$\binom{N}{x}=\frac{N!}{x!(N-x)!}.$$
²

The one-tailed probability of obtaining the observed value or values even more extreme is given by: $\sum_{i=0}^x \binom{N}{i}1/2^i(1-1/2)^{N-i}$.³ In this formula x is the observed value of the number of panelists whose rankings were assigned the less frequently used sign. It is the number of fewer signs. A table which contains these probabilities is available.⁴ If the probability yielded by the test calculation is equal to or less than alpha, then the null hypothesis should be rejected at the alpha level.

In short, if the number of panelists who prefer one product to the other is 25 or less, the following procedure may be used.⁵

¹Ibid.

²Ibid., p. 37.

³Ibid.

⁴Ibid., p. 250.

⁵Ibid., p. 75.

1.

ranking

2.

panels

both pr

3.

associa

extreme

Tables

with a v

(x).¹ T

product

doubled

4.

is equal

null hypo

"If N

distributi

rection f

x is less

So with

1

2

1. Determine the sign of the difference between each panelist's rankings of the two products.
2. By counting, determine the value of N , which is the number of panelists who prefer product A or product B. Those panelists who rank both products alike are not included in the count to determine N .
3. The binomial distribution may be used to determine the probability associated with the occurrence under the null hypothesis of a value as extreme as the observed value of the number of fewer signs available. Tables contain both one-tailed and two-tailed probabilities associated with a value as small as the observed value of the number of fewer signs (x).¹ The one-tailed test is used when the researcher predicts which product the consumers will prefer. The probability values should be doubled for a two-tailed test.
4. If the probability yielded by the binomial distribution calculation is equal to or less than the previously specified alpha level, then the null hypothesis should be rejected.

"If N is larger than 25, the normal approximation to the binomial distribution can be used."² This approximation is improved when a correction for continuity is used. The latter consists of adding .5 to x when x is less than $1/2 N$, and subtracting .5 from x when x is more than $1/2 N$. So with the correction for continuity, $z = \frac{(x \mp .5) - 1/2 N}{1/2 \sqrt{N}}$. This expression

¹Ibid., p. 75.

²Ibid., p. 73.

may be

variance

reference

Example

An

to deter

green pa

response

no effect

was that

trees. The

significa

coated tr

regular tr

both trees

null hypoth

normally c

hypothesi

region of

so extreme

1 Ibid

2 Alfr

Analysis

may be considered to be normally distributed with zero mean and unit variance.¹ Thus, the significance of an obtained z may be determined by reference to any table of the normal distribution.²

Example of the Use of Sign Test

An example of an appropriate use of the sign test is a study conducted to determine whether consumers preferred Christmas trees coated with a green paint to Christmas trees in their natural state. A sample of 125 respondents was drawn. The null hypothesis was that coating trees has no effect upon preferences for Christmas trees. The alternative hypothesis was that the coating of trees has an effect upon preferences for Christmas trees. The statistical test used was the sign test. The alpha level of significance was set at .05. The number of panelists who preferred the coated tree over the regular tree was 86, and those who preferred the regular tree over the coated tree totaled 28. The 11 panelists who gave both trees the same ranking were not included in the analysis. Under the null hypothesis, z , when computed from $z = \frac{x - 1/2N}{1/2\sqrt{N}}$, is approximately normally distributed when N is larger than 25. Since the alternative hypothesis does not state the direction of the predicted preference, the region of rejection is two-tailed. It consists of all values of z which are so extreme that their associated probability of occurrence under the null

¹ Ibid.

² Alfred J. Dixon and Frank J. Massey, Jr., Introduction to Statistical Analysis (McGraw-Hill Book Company, Inc., New York, 1957), p. 381.

hypothesis is equal to or less than alpha equal to .05. To determine whether the null hypothesis should be accepted or rejected, the observed values are inserted in the formula. Thus, one obtains $z = \frac{(28+.5) - [1/2(114)]}{1/\sqrt{114}} = -5.3$. The .5 is added to x, because x is less than 1/2 N. Reference to a table containing the normal distribution reveals that the probability under the null hypothesis of $z \geq |-5.3|$ is at most $p=2(.00001)$. Because this probability is smaller than .05, the decision is to reject the null hypothesis in favor of the alternative hypothesis. One concludes that the coating of trees has a significant effect upon consumers' preferences for trees.

Advantage of Sign Test

The sign test should be used when ordinal measurement within pairs is possible. In other words, the sign test is useful for analyzing data on a variable which can be assumed to be continuous, but which can be measured in only ordinal terms.

Disadvantages of Sign Test

The sign test simply utilizes information about the direction of the differences in preference between product A and product B. If the relative magnitude as well as the direction of the differences can be measured, then more powerful tests, such as the Wilcoxon Signed Ranks Test, should be used.¹

There are several possible interpretations of a panelists' ranking of one product over another.² The ranking may be pure guess which the

¹ Siegel, op. cit., p. 75.

² Platt, op. cit., p. 9.

respondent gave because he was not able to discern a difference between the samples. Then again the respondent may only have a slight preference, and may buy and be satisfied with either product. Also, the respondent may have a strong preference and buy only the preferred product. Similarly, when the respondent gives both products the same rank, he may not be able to detect a difference, or he may have been able to detect a difference, but have no preference for one product over the other.¹

Use of Wilcoxon Test on Successive Interval Scales

If the hedonic scale or the attitude scale methods of questioning are used, the Wilcoxon Signed Ranks Test may be used. In either case, the researcher can determine which product, A or B, the panelist prefers. He can also rank the difference in preferences in order of absolute size. This is possible by assigning numerical values to the points on either the hedonic or attitude scale. For example, on a 9-point hedonic scale, a value of 9 may be assigned to the Like Extremely category, 8 to the Like Very Much category, 7 to the Like Moderately, ... and finally 1 to the Dislike Extremely category.

In the Wilcoxon Signed Ranks Test, the difference between the i^{th} panelists' evaluation of products A and B is indicated by d_i . This test consists of ranking all the d_i 's without regard to sign.² That is, a rank of 1 is given to the smallest d_i in absolute value, a rank of 2 is given to

¹Nair, op. cit., p. 33.

²Siegel, op. cit., p. 76.

the next smallest, etc. The next step involved is to affix the sign of the difference to each rank given above; in other words, indicate which ranks arise from negative d_i 's (where the panelist gives product B a higher evaluation than he does product A) and which ranks arise from positive d_i 's (where product A is preferred over product B).

If the preferences for two products A and B are equal, one would expect to find some of the larger d_i 's favoring product A (indicated by a positive ranking) and some favoring product B (indicated by a negative ranking). Thus, if one summed the ranks having a plus sign and also summed those having a minus sign, he would expect the two sums to be about equal under the null hypothesis which states that there is no difference in the preference for product A or B. However, if the sum of the positive ranks is very much different from the sum of the negative ranks, one would infer that preferences for product A differ from preferences for product B. In this case, the null hypothesis would be rejected. Hence, the null hypothesis is rejected at the particular level of significance set by the experimenter, if either the sum of the ranks for the negative d_i 's or the sum of the ranks for the positive d_i 's is too small.¹ The smaller, in absolute value, of the sums of the like signed ranks is designated as T.

If a panelist assigns the same evaluations to both products, $d=0$, then his response should be dropped from the analysis. In the Wilcoxon test, N equals the number of panelists minus the number of panelists who assign the same evaluation to both products.²

¹ Ibid.

² Ibid.

If two or more d 's are of the same size, they are each assigned the average of the ranks which would have been assigned if the d_i 's had differed slightly.¹ "The practice of giving tied observations the average of the ranks they otherwise would have gotten has a negligible effect on T , the statistic on which the Wilcoxon Test is based."²

In short, the Wilcoxon Test consists of:³

1. Determining the signed difference (d_i) between each panelist's evaluation of the two products being tested.
2. Ranking these d_i 's without respect to sign and assigning the average of the tied ranks to tied d_i 's.
3. Affixing to each rank the plus or minus sign of the d which it represents.
4. Determining T which is the smaller in absolute value, of the sums of the like signed ranks.
5. Determining N which is the total number of d 's having a sign. N does not include any d 's which equal zero.
6. Determining the significance of the observed value of T by referring to tables that have been developed.⁴

"If the observed value of T is equal to or less than that given in the table

¹Ibid.

²Ibid. , p. 77.

³Ibid. , p. 83.

⁴Ibid. , p. 254.

for a particular significance level and a particular N, the null hypothesis may be rejected at that level of significance."¹

If N is larger than 25, the value of z as calculated in $z = \frac{T - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}}$

is approximately normally distributed with zero mean and unit variance.²

The region of rejection consists of all calculated z's which are so extreme that the probability of their occurrence under the null hypothesis is equal to or less than the previously determined alpha level. If the direction of the preference is not stated in the alternative hypothesis, the region of rejection is two-tailed. If the direction of the preference is stated in the alternative hypothesis, the region of rejection is one-tailed.

Example of the Use of the Wilcoxon Test

The Wilcoxon Test was used to analyze consumer preferences for fresh or frozen pork loin roasts.³ Each of the 36 panelists ranked the roasts on a nine-point hedonic scale.

The null hypothesis was that there was no difference in consumer preferences for fresh or frozen pork loins. The alternative hypothesis was that there was a preference for either, but not both, products. The alpha level was set at .05. N was the number of panelists (36) minus any panelist whose d is zero (7) or 29. Under the null hypothesis, the values of

¹Ibid., p. 83.

²Ibid., p. 79.

³Glen Willis Higgins, "Consumer Attitudes and Behavior Toward Frozen Meats" (unpublished Master's dissertation, Michigan State University, 1958), p. 72.

84

z as computed from $z = \frac{T - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}}$ are approximately normally distrib-

uted with zero mean and unit variance. Therefore, a table containing critical values of the normal distribution gives the probability associated with the occurrence under the null hypothesis of values as extreme as an obtained z. The region of rejection consists of all z's which are so extreme (either + or -) that the probability associated with their occurrence under the null hypothesis is equal to or less than $\alpha = .05$.

The calculation of T is shown in Table 1. The calculated value of z in this case is -1.52. The probability associated with the occurrence under the null hypothesis of a z of -1.52 is .128. The null hypothesis was not rejected because the calculated probability is greater than alpha equalling .05. Thus, there does not appear to be a difference in preferences for fresh or frozen pork loins.

Advantages of Wilcoxon Test on Successive Interval Scales

The Wilcoxon Signed Ranks Test gives more weight to the panelist who expresses a large difference between his evaluation of the two products than to a panelist who expresses a small difference between his evaluation of the two products. So while the sign test gives equal weight to all panelists, the Wilcoxon test considers the relative magnitude, as well as the direction of the differences between a panelist's evaluation of the two products. Thus, when measurement is on an ordinal scale, both within and between the product evaluations (as may be the case of the successive interval scales), the Wilcoxon is a more powerful test than the sign test.

TABLE I. Difference in Palatability Ratings Obtained From a Nine-Point Scale of Pork Loin Roasts, Fresh as Compared with Fresh-Frozen, East Lansing Panel of Housewives, April 1957. Session I.

Panelist	Rank of Fresh	Rank of Fresh Frozen	d	Rank of d	Rank With Smallest T
1	7	5	2	17.5	
2	4	5	-1	- 7.0	7.0
3	7	4	3	23.0	
4	7	7	0		
5	9	8	1	7.0	
6	8	9	-1	- 7.0	7.0
7	7	8	-1	- 7.0	7.0
8	8	7	1	7.0	
9	8	8	0		
10	8	6	2	17.5	
11	8	8	0		
12	4	7	-3	-23.0	23.0
13	6	4	2	17.5	
14	7	9	-2	-17.5	17.5
15	6	7	-1	- 7.0	7.0
16	6	6	0		
17	7	8	-1	- 7.0	7.0
18	8	8	0		
19	4	6	-2	-17.5	17.5
20	7	8	-1	- 7.0	7.0
21	7	2	5	25.5	
22	8	6	2	17.5	
23	4	7	-3	-23.0	23.0
24	7	2	5	26.5	
25	8	8	0		
26	7	6	1	7.0	
27	8	8	0		
28	8	7	1	7.0	
29	9	7	2	17.5	
30	6	8	-2	-17.5	17.5
31	6	7	-1	- 7.0	7.0
32	9	8	1	7.0	
33	8	4	4	25.0	
34	8	7	1	7.0	
35	7	7	0		
36	8	8	0		

$$T = \frac{N(N+1)}{4} = \frac{147.5 - 29(30)}{4} = -1.52$$

$$\sqrt{\frac{N(N+1)(2N+1)}{24}} = \sqrt{\frac{29(30)(58+1)}{24}}$$

If efficiency is defined as the inverse ratio of sample sizes needed to attain the same power, then the local asymptotic efficiency of the Sign Test, against normal alternatives, relative to the Wilcoxon Test is .67. This means that .67 is the limiting ratio of sample sizes necessary for the sign test and the Wilcoxon Test to attain the same power. Thus, you can do approximately as well with a sample of size 100 using the Wilcoxon Test as you can with a sample of size 150 using the sign test if the samples are taken from normal distributions.

Disadvantage of Wilcoxon Test

The Wilcoxon Test may only be used when measurement is on an ordinal scale both within and between evaluations for the two products being tested.¹ This means that the consumer must not only be able to rank one product over the other, but he must also be able to rank the degree of intensity with which he prefers one product over the other.

The most fundamental limitation in using the Wilcoxon Test is that the rating scales may not provide an accurate means of measuring attitudes toward product attributes. This is the case if there are attitudes which are not scalable at all. For example, two panelists may have completely different attitudes about the healthfulness of a certain food. The difference may revolve about the circumstances under which the food is healthful, rather than whether it is more or less healthful. In other words, product A may be considered to be healthful under one set of circumstances, and

¹Ibid., p. 93.

product B under another. Any kind of general scaling of one food as more or less healthful than the other may not make sense.

Triangle Test

Test Procedure

The triangle test is most useful in comparing two samples which are almost alike. Three samples are examined in the triangle test, but two of these samples are duplicates.¹ The panelist is first asked to identify the two samples which are alike. Then he is asked to state his preference either for the alike samples or for the unlike sample.

The important consideration one should remember when using the triangle test is that the judges should be able to identify the products only on differences in the particular characteristic under investigation, and not on any other characteristics. For example, if one is investigating whether products A and B differ in flavor, and A and B happen to differ in color, then he must mask out the color differences from the judges.

Although the triangle test is open to the usual criticisms of any subjective test, it does lend itself to statistical analysis. It can be used first to determine whether consumers can discern a difference between the two products. Secondly, if the panelists can discern differences in the products, the results obtained from the triangle test can be used to determine

¹E. B. Roessler, J. Warren, and J. F. Guymon, "Significance in Triangular Taste Tests," Vol. 13, Food Research, pp. 503-505.

a consumer preference for either product. This test may also be utilized to select personnel for trained taste panels, where individual sensitivities to different taste factors must be evaluated.

In the triangle test, the probability "p" of a panelist correctly "guessing" the unlike sample is one-third if there is no difference in the product being tested. The probability (1-p) of an incorrect guess is two-thirds. By chance alone, the expected numbers of correct answers for n panelists is $n/3$ and the expected number of incorrect answers is $2n/3$. The standard error of the distribution is $\sqrt{np(1-p)}$ which is equal to $\sqrt{2n}/3$. If y denotes the observed number of correct answers, then the normal deviate, applying Yate's Correction for Continuity, is: $z = \frac{y - n/3 - 0.5}{\frac{\sqrt{2n}}{3}}$ when y is greater than $\frac{n}{3}$.¹

The observed number of correct answers must exceed $n/3 + .05 + 0.9239\sqrt{n}$ for the one-tailed test at the .025 percent level of significance. In this case, one is only interested in the probability of the frequency of correct choices exceeding a given value, and this is properly a one-tailed test. A significantly small number of correct choices would have no positive bearing upon the panelist's ability to detect the difference between the products being tested. All frequencies of choices less than the significant number of correct choices would constitute negative evidence concerning the panelist's ability to discern the differences in the products.

The number of correct answers calculated in a similar manner, needed

¹ Ibid.

to establish significant differentiation are tabulated for various numbers of panelists in Table II. If the observed y is larger than the calculated (or table) y then discernment is significant at the previously established alpha level.

If the panel shows an ability to discern differences in the two products being tested, and if the difference is significant at the previously specified alpha level; then an effort should be made to ascertain the preferences toward both products.

This may be done by asking the panelists whether they prefer the like samples or the unlike sample. If one assumes that some panelists will prefer the like samples and others will prefer the unlike sample in about equal numbers, then a large number of selections of the one product over the other may indicate, on the basis of the panel's judgment, a significant preference for either product.

The sign test may be used to determine if the panelist's preference for one product, say the like samples, differs significantly from their preference for the other product, say the unlike sample. Only the responses of panelists who could correctly identify the unlike sample should be analyzed by the sign test.

There is a danger that panelists may tend to prefer the like samples or vice versa. So both products should be presented as the like samples one-half of the time, and as the unlike sample the other half of the time.

7

TABLE II. Probability in Triangular Preference Tests

No. of panel-ists	No. of correct answers necessary to establish significant differentiation		No. of panel-ists	No. of correct answers necessary to establish significant differentiation	
	P=0.025	P=0.005		P=0.025	P=0.005
7	5	6	57	27	29
8	6	7	58	27	29
9	6	7	59	27	30
10	7	8	60	28	30
11	7	8	61	28	30
12	8	9	62	28	31
13	8	9	63	29	31
14	9	10	64	29	32
15	9	10	65	30	32
16	10	11	66	30	32
17	10	11	67	30	33
18	10	12	68	31	33
19	11	12	69	31	34
20	11	13	70	32	34
21	12	13	71	32	34
22	12	14	72	32	35
23	13	14	73	33	35
24	13	14	74	33	36
25	13	15	75	34	36
26	14	15	76	34	36
27	14	16	77	34	37
28	15	16	78	35	37
29	15	17	79	35	38
30	16	17	80	35	38
31	16	18	81	36	38
32	16	18	82	36	39
33	17	19	83	37	39
34	17	19	84	37	40
35	18	19	85	37	40
36	18	20	86	38	40
37	18	20	87	38	41
38	19	21	88	39	41
39	19	21	89	39	42
40	20	22	90	39	42
41	20	22	91	40	42
42	21	22	92	40	43
43	21	23	93	40	43
44	21	23	94	41	44
45	22	24	95	41	44
46	22	24	96	42	44
47	23	25	97	42	45
48	23	25	98	42	45
49	23	25	99	43	46
50	24	26	100	43	46
51	24	26	200	80	84
52	25	27	300	117	122
53	25	27	400	152	158
54	25	27	500	188	194
55	26	28	1,000	363	372
56	26	28	2,000	709	722

Example of Triangle Test

The triangle method of sample presentation and analysis was used to determine consumers preferences between ham cured with salt alone and ham cured with salt plus 2 percent sugar. The M.S.U.-W.S.U. consumer panel was used in this experiment. The null hypothesis tested was that consumers cannot distinguish the difference in the two types of ham. The alternative hypothesis tested was that consumers can distinguish the difference in the two types of ham. The alpha level was set as .05.

The triangle method of distinguishing differences seemed to be very satisfactory. The groups came through at various intervals of time, so approximately 20 panelists were in each session. The day was divided into two parts of panel observations, with an afternoon and evening panel. The ham cured with salt alone was presented as the like samples during the afternoon session. The ham cured with salt plus 2 percent sugar was presented as the like samples in the evening session. Each session will be discussed separately and then lumped together for analysis.

In the afternoon session, 58 panelists tasted the ham. Out of this number, 33 could properly discern the unlike samples, which showed that the panels ability to discern differences in the two types of ham was significant at the 5 percent level. In the evening session, 67 panelists participated on the ham panel, and 36 of them could correctly discern the unlike samples. Again, the ability of the panel to discern the difference in the two types of ham was statistically significant at the 5 percent level.

On lumping the evening and afternoon sessions together, 69 out of 125 panel members could discern the difference between the like and unlike samples. Thus, the panel showed the ability to discern differences in the two types of ham, and the difference was significant at the 5 percent level.

The next step was to determine if there was a preference for the hams treated with sugar or those with salt only. The null hypothesis tested was that preferences for both types of ham are equal. The alternative hypothesis was that one type of ham is preferred over the other type of ham. The alpha level was set as .05. In the afternoon session, 24 out of the 33 panel members able to distinguish differences preferred the hams containing sugar. This preference was statistically significant at the 5 percent level.

In the evening, 23 of the 36 panel members showing correct discernment preferred the sugar-cured hams. This difference in preference for the sugar-cured ham was not significant at the 5 percent level. On combining the two panels sessions (Table III) the preference for the sugar-cured hams was significant at the 5 percent level.

TABLE III. Results of Triangular Panel Tests of Ham

Total number of panel members	125
Number of panel members showing discernment	69 ^a
Number of panelists preferring sugar-cured ham	47 ^b
Number of panelists preferring "salt only" ham	22

^aSignificant discernment at the 5 percent level

^bSignificantly different at 5 percent level



Advantages of the Triangle Test

The chief advantage of the triangle test is that the probability of a taster correctly guessing the unlike sample, then preferring either the unlike sample or the like samples is $(1/3)(1/2) = 1/6$. In a rank order test involving two samples, the probability that a panelist will prefer either sample is $1/2$. Thus, the triangle test is a more powerful test as it reduces the probability of agreeing judgments from $1/2$ to $1/6$. That is, it reduces the probability of choosing the "best" sample by chance alone to 16.6 percent, as compared to a 50 percent chance for guessing the best sample in the paired comparison test.

The triangle test alleviates some of the problems which are encountered with a ranking procedure on two products. It has been mentioned that a respondent may give both products the same rank when he cannot detect a difference between the products, or when he can detect a difference between the products but does not have a preference for any one product over the other. It is not possible to determine which situation is occurring when one uses the ranking tests. However, one is easily able to distinguish these situations with the triangle tests.

Bradley indicates that the asymptotic relative efficiency of the triangle test to the duo-trio test is 112 percent.¹ This means that 1.12 is the limiting ratio of sample sizes necessary for the triangle test and the duo-trio test to attain the same power. The duo-trio test consists of the

¹Ralph Bradley, "Some Relationships Among Sensory Difference Tests," Biometrics, Vol. 19, pp. 385-397.

panelist examining a pair of products. The panelist is then given a third sample, told that the sample is identical to one of the two he has just tested, and asked to indicate which one it matches. This test will not be discussed, for the triangle test is superior to it and both tests cost an identical amount to conduct.

Disadvantages of the Triangle Test

One of the disadvantages of the triangle method is the large number of persons which is sometimes required to administer the tests to a large number of panelists. Another disadvantage is that panelists usually require a longer period of time to complete a triangle test than they do to complete a paired comparison test. Also, the task of explaining the triangle test is more difficult than it is in the paired comparison test. The net result is usually that some participants are initially confused as to the procedure they are to follow in triangle tests.

Apparently, the "position effect" is not eliminated by the triangle method of sample presentation, for research conducted in California indicates a bias for the second sample tasted.¹

In general, the primary difference between the two-sample and triangular designs lies in the just noticeable difference. The assumption that in both the two-sample and the triangular tests the point of just noticeable difference is identical is questioned by Filipello.² While the

¹F. Filipello, "A Critical Comparison of the Two-Sample and Triangular Binominal Designs," Food Research, Vol. 21 (March-April 1956), pp. 235-241.

²Ibid.

experiment conducted by Filipello demonstrated the superior power for discrimination of the triangular design, it also indicated that the greater just noticeable difference necessary for the triangular test overcomes any statistical advantage in the determination of absolute or difference thresholds.

He concluded that when quality may be stated unambiguously, the two-sample design is more sensitive than the triangular design. The difference in power of discrimination is explained by the fact that the just noticeable difference for the triangular is higher than for the two-sample design.¹

However, another researcher's results indicate that paired tests and triangular tests are normally about equally powerful.²

Byer and Abrams carried on parallel taste tests, using differing concentrations of dextrose in aqueous solution, to compare odd sample selection by the triangular test with judgment of taste quality intensity by the paired comparison procedure.³ In this study, the paired comparison test results reflected discrimination of higher statistical significance than did those of the triangular tests.

¹Ibid., p. 240.

²N. T. Gridgeman, "Taste Comparisons: Two Samples or Three?" Food Technology, Vol. 9 (March 1955), p. 148.

³Albert J. Byer and Dorothy Abrams, "A Comparison of the Triangular and Two-Sample Taste-Test Methods," Food Technology, Vol. 7, pp. 185-187.

CHAPTER VI

THREE AND FOUR SAMPLE TESTS

Series of Paired Comparison Tests

Test Procedure

When the method of paired comparisons is used, each panel member is given two samples at a time. He is asked which one of the two samples he prefers. This is done until each sample is compared against all the other samples.¹

Several methods can be used to analyze the rankings given to the products being tested. The sign test may be used successfully if the panelist either ranks the samples or checks the preferred product. The Wilcoxon Test should be used if either a hedonic or attitude scale is utilized in the presentation of the paired samples. In either instance, it is important that all panelists complete the rankings for every sample in the series. If several panelists turn in incomplete data, it is difficult to analyze the findings. A common practice in large panels, if incomplete cases are few, is to disregard such responses altogether and to analyze only the responses of panelists who complete the rankings in the entire series.

¹The number of unordered subsets or combinations of r elements is $\frac{n!}{r!(n-r)!}$, where n equals the number of samples and r represents the number of samples judged during one time period. Thus, if there were three samples, $\frac{3!}{2!(3-2)!} = 3$ paired comparisons would be required to test each sample against the other samples. A test involving four samples would necessitate the use of six paired comparisons.

Other methods of analyzing preferences obtained from a series of paired comparison tests are the Bradley-Terry procedure and the Scheffe analysis of variance for paired comparisons. In the Bradley-Terry procedure, the preferences are analyzed as ranks, and the "no preference" response is given a tied rank.¹ If degrees of preference are recorded, one may use the Scheffe method, which gives a little more information than that given by the Bradley-Terry method.²

Advantage of Series of Paired Comparison Tests

The paired comparison method offers the advantage that only two samples are being compared at a time. The panelist does not have to keep in mind a whole series. If a series of signs or Wilcoxon tests are used, he does not have to keep a hedonic or attitudinal scale constant as he applies it in the different series of tests.

A series of paired comparison tests also offers the advantage that intra-rater reliability can be evaluated. In other words, one can determine the reliability or consistency of a panelist's ranking in different series.

For example, if all possible paired comparisons involving four products were made twice by each panelist, then one could cast each panelist's rankings in a 4 by 4 matrix. If the panelist is perfectly consistent throughout

¹Ralph Bradley and Milton Terry, "Rank Analysis of Incomplete Block Designs," Biometrika, Vol. 39 (1952), p. 324-345.

²See Appendix E for a discussion of the Scheffe method.

the 12 paired comparison tests, then the rankings in the first row should be the same as the rankings in the first column. The rankings in the second row should be the same as the rankings in the second column, etc. This is illustrated in Figure IV.

		PRODUCT			
		A	B	C	D
PRODUCT	A		A	A	A
	B	A		B	B
	C	A	B		C
	D	A	B	C	

FIGURE IV. A Perfectly Consistent Panelist's Rankings of Four Products in 12 Paired Comparison Tests¹

The consistency of an individual observer over the series of paired comparisons may also be measured by Kendall's coefficient of consistency. The coefficient of consistency is calculated by obtaining the ratio of the number of circular traids of an observer to the total possible number and subtracting the ratio from unity.² A circular traid is the relationship existing when product A is preferred over product B, product B is preferred over product C, but product C is preferred over product A. High values of the coefficient of consistency indicate consistent observers. Low values denote observer inconsistency.

¹Letters in matrix indicate which product the panelist prefers.

²J. Edward Jackson and Mary Fleckenstein, "An Evaluation of Some Statistical Techniques Used in the Analysis of Paired Comparison Data," Biometrics, Vol. 13 (1957), p. 57.

Disadvantages of Series of Paired Comparison Tests

The principal disadvantage of a series of paired comparisons tests is that the panelist must rank more samples than he would have to rank if all products were ranked simultaneously in one test. For example, if there were four different products to be tested, the panelist would have to examine 12 samples assuming a series of paired comparison tests was used. He would only examine four samples if all were presented simultaneously.

The procedure of using a series of paired comparison tests is not only tedious, but it may lead to fallacious conclusions if it is used to test the null hypothesis that there is no difference in preferences toward all of the products being tested.

Chi Square Test

Test Procedure

Circumstances sometimes require that one must design an experiment so that more than two products or samples can be studied simultaneously. When three or four products are to be compared in an experiment, it is necessary to use a statistical test which will indicate whether there are significant differences among the three or four products. Once it has been determined that there are significant preference differences for the products, one can test the significance of the difference between pairs of products. It is only when an "over-all" test allows one to reject the null hypothesis of no difference in preferences among the products being tested,

that he is justified in employing a procedure for testing significant differences between any two of the three or four products.

The chi square test of independence enables one to examine data which are only classificatory (in a nominal scale) or in ranks (in an ordinal scale).¹ This test may be used to test the hypothesis that two characteristics are independent.² The term independent means that the distribution of one characteristic is the same regardless of the other characteristic. The chi square test may be used to determine if the proportions or frequencies of product rankings are independent of the products being ranked. In other words, the chi square procedure may be used to test the null hypothesis that the various products have the same proportions of individuals in the various ranking categories. This is not the usual problem of a contingency table with fixed border totals for repeated sampling is not a random rearrangement of the c (number of products being tested) n (number of panelists) items. Instead, each of the n panelists acts independently, but has only $c!$ possible preference sequences. In other words, each panelist must examine all c products and use all c ranks. However, Anderson has proven that an adjustment factor of $\frac{c-1}{c}$ times the X^2 calculation obtained by the contingency formula will yield an over-all statistic which has an asymptotic X^2 distribution with the usual $(c-1)(r-1)$ degrees of freedom.³ The r is the number of rows in the table, and is also

¹ Sidney Siegel, op. cit., p. 175.

² Dixon and Massey, op. cit., p. 224.

³ R. L. Anderson, "Use of Contingency Tables in the Analysis of Consumer Preference Studies," Biometrics, Vol. 15 (December 1959), pp. 582-590.

the number of categories describing a panelist's evaluation of the sample. Thus, to apply the chi square test, one first arranges the frequencies in a $c \times r$ table.¹

Under the null hypothesis, the sampling distribution of $\frac{c-1}{c}X^2$ as computed from the contingency table formula is approximated by a chi square distribution with $(c-1)(r-1)$ degrees of freedom. In this case, c is the number of columns, and r is the number of rows. Therefore, the probability associated with the occurrence of values as large as an observed $\frac{c-1}{c}X^2$ is given in any table containing the critical values of chi square.⁴ If an observed value of $\frac{c-1}{c}X^2$ is equal to or greater than the critical value obtained from a chi square table for a particular level of significance and for $(c-1)(r-1)$ degrees of freedom, then the null hypothesis is rejected at that level of significance.⁵ Once "over-all" significance has been established, one can test the significance of the difference between any two products.

¹The null hypothesis is that consumer preferences do not differ toward the c products being tested. In other words, the c products are hypothesized to have come from the same or identical populations with respect to consumer preferences. This hypothesis may be tested by applying the formula: $\frac{c-1}{c}X^2 = \frac{c-1}{c} \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$ where O_{ij} is the observed number

of panelists categorized in the i^{th} row of the j^{th} column, and where E_{ij} is the number of panelists expected under the null hypothesis, to be categorized in the i^{th} row of the j^{th} column.² The E_{ij} may be determined by multiplying the marginal totals common to the cell and dividing this product by the sum of the row totals which is N . It should also be equal to the sum of the column totals. The $\sum_{i=1}^r \sum_{j=1}^c$ directs one to sum over all cells.³

²Anderson, loc. cit. ³Ibid. ⁴Siegel, op. cit., p. 249.

⁵Ibid., p. 175.

It is apparent that the degrees of freedom for the chi square test depend on the number of cells, and not on the number of observations. However, this does not mean that the number of observations can be too small. The chi square test requires that the expected frequencies in each cell be five or larger.¹

The chi square test may be used in studies where the panelists rank the products. It can also be used in studies which make use of either the hedonic or attitude scale. If either scale is used, the table is constructed in exactly the same way, for each category on the scale makes up one row in the *cxr* table. In other words, the categories are substituted for the 1, 2, 3, 4 rankings in the table.

Example of Chi Square Test

A study was made to determine consumer preferences toward pork sausage containing four different levels of sugar (0%, 1.0%, 1.5%, 2.0%). The M.S.U.-W.S.U. consumer preference panel was used in this test. Altogether, 100 panelists ranked the four different sausage samples by placing a 1 beside the sample they liked best, a 2 beside the sample they liked second best, a 3 beside the sample they liked third best, and a 4 beside the sample they liked least of all.

The null hypothesis was that the consumers' rankings of the four sausage products were independent of the four sausage products, e.g.,

¹Lester V. Manderscheid, "An Introduction to Statistical Hypothesis Testing," Department of Agricultural Economics, Mimeograph 867, May 1962, p. 7.

rows are independent of columns. The alternative hypothesis was that the consumers' rankings of the sausage products were dependent on the sausage products.

The statistical test used was the chi square test. The alpha level was set as .10. The number of panelists was 100, but each panelist gave four rankings; hence, N equals 400.

Under the null hypothesis, $\frac{c-1}{c}X^2$, as computed from the formula, is distributed approximately as chi square with $(c-1)(r-1)$ degrees of freedom. Under the null hypothesis, the probability associated with the occurrence of values as large as an observed value of $\frac{c-1}{c}X^2$, is shown in a X^2 table.

The region of rejection consists of all values of $\frac{c-1}{c}X^2$, which are so large that the probability associated with their occurrence under the null hypothesis is equal to, or less than alpha equal to .10.

Table IV gives the number of panelists who gave a 1, 2, 3, or 4 ranking to four different kinds of sausage.

TABLE IV. The Observed and Expected Frequency of Panelists Ranking Four Kinds of Sausage with a 1, 2, 3, or 4

PERCENT OF SUGAR CONTAINED IN SAUSAGE										Row
Ranking		0%	1.0%		1.5%		2.0%		Total	
1.	18	<u>25.75*</u>	25	<u>25.75</u>	33	<u>25.75</u>	27	<u>25.75</u>	103	
2.	19	<u>25.0</u>	20	<u>25.0</u>	31	<u>25.0</u>	30	<u>25.0</u>	100	
3.	29	<u>25.25</u>	27	<u>25.25</u>	20	<u>25.25</u>	25	<u>25.25</u>	101	
4.	34	<u>24.0</u>	28	<u>24.0</u>	16	<u>24.0</u>	18	<u>24.0</u>	96	
Column										
Total	100		100		100				400	

*The expected frequencies are underlined; the observed frequencies are not underlined.

The expected frequencies which appear in the table were obtained by multiplying the marginal totals common to the cell and dividing the product by N. For example, the expected frequency of panelists giving a 1 ranking to sausage containing 0% sugar was found by multiplying (103) (100) and dividing this product by 400, which gives 25.75.

A computation of X^2 was made from the values of Table IV by making use of the formula: $\frac{c-1}{c}X^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(18-25.75)^2}{25.75} + \dots + \frac{(18-24)^2}{24} = 3/4(20.7) = 15.4$

The size of the $\frac{c-1}{c}X^2$ reflects the magnitude of the discrepancy between the observed and the expected values in each of the cells. The table value of a chi square at the alpha .10 level and with (c-1) (r-1) or (4-1) (4-1) equals 9 degrees of freedom, is 14.68. Because the calculated $\frac{c-1}{c}X^2$ is larger than the $X^2_{.90}$ with 9 degrees freedom, the null hypothesis should be rejected. Consumers' rankings of the sausage products appear to be dependent upon the percentage of sugar contained in the sausage. Individual panel tests should be used to determine differences between all possible pairs of comparisons.

Advantage of Chi Square Test

The chi square test has the advantage of enabling a researcher to analyze data which are only classifactory in nature or in ranks. The chi square test should be used if the researcher is interested in determining whether his samples are from populations which differ in any respect at

all, e.g., in location or dispersion or skewness.¹

Disadvantages of Chi Square Test

When one rejects the null hypothesis on the basis of a chi square test, he can assert that the product groups are from different populations, but he cannot say in what specific way or ways the populations differ.² Chi square tests are insensitive to the effects of order when the degrees of freedom are greater than one.³ Therefore, when a hypothesis takes order into account, X^2 may not be the best test.

A further disadvantage of the chi square test is that the expected frequencies in each cell must not be smaller than five.⁴ This requirement is usually met in ranking tests in a large panel. If the hedonic or attitude scales are used, however, there may be some studies where the expected frequency is less than five. In this case, the researcher must combine adjacent categories on the scale so as to increase the expected frequencies in various cells. The adjacent categories, which are combined should have some common property. This common property is necessary to interpret test results after the categories have been combined.⁵ This is not always possible.

¹ Siegel, op. cit., p. 157.

² Ibid., p. 158.

³ Ibid., p. 179.

⁴ Manderscheid, loc. cit.

⁵ Siegel, loc. cit.

Coefficient of Concordance (Friedman Two Way Analysis of Variance)

Test Procedure

The coefficient of concordance, W , is an index of the divergence of the actual agreement shown in the rankings from the maximum possible (perfect) agreement.¹ In other words, W (which is the symbol used to indicate the coefficient of concordance) measures the communality of judgments for all of the panelists. If all the panelists agree in their rankings of the products, W equals 1.00. If they differ very much among themselves, W is small. If there is no agreement in the panelists' rankings of the products, W will equal zero. Thus, W may assume values only between zero and plus one.

Let n equal the number of products to be evaluated and m be the number of panelists ranking the n products. He may then cast the observed ranks in a table having m rows and n columns.

The researcher should then find the sum of the ranks assigned to each of the n products by the m panelists. This sum of ranks assigned to each product is R_j . If the panelists are in perfect agreement about the products and rank the products in the same order, then one product would receive m ranks of 1, another product would receive m ranks of 2, still another product would receive m ranks of 3, and the remaining product (if the test involved four products) would receive m rankings of 4. Therefore, if there is perfect agreement among the m panelists, the R_j for the

¹Ibid. , p. 230.

products would be respectively: m , $2m$, $3m$, and $4m$.¹

The next step consists of finding the mean of R_j . This is done by summing the R_j and dividing this sum by n . One then expresses each R_j as a deviation from the mean of the R_j . These deviations are squared, and then the squares are added to obtain s , which is the sum of squares of the deviations.

One then computes the value of W from the formula: $W = \frac{s}{1/12 m^2 (n^3 - n)}$, where s = sum of squares of the observed deviations from the mean of R_j ; that is, $s = \sum [R_j - \left(\frac{n+1}{2}\right)m]^2$, m equals the number of panelists, n equals the number of products ranked, and $1/12 m^2 (n^3 - n)$ equals the maximum possible sum of the squared deviations, e.g., the sum s , which would occur with perfect agreement among m rankings.²

Tied Rankings

When tied rankings occur, the rankings are each assigned the average of the ranks they would have been assigned had no ties occurred.³

The effect of tied ranks is to depress the value of W as found by the above formula. If the proportion of ties is small, the effect is negligible, and the above formula may still be used.⁴ If the proportion of ties is large, the following correction may be incorporated into the above formula:

¹Ibid. . . .

²Ibid. , p. 231.

³Ibid. , p. 233.

⁴Ibid. , p. 234.

$T = \sum \frac{(t^3 - t)}{12}$, where t equals the number of products in a group tied for a given rank by any one panelist, Σ directs one to sum over all groups of ties within any one of the m rankings.¹ Thus, when the correction of ties is incorporated, $W = \frac{s}{1/12 m^2 (n^3 - n) - n \sum T}$, where $\sum T$ directs one to sum the values of T for all the m rankings.

Testing the Significance of W

The significance of any observed value of W may be tested by determining the probability associated with the occurrence, under the null hypothesis, that there is no difference in consumers preferences for the products being tested, of a value as large as the s with which it is associated.² The distribution of s under the null hypothesis has been worked out, and certain critical values have been tabled.³ One table which gives values of s for W 's significant at the .05 and .01 levels is applicable for 3 to 20 panelists and for 3 to 7 products.⁴ If an observed s is equal to or greater than the appropriate critical value shown in the table, then the null hypothesis may be rejected at that level of significance.

If the number of panelists is larger than 20, a similar test called the Friedman Two Way Analysis of Variance may be used to test whether the agreement among the panelists is higher than it would be by chance. The

¹ Ibid.

² Ibid., p. 235.

³ Ibid., p. 236.

⁴ Ibid., p. 286.

Friedman Test determines whether the rank totals, R_j , differ significantly.

This test involves the computation of a statistic which Friedman denotes

as X_r^2 . When m (the number of panelists) is larger than 10, and n (the number of products being ranked) is 3 or larger, or when m is larger than 5, and n is 4 or larger, X_r^2 may be considered to be distributed approxi-

mately as chi square with $n - 1$ degrees of freedom.¹ X_r^2 is found by the

following formula: $X_r^2 = \frac{12}{m n (n+1)} \sum_{j=1}^n (R_j)^2 - 3m (n+1) = m(n-1)W$ where m equals the number of rows

n equals the number of columns

R_j equals the sum of the ranks in j^{th} column

$\sum_{j=1}^n$ directs one to sum the squares of the sums of ranks over all n products.²

Under the null hypothesis, the sampling distribution of X_r^2 is approximated by the chi square distribution with $n - 1$ degrees of freedom so the probability associated with the occurrence under the null hypothesis of values as large as an observed X_r^2 is shown in a chi square table.³ If the value of X_r^2 as computed from the above formula is equal to or greater than that given in the X^2 table for a particular level of significance and a particular degree of freedom, then the implication is that the sums of the ranks for the various columns differ significantly; thus, the null hypothesis may be rejected at that level of significance.

¹ Ibid., p. 168.

² Ibid.

³ Ibid., p. 249.

If either the hedonic or attitude scales are used, only a slight adaptation is required to analyze the data in the same manner that rank data are analyzed.

The researcher should cast the observed evaluations in a table having m rows and n columns. The symbols m and n still represent the number of panelists and the number of products respectively. The next step is to rank the scores in each row, giving the highest score a rank of 1, the next highest score a rank of 2, etc. This step yields the same table which is used to analyze rank data. Again, tied scores should be assigned the average of the ranks they would have been assigned had no ties occurred.

An example of a transformation from scale scores to ranks is given with Table V containing the scale scores and Table VI containing the rankings obtained from the scale scores.

TABLE V. Scale Scores Obtained From Panelists

	Product			
	A	B	C	D
Panelist No. 1	9	7	3	5
Panelist No. 2	4	7	6	5
Panelist No. 3	5	5	4	7

TABLE VI. Rankings Obtained From Panelists' Scale Scores

	Product			
	A	B	C	D
Panelist No. 1	1	2	4	3
Panelist No. 2	4	1	2	3
Panelist No. 3	2.5	2.5	4	1
R_j	7.5	5.5	10	7

Example of the W and Friedman Test

The data obtained from the study concerning consumer preferences for sausage containing four different levels of sugar will be used to illustrate the use of the W and Friedman Test.

The null hypothesis is the different levels of sugar contained in sausage have no effect upon consumer preferences toward sausage. The alternative hypothesis is that the different levels of sugar contained in sausage do have an effect upon consumer preferences toward sausage.

The statistical test used to analyze the data is the Friedman Two Way Analysis of Variance. The coefficient of concordance will also be calculated to determine the divergency of agreement from perfect agreement.

Since 100 panelists ranked the four samples of sausage m equals 100. The alpha level was set at .05.

As calculated from $X_r^2 = m(n-1) W$, X_r^2 under the null hypothesis is approximately distributed as chi square with $n-1$ degrees of freedom. So the probability associated with the occurrence under the null hypothesis of a value as large as the observed value of X_r^2 may be determined by referring to a X^2 table.¹ The region of rejection consists of all values of X_r^2 which are so large that the probability associated with their occurrence under the null hypothesis is equal to, or less than, alpha equal to .05.

The sum of the ranks for the sausage containing no sugar was 279, for sausage with 1.0% sugar it was 258, for sausage with 1.5% sugar it

¹Ibid.

was 219, and for sausage with 2.0% sugar the sum of the ranks was 234.

$$W = \frac{s}{1/12 m^2 (n^3 - n) - n \sum T^2} = \frac{12 (2156.5)}{(100)^2 (4^3 - 4) - 4(2)} = .04313, \text{ because } s = 2156.5,$$

m equals the number of panelists (100), and n equals the number of products (4). The value for s was found in the table which appears in Appendix A.

The coefficient of concordance value of .04 indicates that the panelists differed among themselves as to the rankings of the products. The Friedman Test was used to determine if the agreement among the panelists was higher than it would be by chance. $X_r^2 = m(n-1) W = 100(3)(.04313) = 12.9$.

Reference to a X^2 table indicates that the critical value of $X_{.95}^2$ with n-1 or 3 degrees of freedom is 7.82. The calculated X_r^2 is larger than this critical value thus the null hypothesis is rejected. There is a difference in consumer preferences for sausage containing four different levels of sugar.

Advantage of the W and Friedman Test

The coefficient of concordance, W, measures the extent of the association among the m sets of panelists' rankings of the n products. It is useful in determining the agreement among the panelists. A significant value of W may be interpreted as meaning that the panelists are applying essentially similar standards in ranking the products being studied.²

¹Ibid.

²Siegel, op. cit., p. 237.

Kendall suggests that the best estimate of the true ranking of the n products is provided, when W is significant, by the order of the various sums of ranks, R_j .¹ If one accepts the criterion which the panelists have agreed upon (as evidenced by the significance of W) in ranking the products, then the best estimate of the "true" rankings of those products, according to this criterion, is provided by the order of the sums of ranks.² Thus, the W has special applications in providing a standard method of ordering products according to a consensus of consumers preferences.

The tests used to determine whether W is significant are over-all tests which determine whether the panelists' rankings depend on the products being tested.

Disadvantage of the W and Friedman Test

When tied rankings occur, the rankings must each be assigned the average of the ranks they would have received had no ties occurred if one is to obtain an accurate X^2 value. The effect of tied ranks is to depress the value of W and X^2_r . Therefore, one must make a special effort to assign to tied scores the average of the ranks that they would have been assigned had no ties occurred.

Individual Paired Tests to be Used if Over-All Test is Significant

The Chi Square Test of Independent Samples, the Friedman Two Way

¹Maurice G. Kendall, Rank Correlation Methods, London, Griffin, p. 87.

²Siegel, op. cit., p. 238.

Analysis of Variance, and a significant coefficient of concordance all indicate that there is a significant difference among the products being tested. None of these tests indicate whether the difference in consumer preference rankings between the other possible pairs of products being tested is significant.

Test Procedure

If, and only if, the statistical analysis of "over-all" differences in preferences is significant, chi square tests may be conducted to determine differences between all possible pairs of comparisons. It is also permissible to apply this test to any pair of products which the researcher specified before viewing the data. The appropriate formula for this chi square test is: $X^2 = 2 \sum_{i=1}^c \frac{(O_i - E_i)^2}{E_i}$, where X^2 is the chi square, O_i is the observed value for either product, E_i is the average or expected value, and c is the number of columns. Under the null hypothesis the sampling distribution of X^2 , as computed from this formula, follows the chi square distribution with $c-1$ degrees of freedom. The null hypothesis tested by this chi square test is that there is no difference in the consumer preferences for the two products being compared. The alternative hypothesis tested is that there is a difference in the consumer preferences for the two products.

To apply the chi square test of significance between paired products, one first arranges the frequencies in a c by 2 table. The two rows consist of the products being compared. The c columns consist of all of the

different ranks or scores given by the panelists to the two products. The number of times each product was ranked with a 1, 2, etc. is placed in the appropriate column of the table.

The next step consists of adding the number of times either product was ranked with a similar number. This involves the addition of the rankings appearing in each column to get an R_j . The R_j is then divided by 2 to obtain the average or expected value for either product in the j^{th} column.

Then the deviation of one observed frequency from the average or expected frequency is found. The deviation for each column is squared and divided by its average or expected value. The X^2 value is two times the sum of these column deviations squared and divided by their expected value.

Example of Paired Test

Consumer preferences for sausage containing four levels of sugar were found to be significant by both the W and Chi Square Test of Independence. Therefore, it is appropriate to use the X^2 paired comparison test to determine if consumers' preferences differ between any two products.

The null hypothesis tested is that there is no difference in consumer preferences for the sausage containing 1.5 percent sugar and sausage containing no sugar. The alternative hypothesis is that there is a difference in consumer preferences for the two kinds of sausage.

The X^2 test for two independent samples was used. The alpha level was set at .05, and N equals the number of panelists, which is 100.

The chi square as computed from $X^2 = 2\sum_{i=1}^c \frac{(O_i - E_i)^2}{E_i}$ has a sampling distribution which is approximated by the chi square distribution which has $c-1$ or 3 degrees of freedom. Critical values of chi squares are given in any X^2 table.¹ The region of rejection consists of all values of X^2 which are so large that the probability associated with their occurrence is equal to, or less than, alpha equal to .05.

Table VII contains the number of times panelists ranked each kind of sausage first, second, third, and fourth. It also includes the calculation of the X^2 value.

TABLE VII. An Example of the Paired Chi Square Test

Product	Observed Number of Times Ranked			
	First	Second	Third	Fourth
(O) Sausage with 1.5% sugar	33	31	20	16
Sausage with no sugar	18	19	29	34
Sum	51	50	49	50
Averaged or expected (E)	25.5	25.0	24.5	25.0
Observed minus expected				
(O-E)	7.5	6.0	- 4.5	- 9.0
(O-E) ²	56.25	36.0	20.25	81.0
$\frac{(O-E)^2}{E}$	2.21	1.44	.83	3.24
$X^2 = 2\sum_{i=1}^c \frac{(O_i - E_i)^2}{E_i} = 15.44$				

¹Ibid. , p. 249.

The $X^2_{.95}$ with c-1 or 3 degrees of freedom is 7.82; thus, the null hypothesis is rejected in favor of the alternative hypothesis. It appears that consumers prefer sausage with 1.5 percent sugar over sausage containing no sugar.

The other paired comparisons were made in a similar manner. A summary of these paired comparisons shows that: the 1.5% level was significantly preferred over the 1.0% level, but not the 2.0% level; the 2.0% level was significantly preferred over both the 1.0% and the 0% levels, and the 1.0% level was not significantly preferred over the 0% level.

For ease in presentation of these results, differences in preferences may be expressed in a manner whereby the kinds of sausage are listed from left to right from the lowest rank (rank 1 was the sample most preferred) to the highest rank. Thus, for this test the data would be expressed as follows: 1.5% 2.0% 1.0% 0%

Note: Any two treatments underscored by the same line are not significantly different at the 5% level.

Any two treatments not underscored by the same line are significantly different at the 5% level. Thus, the odds are 19 to 1 that the differences are not due to chance alone.

Advantage of Paired Test

The chief advantage of the paired chi square test is that it allows one to compare one product with another and to determine whether there is a significant difference among the preferences for each product.

Disadvantage of Paired Test

The paired X^2 test, when used after a significant over-all difference has been determined, may lead to fallacious conclusions. A significant value for a paired comparison, say a significant preference for product A over B, indicates only that preferences for the pair of products being evaluated are significantly different. This significance does not imply anything about the relationship between either A or B and another product, say C. This is true, although product B may have already been shown to be preferred over C. The paired comparison between product A and product C must still be made to draw conclusions on the pair.

CHAPTER VII

FIVE OR MORE SAMPLE TESTS

When a researcher is testing consumer preferences for more than four products, it is desirable to present the products in more than one setting or series. Several reasons for limiting the number of samples presented in one series to no more than four were discussed in Chapter III.

Therefore, when one attempts to determine consumers' preferences for as many as 16 or 20 different types of products, he must use a well organized, efficient design to accomplish his task. This chapter is devoted to a discussion and evaluation of designs and methods of data analysis which may be used when one is conducting consumer preference research on five or more samples.

The application of any of the techniques discussed in this chapter has many qualifications which must be considered when one is conducting a product test.

There are certain limitations which are inherent in the assumptions underlying each design and its related analysis of variance. Assumptions such as: random sampling, normality of distribution, homogeneity of variance, and extensive measurement are violated by data obtained from ranks or rating scales. One plus one on a rating scale does not necessarily equal two. Ratings may skew extremely as preferences cluster toward the positive end of a rating scale.

The role of each of the assumptions relative to the results of the analysis of variance should be considered.

The effect of non-normality is a loss of efficiency in estimation of means. How much efficiency is lost is difficult to assess but it is not thought to be serious.¹ Another effect of non-normality is to change the level of significance. Generally, too many significant effects will be found, e.g., the probability of the Type I error is larger than the tabled level. Instead of obtaining significance at the 5 percent level as read from the table, the true probability may be between 4 percent and 7 percent. The exact probability for any experiment is unknown. This approximation is not serious unless the experiment is a very delicate one. Non-normality usually results in heterogeneous variance.

Heterogeneous within treatment variances result in a loss of efficiency of the analysis of variance and a loss of sensitivity in the tests of significance.

Correlations between the random variations cause biases in the estimates of the population variance which can be positive or negative. Of course, this distorts the tests of significance.

The consequences of any of the assumptions not being met is a loss of efficiency of estimates and distorted tests of significance. These effects are not usually serious.² However, the researcher should realize that the

¹Elmer Remmingea, Mimeographed Outline for Analysis of Variance Course, Colorado State University, Winter Term, 1960, pp. 7-8.

²Ibid.

analysis of variance is an approximation and not an exact procedure.

The assumptions are usually less critical for large sample sizes because many statistical properties of tests can be obtained as the size of the sample approaches infinity.¹ The term large sample size is a relative term which must be left undefined because the distinction depends on the test being discussed. The researcher usually relies on past experience to guide him in choosing his definition of a large sample.

Thus, although the usual procedure followed in each type of design and its related analysis of variance is not entirely suited for the analysis of ranking or rating scale data, it can provide approximate probability statements. Such procedures can yield results which will serve as a basis for determining the direction future testing, marketing, or production should follow.

Balanced Incomplete Block Design Analyzed by Analysis of Variance

Test Procedure

If more than five samples are being tested, one may be able to utilize one of the balanced incomplete block designs as a method of presenting the samples to the panelists. The balanced lattice is a design which allows the researcher to compare each of the different products being tested to each of the remaining products an equal number of times. That is, any pair of products occurs once in the same row. In this design, all the

¹Manderscheid, op. cit., p. 30.

comparisons are made with approximately equal precision as each combination contains the same number of products, and each product occurs the same number of times in the complete design.

The number of products to be tested, must be an exact square unless one or more products are duplicated. The method of grouping into rows and columns is such that the product means can be adjusted for differences among the rows and columns of each square.¹

The 3 by 3 and 4 by 4 balanced lattices are the designs which are most likely to be used, because no more than four products should be compared in any one series. If the 3 by 3 design is used, nine products may be compared. Sixteen products may be compared if the 4 by 4 design is used. Of course, some products may appear more than once in the design, if one is interested in comparing a number of products which is neither 9 or 16. For example, if one wanted to compare eight products, he could place each product twice in the 4 by 4 design. Other replications could be made for tests involving other numbers of products.

The 3 by 3 balanced lattice design is given below:²

Replication I	Replication II	Replication III	Replication IV
<u>Block</u>	<u>Block</u>	<u>Block</u>	<u>Block</u>
(1) 1 2 3	(4) 1 4 7	(7) 1 5 9	(10) 1 8 6
(2) 4 5 6	(5) 2 5 8	(8) 7 2 6	(11) 4 2 9
(3) 7 8 9	(6) 3 6 9	(9) 4 8 3	(12) 7 5 3

¹William G. Cochran and Gertrude M. Cox, Experimental Designs (New York: John Wiley & Sons, 1957), p. 428.

²Ibid.

Where the numbers 1, 2 . . . , 9 identify the nine different types of products which are being tested. Thus, panelist number one will rank products 1, 2, and 3. Panelist number two will rank products 4, 5, and 6. Panelist number twelve will rank products 7, 5, and 3. Then the design is repeated again so panelist number thirteen will rank products 1, 2, and 3. The design can be replicated many times.

An effort should be made to complete the design each time it is repeated. Therefore, the number of panelists used should be a multiple of 12. The replies of all panelists should be checked to make sure that they have ranked all the products in the series.

The results obtained from either a 3 by 3 or a 4 by 4 balanced lattice design may be analyzed by analysis of variance and paired comparison tests. Calvin's method of adjusting the product means for blocks should be used if panelists' give their evaluation on an attitude or hedonic scale.¹ An analysis of variance table may be used to organize the computational formulas for computing the needed sum of squares and degrees of freedom. The total, product effect, and error sums of squares are required in testing for significance. However, one should be interested in any interactions which may occur. The computational formulas one uses to obtain the components in the analysis of variance table can be found in Appendix B. Examples of other balanced incomplete block designs appear in Appendix C.

¹Lyle Calvin, "Doubly Balanced Incomplete Block Designs for Experiments in which the Treatment Effects are Correlated," Biometrics, X (1954), pp. 61-88.

Example of an Analysis of a 4 by 4 Balanced Lattice Experiment

Paired hams were removed from carcasses of pigs weighing between 180 and 220 pounds live weight. A total of 64 hams was used for this study. The hams were then randomized into 16 different types of products as follows:

Product	Salt	Sugar	Product	Salt	Sugar
1	1.5%	0%	9	2.5%	0%
2	1.5%	1%	10	2.5%	1%
3	1.5%	2%	11	2.5%	2%
4	1.5%	3%	12	2.5%	3%
5	2.0%	0%	13	3.0%	0%
6	2.0%	1%	14	3.0%	1%
7	2.0%	2%	15	3.0%	2%
8	2.0%	3%	16	3.0%	3%

Presentation to the panel made use of a 4 by 4 balanced lattice design so that each of the 16 different types of products was compared to each of the other types of products an equal number of times. In this design, there were 20 possible combinations of the 16 types of products, compared four at a time. The 4 by 4 balanced lattice design is:

Replication I

Block

(1) 1 2 3 4
 (2) 5 6 7 8
 (3) 9 10 11 12
 (4) 13 14 15 16

Replication II

Block

(5) 1 5 9 13
 (6) 2 6 10 14
 (7) 3 7 11 15
 (8) 4 8 12 16

Replication III

Block

(9) 1 6 11 16
 (10) 5 2 15 12
 (11) 9 14 3 8
 (12) 13 10 7 4

Replication IV

Block

(13) 1 14 7 12
 (14) 13 2 11 8
 (15) 5 10 3 16
 (16) 9 6 15 4

Replication V

Block

(17) 1 10 15 8
 (18) 9 2 7 16
 (19) 13 6 3 12
 (20) 5 14 11 4

Where the numbers 1, 2, ..., 16 refer to the 16 types of products containing different levels of salt and sugar.

The balanced lattice design was chosen because preferences for the 16 types of ham had to be compared. This design is most efficient in this respect, since it allows one to compare preferences for ham containing 16 different levels of salt and sugar with a minimum of panel members. The balanced lattice design was also chosen because it has the advantage of being an organized method of presenting the 16 types of ham to the large-scale panel.

Each of the 160 panel members received a sample of each of four types of products at each of the two tables from which ham was being served. The samples were tested independently. A randomized design was drawn up, so that the respondent's previously assigned number was used to determine the combination of samples he was to taste. The design was such that no participant tasted the same four samples at both tables. This was done by using the 20 possible combinations of products in a randomized 4 by 4 balanced lattice design in a consecutive manner at the first table (block 1 was given to no. 1 panelist, block 2 was given to no. 2 panelist, ..., and block 20 was given to the no. 20 panelist). That is, panelist no. 1 was given products no. 1, 2, 3, and 4 and panelist no. 20 was given products no. 5, 14, 11, and 4 at the first table. At the second table, block 20 was given to the no. 1 panelist, block 19 to the no. 2 panelist, ..., and block 1 was given to the no. 20 panelist. In other

words, panelist no. 1 received products no. 5, 14, 11, and 4 at the second table. This design was repeated 8 times at each table for the 160 consumers.

The samples were presented to consumers four at a time on coded paper plates. The samples were coded by symbols to prevent the possible influence of ranking association by use of letters or numbers. The four symbols used at the first table were %, (), *, and &. The symbols used at the second table were &, (), %, and #. The types of products were rotated so that each product was placed by each symbol an equal number of times. The position that each symbol occupied on the paper plate was also rotated, so that each symbol appeared in each of the four positions on the plate an equal number of times.

The panelists were asked to rank the four samples in the order of their preference, using 1 for the sample most preferred, 2 for the sample second-most preferred, 3 for the sample third-most preferred, and 4 for the sample least preferred.

The analysis of variance computations appear in Appendix D and are summarized in Table VIII.

The null hypothesis tested by the analysis of variance is that the types of products (different percentages of salt, sugar, and any interaction between them) have no effect on consumer preferences for ham. The alternative hypothesis tested is that the types of products have an effect on consumer preferences. Although the null hypothesis cannot be rejected for salt level and the interaction of salt and sugar levels, there appeared

to be a significant difference in the preferences for ham containing different levels of sugar. The difference was significant at approximately the 1 percent level as the normal analysis of variance computations would indicate that the null hypothesis should be rejected for an F value of 4.16 when the alpha level equals .01 and the degrees of freedom in the F ratio are 3 and 64. The hypothesis that the percentage of sugar has no effect on consumer preferences for ham was rejected. The odds were greater than approximately 99 to 1 that the agreement in preferences between sugar levels was not due to chance alone. Tests for linearity of the salt and sugar treatments were nonsignificant.

TABLE VIII. Analysis of Variance of a 4 by 4 Balanced Lattice Design to Determine Consumer Preferences Among Hams Prepared with Four Levels of Sugar and Four Levels of Salt.

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Square	F
Treatments	15	439.2	29.29	1.48
Sugar	3	247.2	82.40	4.16 ^a
Salt	3	36.7	12.23	2.75
Sugar X Salt	9	155.3	17.25	.87
Error	64	1266.8	19.79	
Total	79	1706		

^aSignificant at the 1% level.

The analysis of variance may be regarded as a preliminary test, since it shows only that the percentage of sugar contained in ham has an effect on consumer preference for ham. Multiple comparison techniques such as

Scheffe's or Turkey's tests should be used to draw more precise conclusions.

Advantage of Balanced Lattice Designs

The balanced lattice design has the advantage of being an organized method of presenting many different products of the panelists. Similar results, obtained with both balanced lattice and triangle methods of testing, indicate that the balanced lattice design can adequately reflect differences in preferences. The advantage of this method of presentation is that it allows one to compare preferences for a large number of products from a minimum number of panel members. This is especially important if one is comparing a large number of products. In such cases, the use of paired comparisons or triangle tests would require a large number of respondents. In this example, use of the paired comparison method would require 120 respondents, if each of the 16 different kinds of ham were to be compared with each other once, and if each respondent made only one comparison. Therefore, it would have taken 1,920 panelists to compare each of the 16 different kinds of ham with each other 16 times. By using the 4 by 4 balanced lattice design, only 20 panel members were required to compare each of the 16 kinds of ham with each other one time. By each panel member making two comparisons of four samples each, 160 panel members compared each of the 16 different kinds of ham with each other 16 times, and significant differences in preferences were obtained.

The balanced lattice design allows one to compare each of the products

being tested to each of the remaining products an equal number of times. It also allows one to determine whether a significant fatigue or order effect is present in the study.

Disadvantages of Balanced Lattice Design

The presentation of samples to panelists by the balanced lattice design is a somewhat complicated procedure, which requires competent researchers, who are careful to present the samples in the proper manner.

The analysis of data obtained from a balanced lattice design is complicated if there are data missing.

It must also be recalled that this method of data analysis is a parametric statistical analysis, which requires independent observations and measurement in the strength of at least an interval scale. However, consumer preference rankings are measured only in an ordinal scale and are not independent. This means that only approximate probability statements can be placed on any of the findings.

Fractional Factorial Design

Test Procedure Illustrated on an Example

When a fractional factorial design is used, only a portion of the total possible factor-level combinations is used. Such a reduction saves a great deal in the cost of the study, however, it increases the risk that important interactions between factors may either be confused with one another, or lost completely.¹

¹Charles C. Beazley, "Product Factors vs. Consumer Acceptance," (General Foods Corporation, Battle Creek, Michigan, Unpublished Mimeograph, 1963), p. 4.

The researcher should first reduce his formula factors to a number which he believes will scan most of the product attributes. Suppose there are six such factors he is interested in. He should also make a decision about the number of levels of each factor. Because of design limitations, the levels chosen are usually 2, 4, or 8 levels of each factor. Table IX shows the factors and their levels in an experiment conducted by the General Foods Corporation.¹

TABLE IX. Factors and Levels of Factors in Experiment

Code Letters	Factor	Levels
A	I	1 low 2 high
BC	II	1 in 2 descending 3 order 4
D	III	1 low 2 high
E	IV	1 low 2 high
FG	V	1 in 2 descending 3 order 4
H	VI	1 low 2 high

The researcher uses this information to aid in his choosing of the formulas to be used in consumer testing. Table X illustrates the formulas

¹ Researchers at General Foods Corporation planned, executed, and analyzed this experiment.

chosen in the example. In this design, each level of each factor appears an equal number of times.

TABLE X. Formulas to be Tested in Experiment Factors, Codes and Levels

Formula Number	A	BC	D	E	FG	B
1	2	3	1	2	3	1
2	2	1	1	1	4	1
3	2	4	2	2	4	2
4	2	2	2	2	1	1
5	1	4	1	1	3	2
6	2	4	2	1	2	1
7	2	2	2	1	3	2
8	1	3	2	2	2	2
9	1	1	2	2	3	1
10	1	4	1	2	1	1
11	2	1	1	2	2	1
12	1	1	2	1	1	2
13	1	2	1	1	2	1
14	2	3	1	1	1	2
15	1	2	1	2	4	2
16	1	3	2	1	4	1

The researcher must then determine how the test formulas should be consumer tested. In this example, a paired comparison test was used, with each of the 16 formulas being tested against all others. The sequence in which the panelists test their two samples should also be balanced so that each pair, such as formula 1 and formula 2, is tested an equal number of times in both orders; e. g. , 1-2 and 2-1.¹ In this example, each pair order is presented to a panelist, so there are 16 x 15 or 240 presentations in all. Table XI illustrates the 240 pair orders.

¹Ibid. , p. 4.

TABLE XI. Sample Pairs Presented to Panelists in Experiment

First Sample (Formula)	Second Sample (Formula)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1																
2																
3																
4																
5																
6																
7																
8																
9																
10																
11																
12																
13																
14																
15																
16																

One presentation is made of each pair shown

In the consumer preference tests, each panelist tests two samples and completes a questionnaire which may contain information on overall preference and degree of preference, attribute preferences, and overall acceptability (hedonic) ratings for the two samples. Table XII illustrates the type of ratings which may be used and the method in which responses may be scored.¹

¹ Ibid., pp. 5, 13.

TABLE XII. Typical Questionnaire Data Used in the Experiment

<u>Overall Preference Score</u>	
<u>Answer</u>	<u>Score</u>
Strong preference for first sample	+2
Slight preference for first sample	+1
No preference	0
Slight preference for second sample	-1
Strong preference for second sample	-2
<u>Overall Acceptability Rating</u>	
<u>Answer</u>	<u>Score</u>
Like it very much	7
Like it quite a bit	6
Like it slightly	5
Neither like nor dislike it	4
Dislike it slightly	3
Dislike it quite a bit	2
Dislike it very much	1

All of the raw scores are analyzed by first finding average scores for each of the 16 formulas tested. The next step consists of using these scores to determine the effects of the formula factors studied.¹ In this example, the results of the initial analysis of overall preference scores and overall acceptability ratings are given in Table XIII. The analysis yields average preference scores and average ratings, both of which relate to the scales contained in Table XII. The "Scheffe" or analysis of variance of scored preferences may be used in the case of preference

¹ Ibid., p. 5.



scores, and the balanced incomplete block analysis in the case of ratings.¹

A brief discussion of the Scheffe method is presented in Appendix E.

TABLE XIII. Preference Scores and Acceptability Ratings Obtained in the Experiment

Formula Number	Average Overall Preference Scores	Average Overall Acceptability Ratings
1	.12	5.78
2	.16	5.73
3	.19	5.93
4	.02	5.97
5	-.28	5.28
6	.08	5.84
7	-.05	5.70
8	-.33	5.14
9	-.19	5.28
10	.23	5.93
11	.02	5.76
12	-.02	5.29
13	-.05	5.93
14	.16	6.03
15	-.19	5.53
16	.12	5.82

Conclusions should not be drawn from the results on single samples, for each sample receives very limited testing. Therefore, each sample's average scores are subject to a large amount of chance variation.

Instead, a procedure called standard factorial analysis of variance should be used to find the real effects of the formula factors. This procedure yields numbers called "factorial contrasts," which are simply the

¹Scheffe's method is presented in an article by Otto Dykstra, Jr., "Factorial Experimentation in Scheffe's Analysis of Variance for Paired Comparisons," American Statistical Association Journal (Vol. 53, June 1958), pp. 529-542.

differences between the rating averages at the high levels of the factor minus the rating averages at the low levels for each factor.¹ Certain contrasts relate to the effects of factors by themselves; others relate to interactions between two factors. This analysis also indicates which of the contrasts are significant.

This is done by ordering the factorial contrasts on the basis of absolute magnitude. The largest factorial contrast is then tested against the variation of the remaining contrasts. Thus, if the factorial contrasts are denoted as $C_1, C_2, \dots, C_n$ where $|C_i| \geq |C_{i+1}|$, then the variance of C_j equals $(\sum_{i=j+1}^{15} C_i^2)/(15-j)$, in this example, where the number of factorial contrasts is 15. One may then test the significance of the contrast by using an F test with 1, 15-j degrees of freedom, as $C_j^2/(\text{Variance of } C_j)$ is approximately distributed as an $F(1, 15-j)$ distribution.² Table XIV illustrates some of the contrasts from the study.

Once the contrasts have been calculated and tested for significance, the researcher should calculate pertinent average scores or ratings to illustrate the effects which were found to be significant.³ Table XV illustrates some of the averages for the study. For example, Factor A is found to be significant in both overall preferences and the overall acceptability ratings. In Table XV it can be seen that the higher level of factor A gives

¹Beasley, op. cit., p. 6.

²Interview with Charles Beasley, January 21, 1964.

³Beasley, op. cit., p. 7.

higher scores and ratings, so the researcher would use this level in future formulas. Similarly, he would use the lower level of factor H.

TABLE XIV. Factorial Contrasts for Selected Consumer Test Results Obtained in the Experiment

Factor	Overall Preference Scores	Overall Acceptability Ratings
Average	.00	5.62
A	.18*	.32*
B	.07	.07
C	.05	-.10
D	-.04	-.12
E	-.03	-.04
F	-.03	-.10
G	.17*	.19
H	-.12	-.20
BC	-.01	.16
FG	.00	.06
BE	.07	-.01
DF	.11*	.23
DE	-.08	-.05
AE	.03	.07
CD	.07	.04

* Indicates significant effect

TABLE XV. Illustration of Some Significant Factor Effects Using Average Scores Obtained in the Experiment

Factor	Overall Preference	Overall Acceptability Ratings
A		
1 low	-.09	5.52
2 high	.09	5.84
H		
1 low	.06	--
2 high	-.06	--

In highly fractionated factorial designs, such as the one illustrated here, it is sometimes difficult to identify exactly which of several interaction effects is significant. Often in such cases, the individual attribute preferences and ratings may hold the key to proper identification, as common sense would tend to eliminate many of the alternatives.¹

Advantage of Fractional Factorial Designs

The objective of fractional factorial designs is to obtain information on the main effects and as many of the interactions as seems necessary with a smaller number of observations than is required by the complete design.

Careful consideration of the best combinations of experimental treatments is needed, and the theory brings out in detail what becomes of the interactions neglected in any particular design, and what are the results if, unexpectedly, they are not negligible in reality.²

Fractional factorial designs are used most successfully in investigations involving several possible factors, of which only a few have an appreciable effect.³ These designs allow one to determine whether a significant order effect is present in the experiment.

Disadvantages of Fractional Factorial Designs

The main disadvantage in the use of these designs in consumer preference testing is that they become complex when many factors are studied

¹Ibid., p. 9.

²Owen L. Davies, The Design and Analysis of Industrial Experiments (Oliver and Boyd, London, 1954), p. 252.

³Ibid., p. 470.

and when samples are presented to a large number of panelists. Another disadvantage is that this is a parametric technique, which required independent observations and measurement of at least an interval scale.

Consumer panel data are not independent and are measured only in an ordinal scale. This means that only approximate probability statements can be placed on any of the findings.

Multiple Comparison Tests

Tukey's Simultaneous Comparisons

Test Procedure

The Tukey Method of Simultaneous Comparisons involves the use of the q statistic where q is equal to the range between the largest and smallest product means divided by the estimated standard deviation, S_p . The q table may be used to test simultaneously hypotheses about all possible differences of the means of the type $\mu_1 - \mu_2, \mu_1 - \mu_3$, etc. as well as all possible linear expressions of the form $a_1\mu_1 + a_2\mu_2 + \dots + a_k\mu_k$ where $a_1 + a_2 + \dots + a_k = 0$ and the sum of the positive a 's equal to 1. The symbol a is used to indicate the coefficient appearing before the means of the products which are being contrasted. From random samples from k normal populations with the same variance, the chance that all comparisons simultaneously satisfy $\frac{-q_{1-\alpha} S_p}{\sqrt{n}} (a_1\bar{x}_1 + a_2\bar{x}_2 + \dots + a_k\bar{x}_k) \text{ minus } (a_1\mu_1 + a_2\mu_2 + \dots + a_k\mu_k) < \frac{q_{1-\alpha} S_p}{\sqrt{n}}$ is equal to $1 - \alpha$, where the value of $q_{1-\alpha}$ is read from the q table.¹

¹ Dixon and Massey, Jr., op. cit., p. 153. Table on p. 442.

Thus confidence limits may be placed on all paired comparisons by using $\bar{X}_A - \bar{X}_B \pm \frac{q S_p}{\sqrt{n}}$. In this formula, $\bar{X}_A$ and $\bar{X}_B$ are the means from the two products being compared, q is obtained from the q table for the desired alpha level where k is the total number of means, and ν is the degrees of freedom in S_p^2 . The error mean square in the analysis of variance table is S_p^2 so $S_p = \sqrt{\text{error mean square}}$. The n refers to the number of observations in each sample. If zero is contained in the computed confidence interval, there is no significant difference between the two means. All confidence intervals are going to be correct at the same time at the specified alpha level, thus, the chance that any one confidence interval is wrong is very small.

Example of Tukey's Simultaneous Comparison

The F test in the study to determine consumer preferences for ham showed an over-all significant difference between ham containing different levels of sugar.

The Tukey method could be used to determine whether ham containing any one level of sugar was preferred over ham with other levels of sugar.

The sum of the rankings for ham cured with the four percentages of sugar were: 826 for 0% sugar, 802 for 1% sugar, 742 for 2% sugar, and 830 for 3% sugar. The greater the preference, the smaller was the rank score, for 1 was given to the most preferred sample. The design was replicated 20 times, so the respective means were: 41.4, 40.1, 37.1, and 41.5.

The null hypothesis is that the population contrasts are equal to zero; e.g., that zero is contained in the confidence interval. In other words, the null hypothesis states that: $\mu_0=\mu_1, \mu_0=\mu_2, \mu_0=\mu_3, \mu_1=\mu_2, \mu_1=\mu_3$ and $\mu_2=\mu_3$. The alternative hypothesis is that the confidence interval does not contain zero, e.g., $\mu_0\neq\mu_1, \mu_0\neq\mu_2, \mu_0\neq\mu_3, \mu_1\neq\mu_2, \mu_1\neq\mu_3$ and $\mu_2\neq\mu_3$. The alpha level is set at approximately .05. The statistic used is the $\bar{X}_A - \bar{X}_B \pm \frac{qS_p}{\sqrt{n}}$ confidence limit. Table XVI shows the population contrasts obtained from: $\bar{X}_A - \bar{X}_B \pm q_{.95}$ (Total number of means, degrees freedom in S_p^2) $\frac{S_p^2}{\sqrt{n}}$ equals $\bar{X}_A - \bar{X}_B \pm q_{.95} (4, 64) 4.49/\sqrt{16}$ equals $\bar{X}_A - \bar{X}_B \pm 3.71(1.112)$ $\bar{X}_A - \bar{X}_B \pm 4.16$.

TABLE XVI. Example of Tukey's Simultaneous Comparisons on Consumer Preferences for 16 Types of Ham

$\bar{X}_0$ (41.3)	$\bar{X}_1$ (40.1)	$\bar{X}_2$ (37.1)	$\bar{X}_3$ (41.5)	Confidence Limits	Population Contrast
1	0	-1	0	$\bar{X}_0 - \bar{X}_2 \pm \frac{qS_p}{\sqrt{n}} = 4.2 \pm 4.16$	$\mu_0 - \mu_2^*$
1	-1	0	0	$\bar{X}_0 - \bar{X}_1 \pm \frac{qS_p}{\sqrt{n}} = 1.2 \pm 4.16$	$\mu_0 - \mu_1$
1	0	0	-1	$\bar{X}_0 - \bar{X}_3 \pm \frac{qS_p}{\sqrt{n}} = -.2 \pm 4.16$	$\mu_0 - \mu_3$
0	1	-1	0	$\bar{X}_1 - \bar{X}_2 \pm \frac{qS_p}{\sqrt{n}} = 3.0 \pm 4.16$	$\mu_1 - \mu_2$
0	1	0	-1	$\bar{X}_1 - \bar{X}_3 \pm \frac{qS_p}{\sqrt{n}} = -1.4 \pm 4.16$	$\mu_1 - \mu_3$
0	0	1	-1	$\bar{X}_2 - \bar{X}_3 \pm \frac{qS_p}{\sqrt{n}} = -4.4 \pm 4.16$	$\mu_2 - \mu_3^*$

* Indicates significance when alpha equals .05.

Therefore, the null hypothesis is rejected for $\mu_0 - \mu_2$ and $\mu_2 - \mu_3$. One can say that ham containing 2% sugar is preferred over ham containing 0% .

or 3% sugar. This can be summarized by arranging the scores as they appear in Table XVII.

TABLE XVII. Consumer Preferences Among Sugar Differences in Ham at Approximately the 5% Level of Significance

Treatment:	2% sugar	1% sugar	0% sugar	3% sugar
Means:	<u>37.1</u>	<u>40.1</u>	41.3	41.5

Note: Any two means not underscored by the same line are significantly different. Any two means underscored by the same line are not significantly different. The smaller the mean, the higher the percentage.

Advantage of Tukey's Method

The Tukey Method of Simultaneous Comparisons gives slightly shorter confidence intervals for paired comparisons than the Scheffe Method when the number of rankings in each mean is equal. The Tukey method has been proven and is an established statistical technique.

Disadvantage of Tukey's Method

The Tukey Method of Multiple Comparisons is applicable only for cases in which the number of rankings is the same for each mean. This is the usual case in preference tests, hence Tukey's method may usually be used appropriately. The Tukey method is a parametric technique which requires independent observations and measurement in the strength of at least an interval scale. However, consumer preference rankings may not be independent and are measured only in an ordinal scale. This means that only approximate probability statements can be placed on any of the findings.

Scheffe's Method of Multiple Comparisons

Test Procedure

Suppose n_i observations are taken randomly from a normal distribution with μ_i and σ^2 , where $(i=1, 2, \dots, k)$. The sample means can be represented by $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k$. The S_p^2 is the estimate of σ^2 and is the error mean squares in the analysis of variance table. The number of degrees of freedom corresponding to S_p^2 is identified as γ . The symbol a_i is used to indicate the coefficients appearing before the means which are being contrasted.

The probability that all possible 100 $(1-\alpha)$ % confidence intervals of the form: $\sum_{i=1}^k a_i \bar{X}_i \pm \sqrt{k-1} F_{1-\alpha}(k-1, \gamma) \sqrt{\sum_{i=1}^k \frac{a_i^2}{n_i}} S_p$ on the linear combination $\sum_{i=1}^k a_i \mu_i$, for all comparisons $\sum_{i=1}^k a_i \mu_i$ are satisfied simultaneously is $(1-\alpha)$. The $\sum a_i = 0$ in this method. The probability that all confidence intervals are correct is 95.

Example of Scheffe's Method

The Scheffe Method of Multiple Comparisons can be used to analyze the individual means for ham containing each level of sugar.

The means for the rankings for ham cured with the four percentages of sugar were: 41.3 for 0% sugar, 40.1 for 1% sugar, 37.1 for 2% sugar and 41.5 for 3% sugar.

The null hypothesis is $\mu_0 - \mu_1 = 0, \mu_0 - \mu_2 = 0, \mu_0 - \mu_3 = 0, \mu_1 - \mu_2 = 0, \mu_1 - \mu_3 = 0$ and $\mu_2 - \mu_3 = 0$. The alternative hypothesis is that all differences between μ_i 's do not equal zero.

The alpha level was set at approximately .05. The test statistic is:

$\sum a_i \bar{X}_i \pm \sqrt{(k-1)F_{1-\alpha}(k-1, \gamma)} \sqrt{\frac{\sum a_i^2}{n_i}} S_p$. The term $\sqrt{\frac{\sum a_i^2}{n_i}}$ is constant for all paired comparisons as n_i is equal in all paired comparisons, e. g. , it equals $\sqrt{\frac{(1)^2 + (-1)^2}{16}}$ or .35. The expression $\sqrt{(k-1)F_{1-\alpha}(k-1, \gamma)} = \sqrt{3F_{.95}(3, 64)} = 2.76$.

TABLE XVIII. Example of Scheffe's Multiple Comparisons on Consumer Preferences for 16 Types of Ham

+1	-1	Confidence Limits	Population Contrast
$\bar{X}_3 - \bar{X}_2$	$= 41.5 - 37.1$	$\pm (2.76) (.35)(4.45) = 4.4 \pm 4.3$	$\mu_3 - \mu_2^*$
$\bar{X}_3 - \bar{X}_1$	$= 41.5 - 40.1$	$\pm (2.76) (.35)(4.45) = 1.4 \pm 4.3$	$\mu_3 - \mu_1$
$\bar{X}_3 - \bar{X}_0$	$= 41.5 - 41.3$	$\pm (2.76) (.35)(4.45) = .2 \pm 4.3$	$\mu_3 - \mu_0$
$\bar{X}_0 - \bar{X}_2$	$= 41.3 - 37.1$	$\pm (2.76) (.35)(4.45) = 4.2 \pm 4.3$	$\mu_0 - \mu_2$
$\bar{X}_0 - \bar{X}_1$	$= 41.3 - 40.1$	$\pm (2.76) (.35)(4.45) = 1.2 \pm 4.3$	$\mu_0 - \mu_1$
$\bar{X}_1 - \bar{X}_2$	$= 40.1 - 37.1$	$\pm (2.76) (.35)(4.45) = 3.0 \pm 4.3$	$\mu_1 - \mu_2$

* Significantly different at approximately .05 level.

The null hypothesis is rejected, because one confidence interval does not include zero. Therefore, this multiple comparison test indicates that ham containing 2% sugar is significantly preferred over ham containing 3% sugar. This is not the same result obtained by the Tukey Method. However, one should note that the confidence interval obtained by Scheffe's Method is longer than the confidence interval obtained by Tukey's Method.

Advantages of Scheffe's Method

The main advantage of Scheffe's method is that the data can be observed

before confidence intervals are determined. That is, one does not have to specify, before the experiment is conducted, which particular paired comparisons he is interested in testing.

Another advantage is that Scheffe's method can be used to test means which arise from either equal or unequal numbers of observations.

Disadvantage of Scheffe's Method

The confidence interval for paired comparisons given by a computation by the Scheffe method is longer than the confidence interval obtained from Tukey's method, if the number of observations going into each mean is equal.

Another disadvantage is that this is a parametric technique, which requires independent observations and measurement of at least an interval scale, but consumer panel data may not be independent and are measured only in an ordinal scale. Thus, only approximate probability statements can be placed on any of the findings.

CHAPTER VIII

TESTS TO DETERMINE WHETHER A PANELIST WILL PAY MORE FOR HIS PREFERRED PRODUCT

When a panelist ranks one product over another product, each of which contains no price level, he indicates his choice if the price is the same for both products. In such a test, he does not indicate that he would pay more to obtain his preferred product instead of the other product. Thus, in many studies, it is desirable to obtain an evaluation of how much the panelist would be willing to give up in order to get his preferred product instead of the other products being tested. The following methods have been used to make just such an evaluation.

Use of the VonNeumann-Morgenstern Model in Measuring the Relative Utility of Two Products

Test Methodology and Assumptions

Throughout a study of this type, it must be assumed that the panelist acts in a rational manner. According to Edwards (1954), one requirement for rationality is that "the consumer can (at least) weakly order the states into which he can get, and he makes his choices so as to maximize something."¹

¹W. Edwards, "The Theory of Decision Making," Psychological Bulletin, Vol. 51 (1954), p. 381.

Two requirements are then necessary before a consumer can place these states into (at least) a weak ordering. First, the consumer must be able to indicate that he prefers one situation to the other or is indifferent with regard to them. The second requirement for weak ordering is that all preferences must be transitive. In other words, if a consumer prefers a known-brand product to an unknown brand product, and this latter brand to two dollars, then he must also prefer the known brand product to the two dollars. Similarly, if he is indifferent with regard to the two brands and with regard to one brand and money, then he must also be indifferent with regard to the other brand and money.

A further requirement for the existence of rationality is that the consumer make his choices in such a way as to maximize something. The fundamental content of the notion of maximization is that the consumer always chooses the best alternative from among those open to him, as he sees it. In the theory of riskless choices, the consumer is usually assumed to maximize utility.

The traditional mathematical notion for dealing with risky decisions is in the notion that choices should be made so as to maximize expected value. The expected value of a random variable is found by multiplying the value of each possible outcome by its probability of occurrence and summing these products across all possible outcomes. In symbols: $EV = p_1 \$1 + p_2 \$2 + \dots + p_n \$n$, where p stands for probability, $\$$ stands for the value of an outcome, and $p_1 + p_2 + \dots + p_n = 1$. The assumption that

people behave in this way is contradicted by observable behavior in risky situations. People are willing to buy insurance or lottery tickets, even though the selling agency makes a profit. Therefore, in this study, the VonNeumann and Morgenstern model was used, and the consumer was assumed to act so as to maximize expected utility rather than expected value. VonNeumann and Morgenstern point out that the usual assumption, that a consumer can always say whether he prefers one state to another or is indifferent between them, needs only to be slightly modified in order to imply cardinal utility. This modification consists of adding that "economic man" can also completely order probability combinations of states. The assumption of VonNeumann and Morgenstern is that imagined events can be combined with probabilities, and therefore the utilities attached to the events can also be combined with probabilities. It appears that consumers should choose the situation with the highest expected utility (EU), where EU is defined as follows: $EU = \sum_i p_i u_i$. The symbol u_i is the utility or subjective value of the i^{th} outcome of the bet. Therefore, this proposal is a suggestion that the utility or subjective value of a given sample may be substituted for its objective value in the calculation of an expected value.

VonNeumann and Morgenstern have used this equation as a definition of utility. This assumes that expected utility maximization is a theory about what people do, rather than a normative theory.

Example of VonNeumann-Morgenstern Method

A study to determine whether consumers will pay a premium for one brand of turkey over another brand of turkey will serve as an example of the manner in which the VonNeumann-Morgenstern model can be applied to consumer preference studies.¹

This study was conducted in conjunction with the M. S. U. -W. S. U. Detroit Consumer Preference Panel. In this study, two metal trays were filled with chipped ice, and a turkey was placed in each for use during the panel. These turkeys were uniform in appearance and weight, but were packaged with differently branded bags. One bag was that of a well-known brand which is sold in the Detroit Market and the other represented a brand which has never been sold in this market. A pair of one dollar bills was then placed next to the well-known brand. This display was identified as Display One. The other tray containing the unknown brand was labeled Display Two. The well-known brand, hereafter referred to as KBT, was identified by the typewriter symbol *. The two dollars was identified by the symbol #, and the unknown brand, hereafter referred to as UBT, was marked as %. Consumers were told that each of the turkeys weighed the same and that the present market price of turkeys of this weight was four dollars.

Each panel member was given a card as shown in Figure V. They were then asked to indicate which of the three items in the two displays would be their first, second, and third selections, if awarded their choice as a

¹James C. Makens, a fellow graduate student, cooperated with the writer in the planning, execution, and evaluation of this study which represents the joint efforts of Mr. Makens and the author.

I. Rank the following products: * ____
 # ____
 % ____

II. Check either Display One or Display Two in each of the 11 stores.

DISPLAY ONE

DISPLAY TWO

<u>Store One</u>	
All chances are Sample * turkey, No chances of Sample # money ()	All chances are Sample % turkey, a sure choice ()
<u>Store Two</u>	
9 chances are Sample * turkey, 1 chance is Sample # money ()	All chances are Sample % turkey, a sure choice ()
<u>Store Three</u>	
8 chances are Sample * turkey, 2 chances are Sample # money ()	All chances are Sample % turkey, a sure choice ()
<u>Store Four</u>	
7 chances are Sample * turkey, 3 chances are Sample # money ()	All chances are Sample % turkey, a sure choice ()
<u>Store Five</u>	
6 chances are Sample * turkey, 4 chances are Sample # money ()	All chances are Sample % turkey, a sure choice ()
<u>Store Six</u>	
5 chances are Sample * turkey, 5 chances are Sample # money ()	All chances are Sample % turkey, a sure choice ()
<u>Store Seven</u>	
4 chances are Sample * turkey, 6 chances are Sample # money ()	All chances are Sample % turkey, a sure choice ()
<u>Store Eight</u>	
3 chances are Sample * turkey, 7 chances are Sample # money ()	All chances are Sample % turkey, a sure choice ()
<u>Store Nine</u>	
2 chances are Sample * turkey 8 chances are Sample # money ()	All chances are Sample % turkey, a sure choice ()
<u>Store Ten</u>	
1 chance is Sample * turkey 9 chances are Sample # money ()	All chances are Sample % turkey, a sure choice ()
<u>Store Eleven</u>	
All 10 chances are Sample # money, No chance of Sample * turkey ()	All chances are Sample % turkey, a sure choice ()

CODE: * is sample of KBT, % is sample of UBT.

FIGURE V. Questionnaire Used in Measuring Relative Utilities of Two Brands of Turkey

prize. Here, then, was an ordinal measurement, since it involved only a choice among three alternatives.

This part of the test was designed to show quickly which consumers were willing to accept two dollars rather than a turkey worth twice this much. In this way, the non-consumers of turkey could be separated from those who may use turkey. The panelists who preferred \$2 to both turkeys completed the entire questionnaire, but their responses were not included in the analysis.

The second part of the test was designed to measure the relative utilities between the brands, or how much a consumer was willing to give up, in terms of utils, to obtain one brand over another.¹ See Figure V.

A set of instructions was given verbally to each panelist after completion of part one. These were as follows: "Assume that you have been awarded a gift and can pick it up in a food store. There are eleven different stores, and in each store there are two display units as you see here. We would like for you to indicate on your cards, from which display you would take your gift in each of the eleven stores. Notice that in the first store, you could reach into display one and receive a turkey (KBT) each time. In display two in this same store, you would always receive this turkey (UBT). However, in store two you could reach into the first display and receive this turkey (KBT), nine times out of ten or 90% of

¹Utils are the units in which utility is measured. Utility denotes the importance or the usefulness which a person thinks he sees in a good or service.

the time, but you would receive \$2.00 one out of ten times. The two dollars is one-half of the market price of the turkeys. In display two in this same store, you would always receive this turkey (UBT). Notice now that in each store, every time that you would reach into display two, you will receive a turkey (UBT). However, in display one there is a chance of receiving the two dollars. The chances of receiving the two dollars increases until you reach store eleven in which you would always receive the two dollars and never receive the turkey (KBT).

After the panel session was completed, the test cards were separated according to the consistency of the replies. Those which were inconsistent or incomplete numbered twenty eight (17%) out of a total of 158 and were disregarded leaving a total of 130 usable cards.

The inconsistent replies were of two types. The first was composed of the cards on which an individual had indicated a preference for a particular sample during the ranking section, but then completely reversed her decision during the section involving chance. The second grouping consisted of those who had shown an illogical pattern during the test involving chance alone. As an example, an individual might have been willing to select from the first display in stores seven and nine but not in store eight. It was assumed that the inconsistent replies indicated that these consumers did not understand the directions.

Since this test was designed to measure utility, it was then necessary to assign a value in utils to the unknown brand turkey (the control) and to the money. Each util was given the monetary value of one cent. Therefore,

the UBT was worth 400 utils, which corresponded with its market value of four dollars. The two dollars was given a value of 200 utils. These two arbitrary definitions of the constants correspond to defining the two undefined constants, which is permissible, since, according to Edwards, cardinal utility is measured only up to a linear transformation.¹

This procedure amounts to placing appropriate restrictions on the pattern of an individual's preferences and expectations. To justify the reasoning under discussion, it was necessary to stipulate conditions strong enough to guarantee that any assignment of numbers to outcomes to measure their relative attraction is unique up to a linear transformation; e.g., unique except for origin and unit. Such an assignment is often called a utility function. Theories which assure the existence of a utility function, unique up to a linear transformation, and which provide in addition for some measure of probability, are often summarized by saying: an individual chooses between alternatives involving risk in such a way as to maximize expected utility.²

Once the utility of the UBT was assigned a value of 400 utils, and the utility of the \$2.00 was defined as 200 utils, the utility of the KBT was calculated by using the concept of expected utility as follows: Utility of UBT = (p) (Utility of KBT) + (1-p) (Utility of \$2.00).

¹Edwards, op. cit., p. 392.

²Donald Davidson and Patrick Suppes, Decision Making, An Experimental Approach (Stanford University Press, Stanford, California, 1957), p. 2.

Using this formula, the utility of the KBT was determined for each of the points at which a consumer would no longer take her prize from Display One but would switch to Display Two. This procedure yielded a calculation of the minimum utility that the consumer placed on the KBT as the estimate of his point of indifference is a conservative estimate. The first point to be measured was the midway at which the chances of receiving a KBT or money was equal. In this case, a consumer would take her prize from the first display in stores one through six, but would then switch to the second display in store seven. Each panelist's point of indifference is assumed to be at store six when his actual point of indifference may be anyplace between store six and seven. This assumption yields a conservative estimate of the panelist's point of indifference. There were eleven of the panel members whose utility for the KBT was determined as follows:

$$400 = (.5) (u \text{ of KBT}) + (.5) (200) \text{ therefore}$$

$$\text{Utility of KBT} = 400 / .5 - .5 / .5 (200) = 600$$

Thus, the cardinal utility of the KBT was determined to be 600 utils for these eleven consumers.

The equations for stores one through six are represented in Table XIX. The number of consumers in each category is given in Table XX.

It can be seen that thirty-one consumers stated that they would prefer the known brand to the unknown one, but would be unwilling to assume a risk to acquire this brand.

TABLE XIX. Calculation of the Minimum Utility Placed on the Known Brand Turkey (KBT)

Panelist Choosing Display I in:	Formula
Store 1 not Store 2	$400 = 1.0 + 1.0 (200)$ $U \text{ of KBT} = \frac{400}{1.0} - \frac{.0 (200)}{1.0} = 400$
Store 2 not Store 3	$400 = .9 + .1 (200)$ $U \text{ of KBT} = \frac{400}{.9} - \frac{.1 (200)}{.9} = 422$
Store 3 not Store 4	$400 = .8 + .2 (200)$ $U \text{ of KBT} = \frac{400}{.9} - \frac{.2 (200)}{.8} = 450$
Store 4 not Store 5	$400 = .7 + .3 (200)$ $U \text{ of KBT} = \frac{400}{.7} - \frac{.3 (200)}{.7} = 486$
Store 5 not Store 6	$400 = .6 + .4 (200)$ $U \text{ of KBT} = \frac{400}{.6} - \frac{.4 (200)}{.6} = 533$
Store 6 not Store 7	$400 = .5 + .5 (200)$ $U \text{ of KBT} = \frac{400}{.5} - \frac{.5 (200)}{.5} = 600$

The consumers who preferred Display Two over Display One in store one evidently felt they could obtain less utility from the KBT than from the UBT. These consumers preferred the KBT over the \$2.00, and therefore the utility of this turkey must be less than 400 utils but more than 200. There were forty-two panelists in this category. This information is included in Table XX along with a list of the number of consumers who demonstrated that the KBT was worth different amounts of utils to them.

The data in this table shows that nine individuals preferred the \$2.00 to the KBT. It is obvious that this turkey was worth less than 200 utils to these persons.

TABLE XX. Number of Panelists and Their Indirect Evaluation of the Value of the Well-Known Brand Turkey in Utils

Number of utils for well-known brand	Number of consumers giving this as maxi- mum utility of well- known brand turkey	Cumulative number of consumers at each utility evaluation
600	11	11
533	5	16
486	13	24
450	13	42
422	2	44
400	31	75
Unknown but more than UBT	4	79
Less than 400 but more than 200	42	121
Less than 200 for well- known brand but more than 200 for unknown brand	3	124
<u>Non-Potential Consumers</u>		
Less than 200 for both turkeys	6	130

At the extreme were four consumers who preferred the KBT to the \$2.00, and at the same time, preferred the \$2.00 to the UBT in all eleven stores. These four consumers were evidently highly brand conscious.

Most of the replies, however, were between these two extremes. There

were eleven consumers who indirectly gave their evaluation of the utility of the KBT as 600 utils or 200 utils more than UBT. Five consumers gave 533 as their evaluation of the utility of the KBT, while thirteen panelists indirectly said their evaluation was 486 and a similar number evaluated the utility at 450. Only two consumers gave an evaluation of 422 but 31 indirectly gave 400 as their utility.

From Table XX it can be seen that 44 out of 126 panelists were willing to give up something (a chance for a lesser valued prize) to receive the KBT rather than the UBT. It can therefore be assumed they were willing to pay extra to obtain this turkey. The amount they were willing to pay extra is their maximum expected utility evaluation minus 400 utils, which were originally assigned in the calculation formula. The 400 utils were originally assigned to the UBT, because this turkey weighed 10 pounds and the market price was 40¢/pound. The assignment of 200 utils to the \$2.00 was on a similar basis of 1 util for every cent of monetary value. A study by Edwards indicated that the utility of money is fairly linearly related to its dollar value over the range of money from -\$5.50 to + \$5.50.¹ Since this is the case, the number of utils obtained for the KBT by solving the equation could be transformed to cents on a 1 util to 1 cent basis.

The null hypothesis is "potential turkey consumers are not willing to pay more for a KBT than for a UBT." Since potential turkey consumers were

¹W. Edwards, "The Prediction of Decisions Among Bets," Journal of Experimental Psychology, Vol. 50 (1955), p. 213.

defined as those who would prefer a turkey over \$2.00, six of the 130 panel members did not meet this requirement. These were the members who stated they would prefer the \$2.00 to either turkey. All other panel members were included, thus giving a total sample of 124. Of these 124 persons, only 45 preferred the UBT to the KBT. Therefore, 79 panelists preferred the KBT to the UBT.

The test of the hypothesis was then conducted as follows: H_0 : There is no difference in potential turkey consumers' preferences toward the KBT and the UBT. H_1 : There is a difference in potential turkey consumers' preferences toward the KBT and the UBT.

The alpha level was set at .05 and a chi-square goodness-of-fit test was selected as the statistical method to test the hypothesis. The 95% confidence level for this statistic with 1 degree of freedom is 3.84, and therefore, the null hypothesis was rejected, since a value of 9.3 was obtained.

The second hypothesis tested was: H_0 : Three fourths of the potential turkey consumers will not pay more for the KBT than for the UBT.¹ H_1 : One fourth or more of the potential turkey consumers will pay more for the KBT than for the UBT.

It can be seen from Table XXI that 48 panelists indirectly indicated they would pay extra for a well-known brand. There were the eleven consumers

¹ Previous work at the Detroit Preference Panel had demonstrated that at least one-fourth of the consumers were willing to pay extra for a brand-name turkey through both a panel ranking test and questionnaire. This was backed up by results of a panel sale.

at 600, five at 533, thirteen at 486, thirteen at 450, and two at 422. In addition, there were four persons whose exact utility is unknown. It was, however, greater than that for the unknown turkey because they preferred the KBT over \$2.00 and \$2.00 over the UBT.

TABLE XXI. Amount Extra the Panel Members Indirectly Indicated They Would Pay for a Well-Known Brand Turkey

Amount Extra/bird	Cents Extra/lb.	No. of Total Sample Population of 124	Percent of Population
2.00	20.0	11	8.9
1.33	13.3	5	4.0
.86	8.6	13	10.48
.50	5.0	13	10.48
.22	2.2	2	1.61
Market price	none	31	25.00
Premium not known but is more than for unknown brand		4	3.22

The calculated chi square value of 12.4 was significant at the 95% level, hence the null hypothesis was rejected.

It is therefore evident that at least one fourth of the potential turkey consumers were willing to pay extra for the KBT.

A regression analysis was also performed on the data in order to ascertain the relationship between the number of utils or cents the consumers were willing to give up to get the KBT, and the number of consumers willing to give up that number of utils or cents. It was assumed that the independent

variable, X , was the number of utils or cents obtained in the equations. The dependent variable, Y , was assumed to be the number of people who gave that number of utils as their evaluation of KBT. It was hypothesized that fewer panelists would give an extra 200 utils for the KBT, and more individuals would be willing to give an extra 22 utils for the KBT. Thus, the regression model hypothesized was:

$Y_i = A + B (X_i - \bar{X}) + E_i$. Where Y_i is the number of consumers who would give additional utils for the KBT, and the X_i is the evaluation of utils for the KBT minus the 400 utils of the UBT. The data was taken from Table XXI and was plotted in Figure VI. The regression analysis calculations yielded the following estimates: $\hat{Y} = 36.17 - 2.84 (x - \bar{X})$ and $r = -.912$, $r^2 = .83$, and $S_{Y.X}^2 = 115.1$.

Thus, 83 percent of the variance of Y is explained by changes in the independent variable X . This estimated regression line is an approximate demand curve for the KBT above the UBT for it indicates that a larger number of consumers would buy the KBT if the cost of doing so were nearer to the UBT.

Advantage of VonNeumann-Morgenstern Method

The VonNeumann-Morgenstern Method allows one to determine whether panelists will pay a premium for one product over another, and, if so, the number of panelists who will pay various premiums in order to obtain their preferred product.

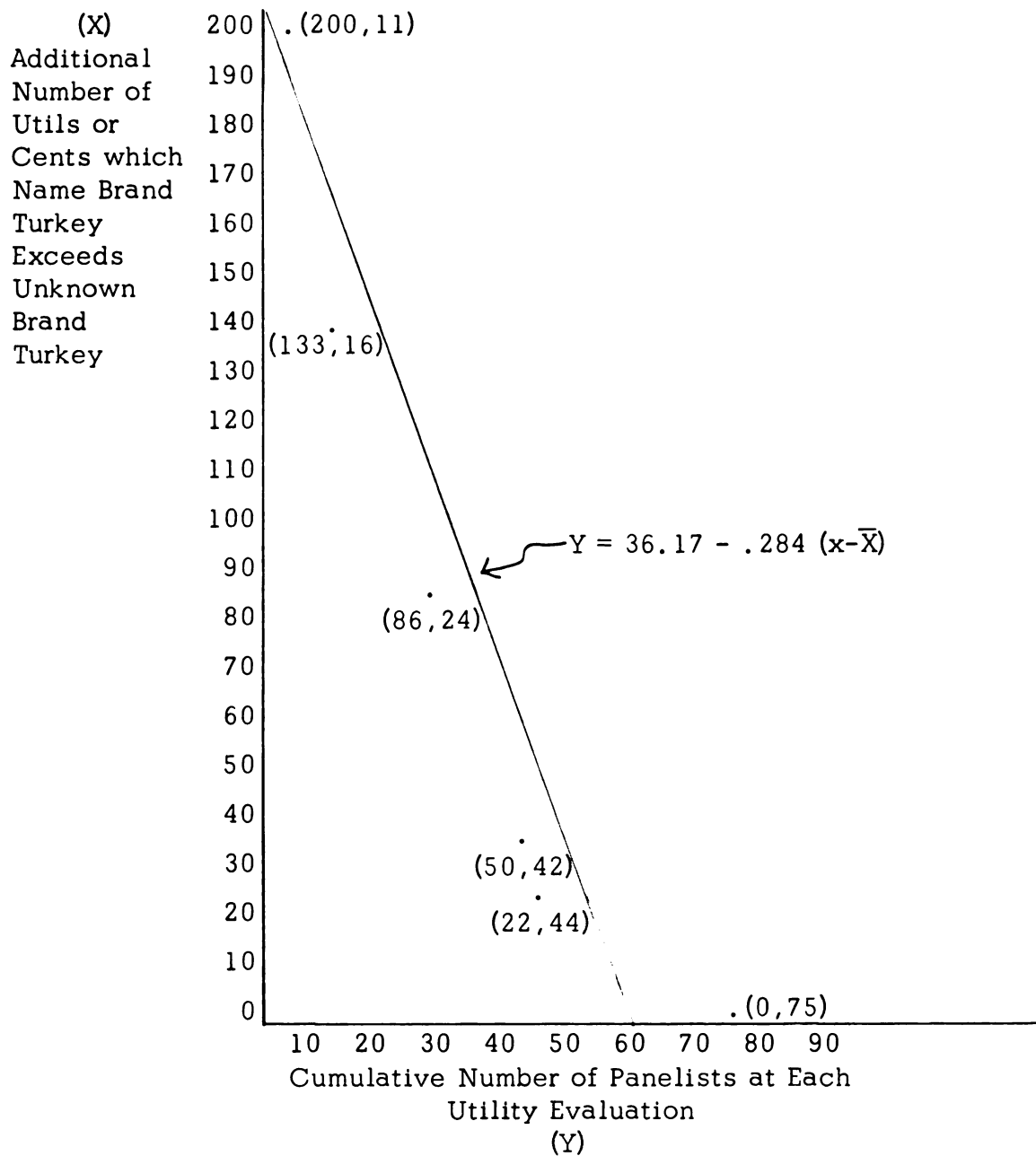


FIGURE VI. Approximate Demand Curve for a Known Brand Turkey in Comparison with an Unknown Brand Turkey

Disadvantage of VonNeumann-Morgenstern Method

Several limitations are inherent in a study of this nature, and should be clearly understood before using this type of test.

The test involves a form of gambling, and may, therefore, by itself offer some degree of positive or negative utility. An individual with negative social or religious convictions concerning gambling might be unwilling to participate in this test, or might consistently select the choice involving no gambling. On the other hand, a person who enjoys games of chance might show strong brand influences, when in actuality this is not the case. During this study, it was assumed that gambling yields zero utility.

It was also assumed that the probabilities by which the utilities were multiplied were objective ones. However, the consumer's estimate of the importance of a particular probability might not have been the same as the numerical value assigned to this probability by the researchers. A consumer might also have a preference for one probability over another. Therefore, in the cases of a probability preference, it would be impossible to measure utility of the items on display.

This research method does not permit a test of the degree of preference for both products. Instead, one product must be used as a constant or control.

The procedure is difficult for consumers to understand, and therefore requires the use of trained personnel who must repeat the instructions to small groups of participants. Apparently, a few consumers failed to

understand even after repeated instructions, and therefore marked their cards in a haphazard fashion or not at all.

It may be difficult to arrive at the quantity of a particular good and the unit of time over which it will be purchased for many products. This was less of a problem with turkeys due to the heavy seasonal demand for this product and a knowledge of the panel members prior purchases.

Use of the Demand-Price Concept in Measuring the Relative Utilities of Two Products

Test Procedure Illustrated on an Example

A modification of the VonNeumann-Morgenstern Method was also used to determine if consumers will pay a premium for one brand of turkey.¹ A panel similar in nature to the M.S.U.-W.S.U. panel was utilized in conjunction with this test.

During the panel, four trays were filled with chipped ice, and a turkey was placed in each. These turkeys were identical in weight and appearance, but were packaged in differently branded bags. Two of the bags bore a known brand and the other two were packaged in unknown brand bags. These trays were then identified as Series I through IV. This test did not involve a display with money. Instead, the panel members were directly asked whether they would prefer the turkey in front of them or various denominations of money. Again, the panel members were told that all the

¹James C. Makens, a fellow graduate student, cooperated with the writer in the planning, execution, and evaluation of this study. This study, therefore, represents the joint efforts of Mr. Makens and the author.

turkeys on display weighed the same and that turkeys of this weight were presently selling for \$4.00.

Each participant was given four cards, one for each turkey. They were instructed to mark on each card whether they would rather have the corresponding turkey or a certain sum of money. There were 13 different amounts of money which ranged from two dollars to six dollars. The panel members were told to mark each of the 13 alternatives. (See Figure VII.) A check for money at any level indicated that the panelist would prefer that amount of money to the turkey. A mark for the turkey indicated that the panelist preferred the turkey over the amount of money listed for the same store, and indicated he would purchase the turkey if it were available for that amount of money. The demand-price may then be determined by finding the largest amount of money the panelist would forego in favor of the product. Demand-price is simply the highest price that consumers will pay to obtain a product.

Two of the cards increased in amount of money from the top of the card to the bottom while the choices listed on the other two cards decreased in value. This was done to avoid bias due to procedure. One of each of these types of cards was used for the unknown and the known brand turkeys.

A total of 97 consumers were given cards for this test. Upon analyzing the results, it was found that 20 of these persons were evidently confused regarding what they were to do. In addition, one member returned only one-half of her cards, thus giving a total of 77.5 consumers. Several left their

cards blank, or checked only one alternative. The results from these individuals were therefore discarded.

MICHIGAN STATE UNIVERSITY

Consumer's Opinion of Product Quality

Series IV Turkeys

Check your preference for money or turkey in each store.

Store 1 \$6.00 - or turkey -	Store 8 \$3.75 - or turkey -
Store 2 \$5.50 - or turkey -	Store 9 \$3.50 - or turkey -
Store 3 \$5.00 - or turkey -	Store 10 \$3.25 - or turkey -
Store 4 \$4.75 - or turkey -	Store 11 \$3.00 - or turkey -
Store 5 \$4.50 - or turkey -	Store 12 \$2.50 - or turkey -
Store 6 \$4.25 - or turkey -	Store 13 \$2.00 - or turkey -
Store 7 \$4.00 - or turkey -	

Name _____ Comments _____

FIGURE VII. Questionnaire Presented to Panelists in a Demand-Price Study

Each of the remaining cards was then analyzed and the last point at which a consumer was willing to take a turkey instead of money was recorded in terms of the amount of money the individual was willing to forego. As an example, if \$4.25 was the highest point at which the alternative was money or a KBT, and this turkey was checked, then the sum of \$4.25 was recorded for it. This is the demand-price for the KBT for this panelist. Since the participants knew that turkeys of this weight were worth \$4.00 on the market, it was apparent that the consumer was willing to forego an additional 25¢ to obtain this turkey. If this same panelist

preferred the UBT to \$3.75, but not to \$4.00, then \$3.75 was recorded for the UBT. The demand-price for the UBT is \$3.75 for this panelist. Once these figures were listed, they were statistically tested. This involved the use of the Wilcoxon Matched-Pairs Signed Ranks Test. The Wilcoxon Test utilized information about the direction and the magnitude of the differences within the pairs of observation of the Known and Unknown brand turkeys for the consumer. Thus, a + \$.50 ($\$4.25 - \$3.75$) would be recorded for this panelist. If preferences for both the KBT and UBT turkeys are equivalent, one would expect to find some of the larger differences favoring the KBT turkey and some favoring UBT. However, if the sum of the ranks of the differences which have a positive sign (indicating KBT is preferred over the UBT by that amount of money), is much different from the sum of the negative ranks, one would infer that the preferences between the two brands of turkeys is different. This test gave a statistical Z value of 3.8, which indicated that the results were significant at an alpha level of .01.

Therefore, the null hypothesis, that there is no difference in potential turkey consumers' preferences toward the KBT and the UBT, was rejected.

The next step in the analysis of these data was to test the null hypothesis, that at least 75% of the potential turkey consumers are unwilling to pay extra for a KBT. Again a chi square goodness-of-fit test was used, and again the null hypothesis was rejected, thus indicating that 25% of the consumers were willing to pay extra for KBT. Since each panel member evaluated two pairs of known and unknown brands, it was necessary to

multiply the numbers of participants whose cards were retained (77.5) by the number two. This then gave a figure of 155 as the total sample population. This figure was employed in the chi square analysis.

It can be seen in Table XXII that slightly more than 37% of the panelists indicated they were willing to pay at least 25¢ per bird extra for the KBT. This amounts to a premium of 2.5 cents per pound. Twenty-five percent of this population indicated they would pay a premium of five cents a pound (50¢ per bird). Seventeen percent were willing to pay 7.5 cents per pound extra, thirteen percent indicated they would pay 10 cents, seven percent were willing to pay 15¢, 3.8 percent would pay 17.5¢, 2.6 percent would pay 20¢, and less than 1 percent would pay 35¢ extra.

The results also indicate that an additional 14.8 percent or 23 persons indicated they were willing to pay extra for the unknown brand turkey. This provides evidence that more than 50 percent of the potential consumers of turkey were willing to pay at least 2.5¢ per pound premium for a turkey they considered to be of superior quality.

A regression analysis was also performed upon the data for the known and the unknown brand turkeys. The analysis was conducted to determine the relationship between the price per turkey and the number of panelists willing to take the turkey at that price. These data are presented in Appendix F.

The independent variable, X , was assumed to be the price per turkey. The dependent variable, Y , was assumed to be the number of panelists

who would forego the different amounts of money to receive a particular turkey.

TABLE XXII. Amount of Money Panel Members Indicated They Would Forfeit to Receive a KBT Turkey

Amount	Number of Replies	Cumulative	% of Total Potential Turkey Consumers	Cumulative %
0*	74	155	47.74	85.14*
.25	19	58	12.25	37.40
.50	12	39	7.74	25.15
.75	6	27	3.87	17.41
1.00	10	21	6.45	13.54
1.50	5	11	3.23	7.09
1.75	2	6	1.29	3.86
2.00	3	4	1.93	2.57
3.50	1	1	.64	.64

The regression analysis on the data concerning the panelists preference for an unknown turkey gave the following estimates $\hat{Y} = 85.83 - .541(x - \bar{X})$ and $r = -.931$, $r^2 = .867$, and $S_{y \cdot x}^2 = 521.4$.

The r^2 value indicates that approximately 86 percent of the variance in the results of the number of panelists willing to take the UBT is explained by linear changes in the independent variable, which is the price per turkey.

The b value of $-.541$ indicates that if the price is increased by 10 cents per turkey, the number of panelists willing to take the UBT will decrease by 5.41.

The estimated regression line was then drawn as shown in Figure VIII. This estimated regression line is an approximate demand curve for the UBT. It shows the number of consumers who will take the unknown brand turkey at each price per turkey.

A similar regression analysis was performed on the data for the KBT turkey, which gave the following estimates: $\hat{Y} = 83.08 - .527(x - \bar{X})$ and $r = -.921$, $r^2 = .85$ and $S^2_{y.x} = 691.18$. In this case, 85 percent of the variance in the number of panelists willing to take the KBT is explained by changes in the price per turkey. The b value indicates that a ten cent increase in the price per turkey (1 cent per lb.) will decrease the number of consumers willing to take the KBT turkey by 5.27. This is equivalent to 3.4 percent fewer consumers who would take the KBT turkey at this price. This was determined by dividing 5.27 by 155 (the number of panelists).

An estimated regression line for the KBT was also drawn on Figure VIII. This line is an approximate demand curve for the KBT, since it indicates the number of consumers who were willing to take this bird at each price.

It should be noted that the slopes of the two approximate demand curves are nearly identical ($-.541$ for the unknown brand and $-.527$ for the KBT). In other words, the approximate demand curves are nearly parallel. The differences between the two curves lies in the fact that the approximate demand curve for the KBT lies above the demand curve for the UBT at all points. The amount of this difference is due to the preference

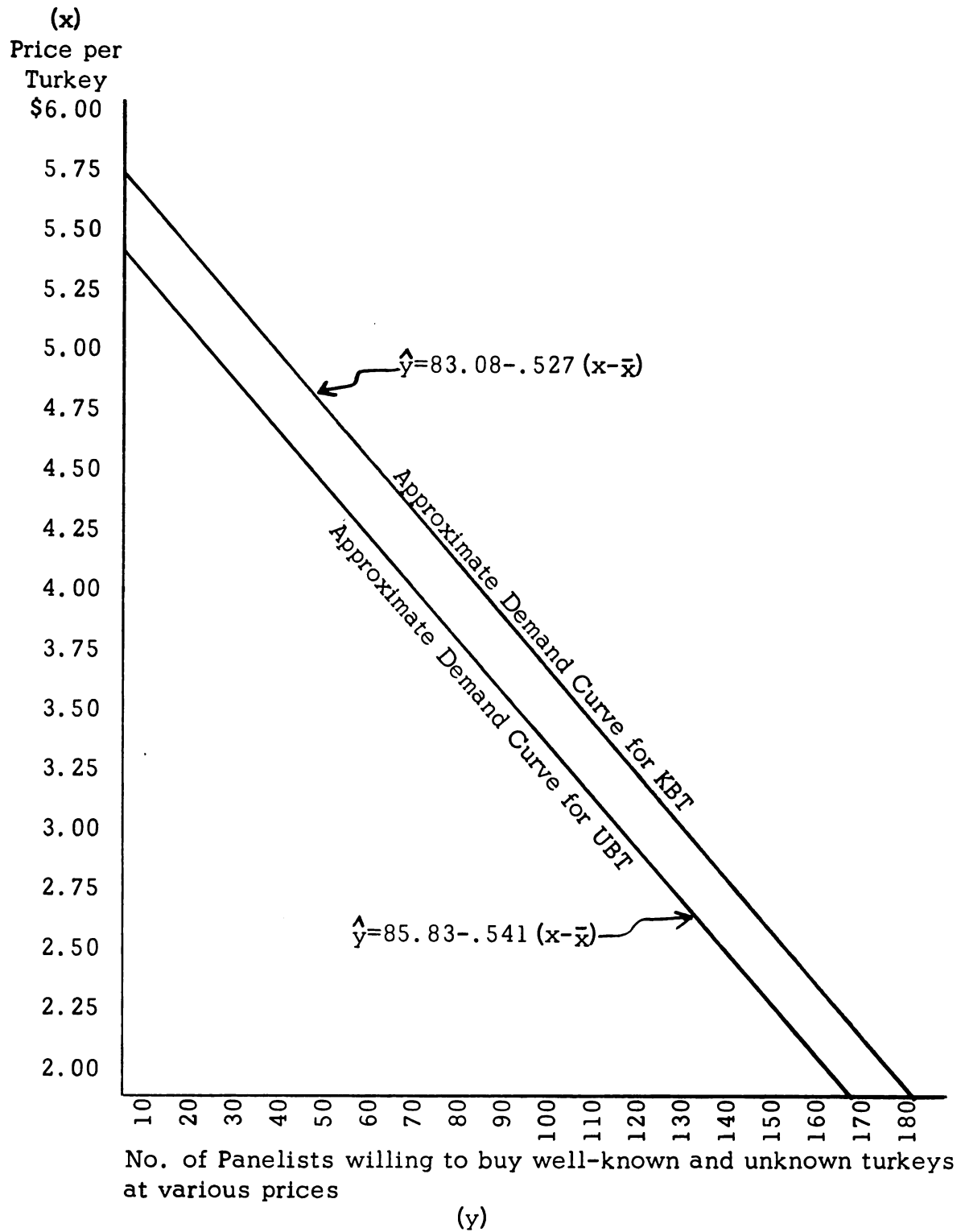


Figure VIII. Approximate Demand Curves for Well-Known and Unknown Brand Turkeys as Derived from Regression Analysis

for the KBT turkey over the UBT. This can be assumed since the only difference between the two brands was the package and not the bird. Therefore, the difference must be due to a preference for the KBT over the UBT.

The approximate demand curve is a relative concept as the dependent variable is the number of panelists willing to forego the various amounts of money in order to receive that particular product. It does not indicate the maximum quantity that will be consumed per unit of time at the various prices. One may obtain an estimate of the quantity that could be sold during a specified time period by weighting each panelist's rankings by the number of turkeys the panelist believes he will consume during this period.

Advantage of Demand-Price Method

The Demand-Price Method of measuring relative utilities of two products has the advantage that the questionnaires are easier to explain to the panelists than is the case under the VonNeumann-Morgenstern Method. This means that the panelists are likely to understand more clearly what they are supposed to do when presented with the questionnaire under the new method. This method allows one to study preferences at as many as 15 price levels during one panel session.

Another advantage of this method is that it avoids gambling entirely. That is, the assumption that gambling yields zero utility is not needed in this method of measuring relative utilities.

Disadvantage of Demand-Price Concept

The primary disadvantage of this new method is that sometimes it does not yield any indication of which product the panelist will choose if the same price is attached to both products. In other words, this method does not give such information when the panelist gives both products the same evaluation in monetary terms.

Series of Rankings Involving Different Prices on Identical Products

Test Procedure Illustrated on an Example

Another method of determining whether the panelist will pay more for his preferred product is to utilize a series of rankings which involve the placing of different prices on identical products. In this method, one replication of each product is included in each series of samples presented to the panel. However, a different price is attached to identical products in the different series. There are as many series as there are different price level combinations on identical samples.

An example of the manner in which this method is used is a study of consumers preferences for varying sizes of eggs. This experiment was conducted in conjunction with the M. S. U. -W. S. U preference panel. Five sizes of eggs were presented to the panel in two series. In the first series, the panelists rated the five different samples on the basis of their preferences for the samples without consideration of the price element.

In other words, the five samples presented in series one had no price attached to them, and the panelists ranked the samples as though they were all available at the same price.

Series two, which was conducted at a different location in the same room, consisted of the same five sizes of eggs which were presented to the panelists in series one. However, in series two, prices were attached to each size of egg. The Jumbo eggs were priced at 75¢ per dozen, Extra Large eggs at 69¢ per dozen, Large eggs at 65¢ per dozen, Medium eggs at 50¢ per dozen, and Small eggs at 40¢ per dozen.

The null hypothesis tested in series one was that consumers preferences toward the five different sizes of eggs are equal. The alternative hypothesis tested was that preferences toward the five different sizes of eggs differ. A chi square test based on the coefficient of concordance indicated that this null hypothesis should be rejected at the alpha .05 level.

A null hypothesis tested in series two was that consumers do not prefer largest sizes of eggs (Jumbo, and Extremely Large), to smallest egg sizes (Medium and Small), when the largest sized eggs are priced 29 to 35¢ higher per dozen than the smallest eggs. The alternative hypothesis is that consumers do prefer largest sizes of eggs to smallest sizes of eggs under these conditions.

The chi square test for individual comparisons may be used to test this hypothesis if the responses are combined into largest and smallest egg categories as illustrated in Tables XXIII and XXIV. The X^2 calculation

was significant at the alpha .05 level. Several other hypotheses may be tested in a study of this type. The nature of the problem determines which hypothesis should be tested.

TABLE XXIII. Number of Panelists Ranking Different Sizes of Eggs as First, Second, Third, Fourth and Fifth Choice

	Ranking				
	1	2	3	4	5
Jumbo					
Series I no price attached	31	7	7	1	5
Series II 75¢/dozen	17	10	17	11	13
Extra Large					
Series I no price attached	19	23	6	0	2
Series II 69¢/dozen	25	30	7	6	0
Large					
Series I no price attached	4	15	25	5	1
Series II 59¢/dozen	17	19	28	3	1
Medium					
Series I no price attached	2	4	6	26	12
Series II 50¢/dozen	3	7	13	26	19
Small					
Series I no price attached	2	3	5	15	25
Series II 40¢/dozen	7	2	4	21	34

TABLE XXIV. Number of Panelists Ranking Largest (Jumbo and Extra Large) Sizes of Eggs and Smallest (Medium and Small) Sizes of Eggs as First, Second, Third, Fourth, and Fifth Choice When the Largest Sized Eggs are Priced 29 to 35¢ Higher per Dozen than Smallest Eggs

	Ranking				
Egg Size	1	2	3	4	5
Largest	42	40	24	17	13
Smallest	10	9	17	47	53

If more than two series of rankings are obtained with each series containing different price combinations on identical products, one may draw an approximate demand curve for each size of egg. The procedure is similar to the method presented in the previous section of utility.

A regression analysis will determine the relationship between the price of a dozen eggs and the number of panelists willing to take the eggs at that price. In this example, five separate regression computations will yield the approximate demand curves for the five sizes of eggs.

In each case, the independent variable is assumed to be the price per dozen eggs. The dependent variable is assumed to be the number of panelists who would prefer the particular egg size at each price. The price can be plotted on the vertical axis, and the number of panelists preferring the particular size of egg at the various prices can be plotted on the horizontal axis. The five estimated regression lines can then be drawn on this graph as approximate demand curves. Each approximate demand curve will show the number of panelists willing to take that particular size of egg at each price recorded on the vertical axis.

Advantage of Series of Rankings

The method of using a series of rankings, each with different prices on identical products, avoids gambling entirely. It is an easy method to explain to panelists, for the panel members are familiar with ranking procedures, and that is all that is required of them.

Disadvantage of Series of Rankings

The disadvantage of using a series of ranking tests is that preference at only a few price variations may be investigated during one panel session. One series is required for each different price level, so the fatigue effect limits the number of price levels that can be studied.

CHAPTER IX

DISCUSSION

Optimum Degree of Testing

The chief task of the market researcher is to make a prediction of future outcomes in terms of either net profit or relative profit for each product being evaluated. This process involves obtaining an estimate of a probability distribution of probable net profits for each product. This profit probability distribution may be visualized as a graph with various probabilities on the vertical axis and various levels of profit on the horizontal axis. In this process, the market researcher is very interested in determining the variance and mean of the distribution.

The most information the market researcher can hope to obtain about each product's profit probability distribution is still likely to be incomplete and sometimes unreliable. However, if the firm is to obtain and use the highest possible level of information available, then it must pay the substantial costs which are associated with the gathering of this information.

Just where the margin of profitable estimation lies may continue to be a matter of guesswork, for it is nearly impossible to tell accurately whether more sophisticated testing will yield enough improvement in income to pay for itself. However, a choice does exist between less reliable estimates at lower costs and more reliable estimates at higher costs. A decision must be reached in each project, and it would be absurd to say that "rationality"

in business calls for the most accurate estimates available regardless of their cost.

Thus, the problems of which test to choose and how many panelists must be interviewed are largely a matter of economics. If marketing management is of the "maximin" type, it may wish to narrow the profit probability distribution by reducing the variance.¹ For example, marketing management may say that it is worth \$10,000 to reduce the variance of a profit probability distribution for a product by a certain amount. The \$10,000 figure may have been arrived at by maximin followers who found that a reduction in the variance by this amount increases the maximin solution by at least \$10,000.

This type of thinking can lead to a demand curve for consumer preference tests. In other words, the demand curve for tests of varying degrees of reliability is derived from marketing managements' evaluation of what it is worth, in dollar terms, to reduce the variance of the profit probability distribution. The degree of reliability obtained from testing is affected by at least the following factors: (1) type of panel, (2) type of design and analysis, (3) number of panelists, and (4) degree of bias in sample of panelists.

The usual cost curves can be derived for consumer preference tests which reduce the variance of the profit probability distribution to various

¹A maximin management selects the product which, if the worst of all possible outcomes are compared, is the best of the worst outcomes.

amounts. The marginal cost curve would probably begin at a relatively high level as the consumer preference testing begins. This higher cost would most likely be due to the time it takes for marketing management to familiarize the market researcher with the problem which is to be investigated. Then the shape of the marginal cost curve is likely to: (1) decline, (2) rise at a slow rate, and (3) finally, rise at a more rapid rate.

Figure IX shows the diagram of the possible curves derived from this type of analysis. The quantity axis is $1/(\text{variance of the profit probability distribution})$ as marketing management prefers the smallest variance it is possible to obtain. Thus, they may be satisfied with tests yielding a large variance, or they may not be satisfied unless tests yield a small variance. Their preferences for variances of different sizes is derived from the assurance of profits or avoidance of losses which these variances indicate. Marketing management will then choose the degree of testing at which marginal cost equals marginal revenue. For example, the minimum variance may be obtained by using the demand-price approach in a personal interview with a large number of consumers who were selected by strict random sampling techniques. This procedure may be indicated at the extreme right end of the quantity axis. However, the largest variance may arise from a simple one-sample test involving a few of the firms' employees. This technique may be the degree of testing indicated at the extreme left end of the quantity axis.

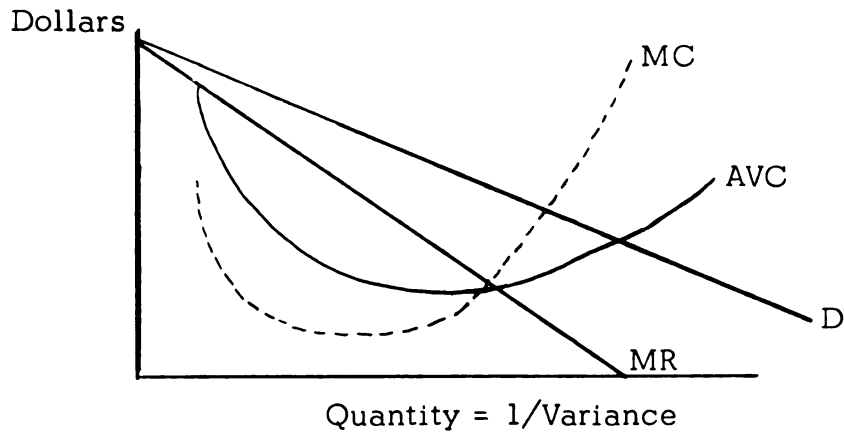


FIGURE IX. Demand and Cost Curves for Different Degrees of Testing

The costs of the various testing methods may be determined on a relative basis. If one is interested in reducing the variance of the profit probability distribution, he may consider the most efficient test to be the one yielding the minimum variance. If all procedures yield an unbiased estimate, then the test which the market researcher will want to choose will be the test which yields the smallest variance per dollar of research expenditure.

Testing Sequence

The exact sequence in which the tests presented in Chapters IV to VIII should be used varies with the type of study being conducted. Usually, a research project originates with a request to find the most preferred product among many alternative products. If this is the case, the researcher is forced to use a balanced lattice or factorial test to obtain a general idea

of which products are most preferred, and which products may be eliminated in further testing. It may be recalled that individual comparisons can be obtained from these tests. If the panel data did meet the assumptions of these parametric tests, then no further testing would be necessary. However, the meaningfulness of the results of a parametric test depends on the validity of at least the following assumptions.¹

1. The observations must be independent.
2. These populations must have the same variance.
3. The observations must be drawn from normally distributed populations.
4. The variables involved must have been measured in at least an interval scale, so that it is possible to use the operations of arithmetic on the scores.

Data obtained from preference tests do not meet these assumptions. Because these assumptions are not realistic in the preference testing, one must place a lower reliance on parametric tests. Therefore, it seems reasonable to follow the procedure of using the parametric tests as a means of reducing the number of products to be tested to four or less. Once the number of products to be tested is less than five, nonparametric tests can be used. All of the tests presented in the two, three, and four sample cases are nonparametric tests. A nonparametric test is also based upon certain underlying assumptions, e.g. that the observations are independent

¹Siegel, op. cit., p. 19.

and that the

these assum

metric test

so strong a

tests apply

Eventu

two. If it

to the othe

if the con

product.

Price Met

with differ

This

Exac

follow th

jectives

If,

found to

the mar

the mar

done b

1

2

and that the variable under study has underlying continuity.¹ However, these assumptions are fewer and weaker than those associated with parametric tests. "Moreover, nonparametric tests do not require measurement so strong as that required for the parametric tests, as most nonparametric tests apply to data in an ordinal scale."²

Eventually, the number of products being tested should be reduced to two. If it has been determined that consumers prefer one of the products to the other, then it may be within the objectives of the test to determine if the consumer will pay more for his preferred product than for the other product. In this case, the VonNeumann-Morgenstern Model, the Demand Price Method of Utility Measurement, or a series of rankings of products with different attached prices may be used.

This sequence of testing is illustrated in Figure X.

Exactly where a particular study will enter and how closely it will follow this testing sequence will vary with the situation, budget, and objectives of the project.

If, after the various phases of consumer testing, a certain product is found to be the most acceptable of the entire group available for testing, the market researcher must establish whether that product has a share of the market which makes it an attractive account proposition. This can be done by test-marketing which is conducted over a period of time long enough

¹Ibid., p. 31.

²Ibid.

for novelty

ket, follow

and market

During the

tising show

wide basis

Five or more
(balanced)

Three or more
(coefficient
chi square)

Two sample
comparisons

Tests to
more for
Neuman
of rank
price m

Fi
of

Sta

importa

are sta

Th

for novelty effects to wear off. The next stage is usually a regional market, followed by first a limited and then a full-scale nationwide promotion and marketing program if the company has that distribution objective. During the test-marketing, the intensity and type of promotion and advertising should usually be typical of what will later be practiced on a nationwide basis.

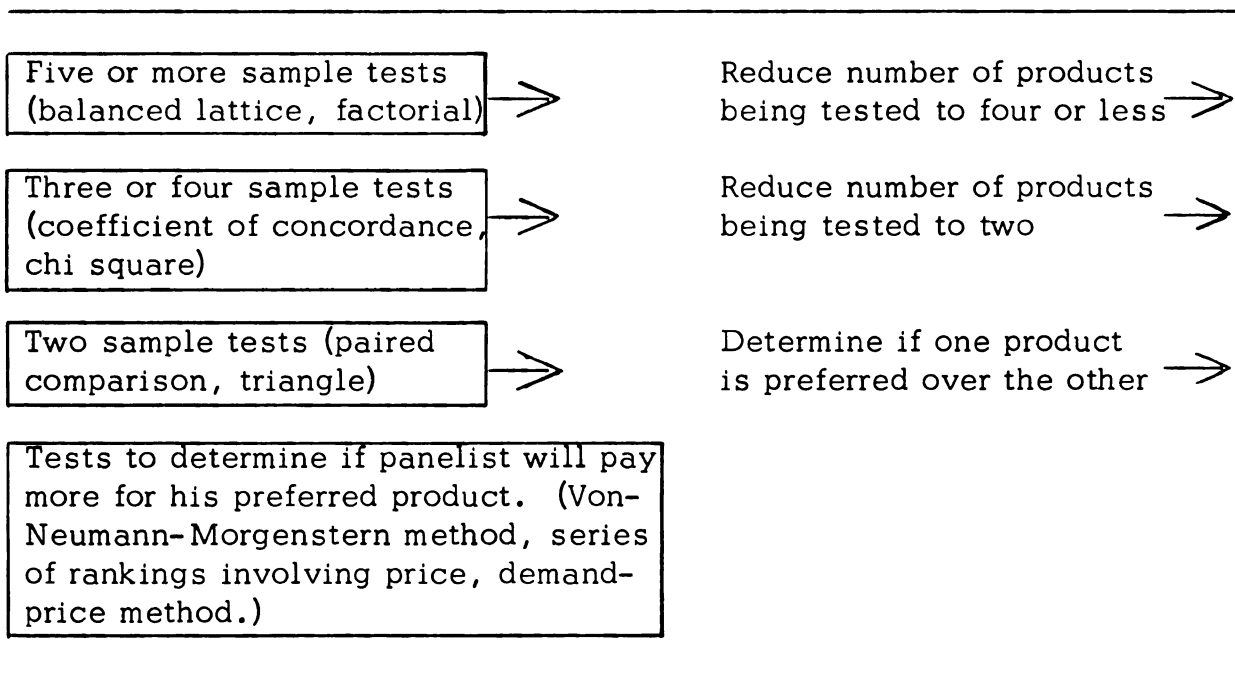


FIGURE X. Testing Sequence in Consumer Preference Testing of Several Products

Interpretation of Results

Statistical significance may not always parallel the relative economic importance of differences in preferences. In fact, some differences which are statistically significant may have absolutely no economic importance.

The market researcher should always interpret the results obtained from

consumer

that produ

in a prefe

larger pro

always co

are alreac

which 60

dies were

40 percen

marketed

Anot

that the c

this case

large a p

amount (

uct A. T

this perc

warrant

If o

(B), the

preferre

consumer preference tests in view of the current market situation. The fact that product A has been found to be significantly preferred over product B in a preference test does not necessarily mean that the firm will make a larger profit producing and selling product A. The market researcher should always consider how many other products similar to the preferred product are already on the market. For example, one may not produce candy A, which 60 percent of the panelists preferred, if seven or eight similar candies were already on the market. Instead he may produce candy B, which 40 percent of the panelists preferred, because no similar candies are being marketed.

Another reason that product B may yield more profit than product A is that the cost of producing product A may be higher than for product B. In this case, the tests presented in Chapter VIII may be used to indicate how large a percentage of the consumers would be willing to pay an additional amount (equal to the added cost of producing A instead of B) to obtain product A. The firm would be farther ahead to produce and sell product A to this percentage of consumers, provided it is a large enough segment to warrant economical production and distribution.

If one product (A) is cheaper to produce and sell than another product (B), then the firm should produce product A provided that A and B are equally preferred, or A is preferred to B.

The v
been ques
validity o
question
what they
actually p
"Is the pr
the same
The latter
most rele
the propo
product.
to evaluat
been con
preferenc
agreemen
tests to
lidity ob

1. W
Preferen
the Mich
1960), p

Validity of Panel Results
in Predicting Sales Results

The validity of predicting sales results from panel results has often been questioned. Actually two questions are involved in evaluating the validity obtained in predicting sales from consumer rankings. The first question is, "How many of the panel members are inconsistent between what they say they will buy, as reflected by rankings, and what they will actually purchase?" The second question which requires answering is, "Is the proportion of panelists ranking each product as their first choice the same as the proportion of panelists actually purchasing that product?" The latter question, which is essentially of the aggregative type, is the most relevant from the point of view of a researcher attempting to predict the proportion of the statistical population who will purchase a particular product. These are the crucial questions which must be answered if one is to evaluate the validity of panel results. Few commercial sales tests have been conducted to verify panel results. Mills, in his studies on consumer preferences for graded turkeys, obtained sales test results which were in agreement with previously obtained preference panel findings.¹ Sales tests to panel participants provide some additional insights into the validity obtained in predicting sales from consumer rankings.

¹W. C. Mills, Jr., L. E. Dawson, and H. E. Larzelere, "Consumer Preferences for Turkeys with Various Grade Labels," Quarterly Bulletin of the Michigan Agricultural Experiment Station, Vol. 43, No. 2 (November 1960), p. 258.

Three

M.S.U-W

one on eg

turkey bra

An an

that 21 of

turkeys w

the basis

chased eg

ber on the

purchase

the exped

8.3.

Thes

consiste

first cho

To a

tion of p

the prop

two pop

be equal

are not

1 M

Three sales tests were conducted at the conclusion of the regular M.S.U.-W.S.U. panel sessions. Two of the tests were performed on eggs; one on egg size and one on egg grade. The other test was performed on turkey brands.¹

An analysis of these studies with regard to the first question shows that 21 of the 39 panelists who purchased either known or unknown brand turkeys were consistent. The expected number of consistent replies on the basis of chance alone is 19.5. Twelve of the 42 panelists who purchased eggs among five sizes of eggs were consistent. The expected number on the basis of chance alone is 8.4. Fourteen of the 25 panelists who purchase eggs among the three egg grades were consistent. In this study the expected number of consistent replies on the basis of chance alone is 8.3.

These studies indicate that most individual panelists were not very consistent between what they said they would buy, as reflected by their first choice ranking, and what they actually purchased.

To answer the second question, one may hypothesize that the proportion of panelists ranking each product as their first choice is the same as the proportion of panelists purchasing that product. In other words, the two populations (first choice ranking and purchasing) are hypothesized to be equal. The alternative hypothesis tested is that the two populations are not the same.

¹Makens, op. cit.

Th

Size of

No. of

No. of

Rankin

Th

ulation

not sig

not re

Th

Egg Gr

Numbe

Numbe

Th

square

Th

Number

Number

Place

Th

between

Al

that th

is nea

The following results were obtained from the study involving egg size.

Size of Egg	Jumbo	Large	Medium	Small	Large
No. of Purchases	12	11	9	5	5
No. of First Choice Rankings	15	12	3	6	6

The chi square test was used to test the hypothesis that the two populations are the same. The calculated chi square value is 2.84 which is not significant at the alpha .10 level. Therefore, the null hypothesis was not rejected. The $X^2_{.90}$ with 4 degrees of freedom is 7.78.

The results obtained from the study on egg grades were:

Egg Grade	A	B	C
Number of Purchases	14	4	7
Number of First Choices	11	9	5

The same hypothesis was not rejected because the calculated chi square value was 1.74 and the $X^2_{.90}$, 2 d.f. is 4.60.

The results obtained from the study on turkey branding were as follows:

	Known Brand Turkey	Unknown Brand Turkey
Number of Purchases	26	12
Number of First Place Rankings	26	12

The calculated chi square value of zero indicates perfect agreement between the two populations.

Although these studies do not offer conclusive proof, they do suggest that the proportion of panelists ranking each product as their first choice is nearly the same as the proportion of panelists purchasing that product.

Thus
between
ranking

It is
can be r
several
that a m
tions are
manner i
phases o

This
is instr
tirely n
from pro

If
time, t
This m
consid
they c
or rank

Thus, it appears that most of the individual panelists are inconsistent between rankings and purchases, but they are consistent as a group in ranking and purchasing of the same products, under the same conditions.

Conclusions

It is obvious from this study that no one marketing research procedure can be recommended for use in all testing situations. However, there are several general problem classifications which will include most of the tests that a market researcher may be involved in. That is, certain considerations are important in any study included in a general classification. The manner in which several typical problems may progress through various phases of preference testing will now be discussed.

New Product Development Problems

This type of problem is defined as one in which the research worker is instructed by management to attempt to produce a product which is entirely new in conception, or essentially different in some characteristics from products already on the market.

If the research worker produces his experimental variants one at a time, these can be evaluated by both trained and untrained panelists. This may be done by obtaining a single evaluation on each characteristic considered important. If the variants are produced several at a time, then they can be evaluated on each important characteristic by hedonic scales or ranking methods, including incomplete block designs. At this early

stage of development, the trained and untrained laboratory panels can be used to indicate acceptability. The trained and untrained laboratory panels should be regarded as sieves for screening out all but the most promising products. The latter are submitted to the several stages of consumer testing. The importance of test marketing and of allowing novelty effects to wear off should be pointed out. Because the Type I and Type II errors should be small, and because the untrained laboratory panel is being used as an indicator of consumer opinion, the panel size should be large.¹ The cost of making both types of errors is quite large. If a good product is incorrectly rejected by the panel, then the company may lose a good source of profit. If a less desirable product is unnecessarily consumer tested, then the cost may also be quite large. Similar considerations apply at the market test stage.

It should also be pointed out that the product's marketing environment and the company's marketing policy influence the selection of research designs and product test techniques. A new product, in a class by itself, may demand one form of research, but another new product, competitive with products now on the market, may require another research approach.

Evaluation of Product Popularity Against Competition Problems

It is usual to have the trained laboratory panel evaluate the products

¹The Type I error is the error made by rejecting the null hypothesis when in fact the null hypothesis is true. The Type II error is the error made by accepting the null hypothesis when in fact the alternative hypothesis is true.

involv

then c

consu

found

answe

ber of

brand

answe

very u

answe

etc. m

Th

elemen

search

another

Instead

market

each b

must b

Th

sumer

involved for technical differences. If no significant differences are found then consumer testing of the product is not generally necessary. Either a consumer taste test or a use test may be run if a significant difference is found by trained testers. The two types of tests can give quite different answers. If a controlled test is being run, it is necessary to take a number of precautions (repackaging, changing recipe wording, etc.) to prevent brands from being recognized. Extended use tests may again give different answers than ordinary use tests. Repeated paired comparison tests are very useful in such situations. Test marketing may give still different answers, because such things as the eye-appeal of the package, branding, etc. may reverse the superiority in the quality of the product.

Evaluation of the Effect of Branding, Advertising, Promotion or Packaging Problems

This type of problem consists of evaluating the effect any one of these elements of the marketing mix has upon consumer preferences. Such research consists of more than determining that one brand is preferred over another, or that one promotional alternative is preferred over another. Instead, this type of research is concerned with the determination of which market segments within the population are most favorably inclined to buy each brand under the stated conditions. Thus, the panelists' evaluations must be interpreted by population segments.

This research is also conducted to determine how much extra the consumer is likely to pay to obtain the brand which he prefers, or the advertised

product versus the non-advertised product. Although the initial testing may be done by having consumers evaluate the samples on the basis of appearance or use, it eventually becomes necessary to use tests which can yield indications of how much extra money the panelist would be willing to pay for his preferred product. The VonNeumann-Morgenstern Model, the demand-price concept, or retail sales testing may be used for this purpose.

Cost Reduction Problems

Such programs arise when the objective is to reduce the cost of a product already on the market by using less expensive ingredients in the formulation, or by using a less expensive process, without significantly lowering acceptance by the consumer.

The purpose of the trained laboratory panel tests is to ensure that any cost-reduced products submitted by the research worker are evaluated against the established products with a design and panel size that guarantee probabilities of Type I and Type II errors of only a tolerable magnitude. Any product found to differ significantly from the established product is not allowed to pass on to the next stage of consumer testing (except when the variant, though different and less expensive, is found to be superior in quality). Those variants found not significantly different are prepared for consumer testing. As the cost of unnecessary consumer tests may be relatively expensive and the researcher can readily go back to the laboratory

and produce a better cost-reduced sample, the probability of a Type II error may be small and the probability of a Type I error moderately small. The method of repeated paired comparisons should be used at the later stage of consumer testing of the promising products, for the principal interest is in estimating the proportion of the population having no consistent preference. The design should allow for the high cost involved in letting an inferior product be incorrectly marketed as against the lower cost of having to do more work finding a better cost-reduced product.

Quality Improvement Problems

The purpose of such programs is to achieve a significant improvement in consumer preference by improving the quality of a product already on the market. The improved product usually costs more to produce than the current one, thus, in a sense, this program is the opposite of the cost reduction problem.

The trained laboratory panel serves to ensure that, of the improved samples submitted for evaluation against the product now in production, only those samples which are sufficiently better than the product being produced will be consumer tested. The emphasis should be placed on a low probability of a Type I error, while a moderately low probability of a Type II error can be tolerated if the cost of developing a better product is less than the cost involved in consumer preference testing. This is to ensure that no needless consumer tests are run, and to a lesser extent,

that th

The co

produc

suffici

On the

research

Variation

Wh

product

the pro

changes

Routine

Wh

to insur

ceptanc

tine qua

Storage

So

life of

It is us

drawal

drawn

tion ag

that the research worker not be sent back to his laboratory unnecessarily.

The consumer test phase should allow for the risk of recommending that production be replaced by a new product, when in fact that product is not sufficiently superior to compensate for the extra cost with extra sales.

On the other hand, one should weigh the risk of unnecessarily sending the research staff to its laboratory to produce a new, improved product.

Quality Control Problems

Variation in Source of Ingredients

When the source, nature, or proportion of ingredients in a current product has to be changed, or when a new piece of equipment is added to the procedure, it is necessary to ensure that the product is not significantly changed.

Routine Quality Control

When no objective tests are available, or when they are insufficient to insure that outgoing consignments of a given product are alike in acceptance, it may be necessary to use the trained laboratory panel for routine quality control.

Storage Studies

Sometimes it is intended to discover whether the shelf-life or storage life of a given product is maintained with reference to a given standard. It is usual to place samples of this product into storage and to make withdrawals of random samples from storage at periodic intervals. The withdrawn samples are then submitted to the trained laboratory panel for evaluation against fresh material or a parallel product which must be stored at

such low temperature that no appreciable loss in quality could have occurred.

In all three quality control problems, the testing usually begins and ends with the trained laboratory panel. No consumer tests are necessary except when, despite the trained laboratory panel finding a significant difference, it is hoped to show no significant difference in consumer acceptance. Because the trained panel works in the direction of trying to show no significant difference, the emphasis in design is on a small probability of a Type II error. In this particular application in the trained laboratory panel, it is often found useful to apply sequential triangle tests. By using this method of testing, one can realize prespecified Type I and Type II errors more readily than with single-stage designs. The plan is also more economical of judges' time.

It should be pointed out that the generalizations mentioned in the above presentation have many working exceptions. Considerations such as type of panel, research designs, types of data analysis, sample size, etc. are often set by the organization's budget, the time period in which the project must be completed, or the level of precision which the firm desires.

It is hoped that this study will provide information which people engaged in consumer preference testing can utilize when they undertake a study in this area. More specifically, it is hoped that this study has provided the information which these people can use as an aid in selecting the procedure that will best satisfy the project objective, while remaining within the organization's budget, time, and talent restraints.

It is also hoped that this study will make people performing consumer preference research aware of the advantages and limitations of the various types of panels, designs, and methods of data analysis.

APPENDICES

APPENDIX A

Sum of the Ranks for Sausage Containing Four Different Levels of Sugar

Percent Sugar in Sausage	Sum of the Ranks for this Sausage	Deviations From Average Rank	Squared Deviations From Average Rank
0.0%	282.5	32.5	1056.25
1.0%	259.5	9.5	90.25
1.5%	221.0	-29.0	841.0
2.0%	237.0	-13.0	169.0
Total	1000.0		s = 2156.5

Source

Row

Column

Product

Treatment

Treatment

or Factor

Treatment

or Factor

A x B

Error

Total

In this

are de

k, ru

from

$g = 1$

and

APPENDIX B

Analysis of Variance Table for Calculation of Sums of Squares from Balanced Lattice Designs

Source	Degrees of Freedom	Sums of Squares	Mean Squares	Expected Mean Square
Row	$r-1$	$\sum_{i=1}^r (\bar{X}_{i..} - \bar{X}...)^2$	$RSS/r-1$	$\sigma^2 + P\sigma_r^2$
Column	$c-1$	$\sum_{j=1}^c (\bar{X}_{.j.} - \bar{X}...)^2$	$CSS/c-1$	$\sigma^2 + P\sigma_c^2$
Products or Treatments	$ab-1$	$\sum_{k=1}^m (\bar{X}_{..k} - \bar{X}...)^2$	$TrSS/rc-1$	$\sigma^2 + Q\sigma_T^2$
Treatment A or Factor	$a-1$	$\sum_{s=1}^a (\bar{X}_{..s} - \bar{X}...)^2$	$T_aSS/a-1$	$\sigma^2 + P\sigma_A^2$
Treatment B or Factor	$b-1$	$\sum_{g=1}^b (\bar{X}_{..g} - \bar{X}...)^2$	$T_bSS/b-1$	$\sigma^2 + P\sigma_B^2$
A x B	$(a-1)(b-1)$	$T_rSS - (T_aSS + T_bSS)$	$ABSS/(a-1)(b-1)$	$\sigma^2 + P\sigma_{AB}^2$
Error	Subtract	$ToTSS - (TrSS + RSS + CSS)$	ESS/EDF	σ^2
Total	$rcq-1$	$\sum_{ijk} (X_{ijk} - \bar{X}...)^2$		

In this table, the rows are designated by i running from $i = 1$ to r , columns are designated by j , running from $j = 1$ to c , products are designated by k , running from $k = 1$ to m , levels of product A are designated by s , running from $s = 1$ to a , levels of product B are designated by g , running from $g = 1$ to b , P indicates the number of times the entire design was repeated, and Q indicates the number of replicates in the design.

T
of five

(
pair o

this d

ber of

(
pair o

(
that e

equal.

P
RANGI

Reps.

Block

(1)

(2)

(3)

(4)

(5)

APPENDIX C

Examples of Other Balanced Incomplete Block Designs that Can Be Used to Evaluate Five or More Products

The following types of designs may be used to guide the presentation of five or more samples to a consumer preference panel:

(1) The balanced lattice designs which are arranged so that every pair of products occurs together once in the same block in the design. In this design the number of products must be an exact square while the number of products presented in one block is the corresponding square root.

(2) The balanced incomplete block designs are arranged so that every pair of products occurs together once in some block in the design.

(3) The doubly balanced incomplete block designs are arranged so that each pair of products and each trio of products appears together equally often in the design.

PLAN FOR EVALUATING FIVE PRODUCTS: BALANCED DESIGN ARRANGED IN RANDOMIZED INCOMPLETE BLOCKS¹

Reps. I, II, and III

Reps. IV, V, and VI

Block Products

Block Products

(1) 1 2 3
(2) 1 2 5
(3) 1 4 5
(4) 2 3 4
(5) 3 4 5

(6) 1 2 4
(7) 1 3 4
(8) 1 3 5
(9) 2 3 5
(10) 2 4 5

¹ Cochrane and Cox, op. cit., p. 471.

PLAN FOR EVALUATING SIX PRODUCTS: BALANCED DESIGN ARRANGED
IN RANDOMIZED INCOMPLETE BLOCKS¹

Reps. I and II

<u>Block</u>	Products	<u>Block</u>	Products	<u>Block</u>	Products
(1)	1 2 3 4	(4)	1 2 3 5	(7)	1 2 3 6
(2)	1 4 5 6	(5)	1 2 4 6	(8)	1 3 4 5
(3)	2 3 5 6	(6)	3 4 5 6	(9)	2 4 5 6
<u>Block</u>	Products	<u>Block</u>	Products		
(10)	1 2 4 5	(13)	1 2 5 6		
(11)	1 3 5 6	(14)	1 3 4 6		
(12)	2 3 4 6	(15)	2 3 4 5		

PLAN FOR EVALUATING SEVEN PRODUCTS: BALANCED DESIGN AR-
RANGED IN RANDOMIZED INCOMPLETE BLOCKS²

Reps. I, II, III, and IV

<u>Block</u>	Products
(1)	3 5 6 7
(2)	1 4 6 7
(3)	1 2 5 7
(4)	1 2 3 6
(5)	2 3 4 7
(6)	1 3 4 5
(7)	2 4 5 6

PLAN FOR EVALUATING EIGHT PRODUCTS: DOUBLY BALANCED IN-
COMPLETE BLOCK ANALYSIS³

Number of replications equal seven

<u>Block</u>	Products	<u>Block</u>	Products
(1)	1 2 3 4	(8)	2 3 5 8
(2)	5 6 7 8	(9)	1 2 5 6
(3)	1 2 7 8	(10)	3 4 7 8
(4)	3 4 5 6	(11)	1 3 5 7
(5)	1 3 6 8	(12)	2 4 6 8
(6)	2 4 5 7	(13)	1 4 5 8
(7)	1 4 6 7	(14)	2 3 6 7

¹Ibid., p. 472.

²Ibid., p. 473.

³Calvin, op. cit., p. 83.

PLAN FOR EVALUATING NINE PRODUCTS: 3 BY 3 BALANCED LATTICE DESIGN GIVEN IN CHAPTER VII.

PLAN FOR EVALUATING TEN PRODUCTS: DOUBLY BALANCED INCOMPLETE BLOCK ANALYSIS¹

Number of replications equal 12

<u>Block</u>	Products	<u>Block</u>	Products
(1)	7 8 9 10	(16)	1 3 7 9
(2)	3 6 9 10	(17)	2 5 7 9
(3)	1 5 9 10	(18)	4 6 7 9
(4)	2 4 9 10	(19)	3 4 5 9
(5)	2 5 8 10	(20)	1 2 6 9
(6)	3 4 8 10	(21)	1 2 7 8
(7)	1 6 8 10	(22)	3 6 7 8
(8)	1 4 7 10	(23)	4 5 7 8
(9)	3 5 7 10	(24)	2 4 6 8
(10)	2 6 7 10	(25)	1 3 5 8
(11)	4 5 6 10	(26)	1 5 6 7
(12)	1 2 3 10	(27)	2 3 4 7
(13)	1 4 8 9	(28)	2 3 5 6
(14)	2 3 8 9	(29)	1 3 4 6
(15)	5 6 8 9	(30)	1 2 4 5

PLAN FOR EVALUATING ELEVEN PRODUCTS: BALANCED DESIGN ARRANGED IN RANDOMIZED INCOMPLETE BLOCKS²

Number of replications equal 5

<u>Block</u>	Products
(1)	1 2 3 5 8
(2)	2 3 4 6 9
(3)	3 4 5 7 10
(4)	4 5 6 8 11
(5)	5 6 7 9 1
(6)	6 7 8 10 2
(7)	7 8 9 11 3
(8)	8 9 10 1 4
(9)	9 10 11 2 5
(10)	10 11 1 3 6
(11)	11 1 2 4 7

¹Ibid., p. 84.

²Cochrane and Cox, op.cit., p. 476.

PLAN FOR EVALUATING TWELVE PRODUCTS: DOUBLY BALANCED IN-COMplete BLOCK ANALYSIS¹

Number of replications equal 11

<u>Block</u>	Products						<u>Block</u>	Products					
(1)	1	2	3	5	8	12	(12)	4	6	7	9	10	11
(2)	2	3	4	6	9	12	(13)	1	5	7	8	10	11
(3)	3	4	5	7	10	12	(14)	1	2	6	8	9	11
(4)	4	5	6	8	11	12	(15)	1	2	3	7	9	10
(5)	1	5	6	7	9	12	(16)	2	3	4	8	10	11
(6)	2	6	7	8	10	12	(17)	1	3	4	5	9	11
(7)	3	7	8	9	11	12	(18)	1	2	4	5	6	10
(8)	1	4	8	9	10	12	(19)	2	3	5	6	7	11
(9)	2	5	9	10	11	12	(20)	1	3	4	6	7	8
(10)	1	3	6	10	11	12	(21)	2	4	5	7	8	9
(11)	1	2	4	7	11	12	(22)	3	5	6	8	9	10

PLAN FOR EVALUATING THIRTEEN PRODUCTS: BALANCED DESIGN ARRANGED IN RANDOMIZED INCOMPLETE BLOCKS²

Number of replications equal 4

<u>Block</u>	Products				<u>Block</u>	Products			
(1)	1	2	4	10	(8)	8	9	11	4
(2)	2	3	5	11	(9)	9	10	12	5
(3)	3	4	6	12	(10)	10	11	13	6
(4)	4	5	7	13	(11)	11	12	1	7
(5)	5	6	8	1	(12)	12	13	2	8
(6)	6	7	9	2	(13)	13	1	3	9
(7)	7	8	10	3					

¹Calvin, op. cit., p. 87.

²Cochrane and Cox, op. cit., p. 477.

PLAN FOR EVALUATING FIFTEEN PRODUCTS: BALANCED DESIGN ARRANGED IN RANDOMIZED INCOMPLETE BLOCKS¹

Rep. I		Rep. II		Rep. III	
<u>Block</u>	Products	<u>Block</u>	Products	<u>Block</u>	Products
(1)	1 2 3	(6)	1 4 5	(11)	1 6 7
(2)	4 8 12	(7)	2 8 10	(12)	2 9 11
(3)	5 10 15	(8)	3 13 14	(13)	3 12 15
(4)	6 11 13	(9)	6 9 15	(14)	4 10 14
(5)	7 9 14	(10)	7 11 12	(15)	5 8 13
Rep. IV		Rep. V		Rep. VI	
<u>Block</u>	Products	<u>Block</u>	Products	<u>Block</u>	Products
(16)	1 8 9	(21)	1 10 11	(26)	1 12 13
(17)	2 13 15	(22)	2 12 14	(27)	2 5 7
(18)	3 4 7	(23)	3 5 6	(28)	3 9 10
(19)	5 11 14	(24)	4 9 13	(29)	4 11 15
(20)	6 10 12	(25)	7 8 15	(30)	6 8 14
Rep. VII					
<u>Block</u>	Products				
(31)	1 14 15				
(32)	2 4 6				
(33)	3 8 11				
(34)	5 9 12				
(35)	7 10 13				

PLAN FOR EVALUATING SIXTEEN PRODUCTS: 4 BY 4 BALANCED LATTICE DESIGN GIVEN IN CHAPTER VII

¹Ibid. , p. 478.

APPENDIX D

Analysis of Variance Computations for the 4 by 4 Balanced Lattice Experiment on Consumer Preferences for Ham

An analysis of variance computation revealed that the total sums of squares equals 129,706- the correction term of 128,000, which is 1706.

The treatment or product sums of squares = $(200)^2 + (187)^2 + (179)^2 + (212)^2 + (206)^2 + (214)^2 + (171)^2 + (209)^2 + (210)^2 + (200)^2 + (197)^2 + (202)^2 + (210)^2 + (201)^2 + (195)^2 + (207)^2 = 642,196 + 5 = 128,439.2$, which minus the correction term of 128,000, equals 439.2. Where 200 is the sum of the rankings for product one, 187 is the sum of the rankings for product 2, ...and 207 is the sum of the rankings for product 16.

The sums of squares for the sugar effect = $(826^2 + 802^2 + 742^2 + 830^2) / 20 = 128,247.2$, from which the correction term of 128,000 is subtracted, giving 247.2 as the sum of squares, where 826 is the sum of rankings for all products containing 0% sugar, 802 is the sum for all products containing 1% sugar, 742 is the sum of the rankings for all products containing 2% sugar, and 830 is the sum for all products containing 3% sugar.

The sum of squares for salt = $(778^2 + 800^2 + 809^2 + 813^2) / 20 = 128,036.7$, from which the correction of term of 128,000 is subtracted, giving 36.7 as the salt sums of squares. The 778 is the sum of all rankings for all products containing 1.5% salt, 800 is the sum of all rankings for the products containing 2.0% salt, 809 is the sum of all rankings for all products containing 2.5% salt, and 813 is the sum of all rankings for all products containing 3.0% salt.

APPENDIX E

The Scheffe Analysis of Variance

The Scheffe method is essentially an analysis of variance using scored data. The scoring systems used rate the products according to the degree of preference, rather than by a preferential selection. The five point preference scale contained in Table XIX illustrates the scoring of product i when it is compared to product j.

In this method, the experiment is completely replicated with the order of presentation reversed in the second replication. In other words, if product i is observed before product j for the first time, then product i is presented after product j the second time.

One can let x_{ijk} be the k^{th} observation on the ordered pair (i, j), and assume all x_{ijk} are independent random variables with mean μ_{ij} and variance σ^2 .

The mean preference for treatment i to treatment j is denoted as μ_{ij} when presented in the order (i, j) and $-\mu_{ji}$ in the order (j, i). The average preference π_{ij} and the average difference due to the order of presentation δ_{ij} are given by:

$$\begin{aligned}\pi_{ij} &= 1/2 (\mu_{ij} - \mu_{ji}), & \pi_{ji} &= -\pi_{ij}, \\ \delta_{ij} &= 1/2 (\mu_{ij} + \mu_{ji}), & \delta_{ji} &= \delta_{ij}.\end{aligned}$$

The average order effect is given by:

$$\delta = \sum_{i \neq j} \mu_{ij} / (2M) \text{ which equals } \sum_{i < j} \delta_{ij} / M.$$

In this equation, M equals the number of pairs; e.g. $M=m(m-1)/2$ and m equals the number of products. Each pair is presented to $2n$ judges, to r judges in the order (i, j) and to r in the order (j, i) .

Scheffe's "hypothesis of subtractivity" postulates that parameters $\alpha_1, \alpha_2, \dots, \alpha_m$ characterize the m products such that the average preference π_{ij} for the i to j is the difference between the corresponding parameters $\pi_{ij} = \alpha_i - \alpha_j, \pi_{ii} = 0$; thus, the deviations from subtractivity γ_{ij} are given by: $\gamma_{ij} = \pi_{ij} - (\alpha_i - \alpha_j)$. The α_i may be regarded as the main effects. The "hypothesis of subtractivity" states that the results for any pair, after order effects are eliminated, can be attributed entirely to the difference of the two main effects of the two brands in the pair.

The sources of variation, their corresponding degrees of freedom, and sum of squares are given in Table XXV.

TABLE XXV. Scheffe Analysis of Variance Procedure¹

Source	Degrees of Freedom	Sum of Squares	H_0 :
Main effects ($\alpha_i - \alpha_j$)	$m-1$	S_α	$\alpha_i = 0$
Deviations from subtractivity (γ_{ij})	$M - (m-1)$	S_γ	$\gamma_{ij} = 0$
Average preferences (π_{ij})	M	S_π	
Order effects (δ_{ij})	(M)	(S_δ)	
Av. order effect (δ)	1	$2rM\delta^2$	$\delta = 0$
Diff. among order effects ($\delta_{ij} - \delta$)	$M-1$	$S_\delta - 2rM\delta^2$	$\delta_{ij} - \delta = 0$

¹Otto Dykstra, Jr., "Factorial Experimentation in Scheffe's Analysis of Variance for Paired Comparisons, " American Statistical Association Journal (Vol. 53, June 1958), pp. 529-542.

TABLE XXV--(continued)

Source	Degrees of Freedom	Sum of Squares	H ₀ :
Means (μ_{ij})	2M	S_{μ}	
Error (e_{ijk})	2M(r-1)	S_e	
Total (x_{ijk})	2rM	S_t	

The following formulas can be used to estimate the various effects and to compute the sums of squares:

EFFECT	ESTIMATION	SUMS OF SQUARES
x_{ijk}		$S_t = \sum_{i \neq j} \sum_k x_{ijk}^2$
μ_{ij}	$r\hat{\mu}_{ij} = \sum_{k=1}^r x_{ijk}$	$S_{\mu} = \frac{1}{r} \sum_{i \neq j} 2(r\hat{\mu}_{ij})^2$
π_{ij}	$2_r\hat{\pi}_{ij} = r\hat{\mu}_{ij} - r\hat{\mu}_{ji}$	$S_{\pi} = \frac{1}{2r} \sum_{i < j} (2_r\hat{\pi}_{ij})^2$
δ_{ij}	$2r\hat{\delta}_{ij} = r\hat{\mu}_{ij}$ plus $r\hat{\mu}_{ji}$	$S_{\delta} = \frac{1}{2r} \sum_{i < j} (2r\hat{\delta}_{ij})^2$
ϕ	$2rM\hat{\phi} = \sum_{i \neq j} \sum_k x_{ijk}$	$2rM\hat{\phi}^2 = \frac{1}{2rM} \left(\sum_{i \neq j} \sum_k x_{ijk} \right)^2$
α_i	$2rm\hat{\alpha}_i = \sum_{j=1}^m 2_r\hat{\pi}_{ij}$	$S_{\alpha} = \frac{1}{2rm} \sum_i (2rm\hat{\alpha}_i)^2$

Scheffe gives a convenient way of tabulating the data and calculating the sums of squares.¹

¹Henry Scheffe, "An Analysis of Variance for Paired Comparisons," American Statistical Association Journal (September 1952), pp. 390-395.

APPENDIX F

Regression Data for UBT and KBT in Demand-Price Study

List of Prices and the Number of Panelists Who Would Purchase the UBT and KBT at Each Price

Price	Number of panelists who would purchase the UBT	Number of panelists who would purchase the KBT
\$2.00	154	155
\$2.50	149	155
\$3.00	144	153
\$3.25	136	150
\$3.50	132	147
\$3.75	122	136
\$4.00	91	115
\$4.25	40	54
\$4.50	28	34
\$4.75	18	25
\$5.00	10	13
\$5.50	6	8
\$6.00	0	7

BIBLIOGRAPHY

- Anderson, R. L. "Use of Contingency Tables in the Analysis of Consumer Preference Studies," Biometrics, XV (December 1959), 582-59.
- Arnold, C. L. "Do Consumers Have Good Taste?" Food Industries, XIII (March 1941), 45.
- Bayton, James A. "Consumer Preference Research in the Development of Agriculture," Agriculture Economics Research, II (October 1950), 105-112.
- Bayton, James A. "Discussion: New Techniques in Consumer Preference Research," Journal of Farm Economics, XXXVIII (1956).
- Bayton, James A., and Bell, H. P. Preference for Canned Orange Juices that Vary in Brix-Acid Ratio. U.S.D.A. Market Research Report No. 178, 1954.
- Beazley, Charles C. "Product Factors vs. Consumer Acceptance." General Foods Corporation, Battle Creek, Michigan, 1963. (Unpublished, Mimeographed.)
- Beazley, Charles C., and Elsesser, Charles C. "An Industrial Example--Examining Six Variables Simultaneously." General Foods, 1963. (Mimeographed.)
- Bengtsson, K., and Helm, E. "Principles of Taste Testing," Wallerstein Laboratorys Communication. 1946.
- Benson, Purnell H. "A Model for the Analysis of Consumer Preference and Exploratory Tests," Journal of Applied Psychology (October 1955), 375-381.
- Berg, H. W., Filipello, F., Hinreiner, E., and Sawyer, F. M. "Consumer Wine Preference Methodology Studies at California Fairs," Food Technology, IX (1955).
- Bradley, Ralph, and Terry, Milton. "Rank Analysis of Incomplete Block Designs," Biometrika, XXXIX (1952), 324-345.
- Bradley, Ralph. "Some Relationships Among Sensory Difference Tests," Biometrics, XIX (1963), 385-397.

- Brown, Geo. E. "Measuring Consumer Attitudes Toward Products," Motivation and Market Behavior. "ed." Robert Ferber and H. G. Wales. Homewood, Illinois: Irwin, 1958.
- Brown, Geo. H. "Measuring Consumer Attitudes Toward Products," Journal of Marketing (April 1950), 691-698.
- Brunk, Max E. "Discussion of Research on Consumer Behavior and Preferences," Market Demand and Product Quality, A Report of the Marketing Research Workshop. Michigan State College. July 1951.
- Byer, Albert J., and Abrams, Dorothy. "A Comparison of the Triangular and Two-Sample Taste-Test Methods," Food Technology, VII (1953), 185-187.
- Calvin, Lyle. "Doubly Balanced Incomplete Block Designs for Experiments in which the Treatment Effects Are Correlated," Biometrics, X (1954), 61-88.
- Carroll, M. B. "The Use of Subjective Evaluation in Product Control and Development." General Foods Research Center, 1963. (Mimeographed.)
- Carter, Roy E. Jr., Troidahl, Verling C., and Schuneman, R. Smith. "Interviewer Bias in Selecting Households," Journal of Marketing (April 1963), 27-34.
- Clements, F. E., Bayton, J. A., and Bell, H. P. "Method of Single Stimulus Determinations of Taste Preference," Journal of Applied Psychology, XXXVIII, No. 6 (1954).
- Cochran, William G. Sampling Techniques. John Wiley & Sons, Inc.: New York and London, 1953.
- Cochran, William G., and Cox, Gertrude M. Experimental Designs. John Wiley & Sons, Inc.: New York, 1957.
- "Consumer Preferences for Frozen Peas in Relation to Standards for Grades," Marketing Research Report No. 280 (U.S.D.A., A.M.S., 1958).
- Cowan, Donald G. "Developing and Improving Foods by Consumer Testing," Food Industries, XIII (1941).
- Crocker, E. C. Flavor. New York and London, 1945.
- Davidson, Donald, and Suppes, Patrick. Decision Making, An Experimental Approach. Stanford University Press: Stanford, California, 1957.

- Davies, Owen L. The Design and Analysis of Industrial Experiments. Oliver and Boyd: London, 1954.
- Dawson, Elsie H., and Harris, Betsy L. Sensory Methods for Measuring Differences in Food Quality, Agriculture Information Bulletin No. 34, U.S.D.A., Washington, D. C., 1951.
- Dixon, Wilford, and Massey, Frank J. Jr. Introduction to Statistical Analysis. McGraw-Hill Book Company, Inc.: New York, 1957.
- Dove, W. F. "Food Acceptability--Its Determination and Evaluation," Food Technology, I (1947), 39-50.
- Dykstra, Otto Jr. "Factorial Experimentation in Scheffe's Analysis of Variance for Paired Comparisons," American Statistical Association Journal, LIII (June 1958), 529-542.
- Eastwood, R. P. Sales Control by Quantitative Methods, Columbia University Press, 1940, 288-293.
- Edwards, W. "The Prediction of Decisions Among Bets," Journal of Experimental Psychology, L (1955), 213.
- Edwards, W. "The Theory of Decision Making," Psychological Bulletin, LI (1954), 381.
- Filipello, F. "A Critical Comparison of the Two-Sample and Triangular Binomial Designs," Food Research, XXI (March-April 1956), 235-241.
- Filipello, F., Berg, W. H., Hinreiner, E., and Webb, A. D. "Reproducibility of Results in Consumer Wine-Preference Survey," Food Technology, IX (1955).
- Franzen, Raymond, and Teilhet, Darwin. "A Method for Measuring Product Acceptance," Journal of Marketing, V (October 1940), 156-161.
- Garnatz, George. "But What Do the Consumers Say?" Food Industries, XXII (1950).
- Greenberg, A., and Manfield, M. N. "On the Reliability of Mail Questionnaires in Product Tests," The Journal of Marketing, XXI (1956), 342-345.
- Gridgeman, N. T. "Taste Comparisons: Two Samples or Three?" Food Technology, IX (March 1955), 148.

- Grieg, W. Smith. "Quality Competition and Product Development." Agricultural Market Analysis: Development, Performance, Process. Chapt. VIII. (To be published.)
- Grieg, W. Smith, and Larzelere, H. "Consumer Taste Preferences Among Dehydrated Mashed Potato Products," National Potato Council News, V, No. 2 (1954).
- Hanson, Helen Louise. Sensory Difference Tests in the Evaluation of Quality. A Report of the Marketing Research Workshop, Michigan State College, July 13 to 21, 1951. Prepared by the United States Department of Agriculture, Washington: Agricultural Research Administration, 1951, 73-79.
- Hanson, H. L. "Sensory Test Methods II, Effect of Previous Tests on Consumer Response to Foods," Food Technology, IX (1955).
- Helm, E., and Trolle, B. Selection of a Taste Panel, Wallerstein Laboratorys Communication, IX (1946), 181-184.
- Higgins, Glen Willis. "Consumer Attitudes and Behavior Toward Frozen Meats." Unpublished Master's dissertation, Michigan State University, 1958.
- Hopkins, J. W. "Precision of Assessment of Palatability of Foodstuffs by Laboratory Panels," Canadian Journal of Research, Section F, Technology 24 (1946), 203-214.
- Hunter, J. Scott. "Consumer Preference for a 6 to 1 Apple Juice Concentrate," U.S.D.A. Marketing Research Report No. 343, Washington: 1959.
- Jackson, J. Edward, and Fleckstein, Mary. "An Evaluation of Some Statistical Techniques Used in the Analysis of Paired Comparison Data," Biometrics, XIII (1957), 57.
- Johnson, Gerald C. "Effective Marketing Begins on the Design Board," Journal of Marketing, XIII (July 1948), 33.
- Jones, Ralph W. "Management of New Products," Managerial Marketing: Perspectives and Viewpoints. "ed." William Lazer and Eugene Kelley, Homewood, Illinois: Richard D. Irwin, Inc., 1962, 444.
- Kendall, Maurice G. Rank Correlation Methods. London: Griffin Co., 1948.

- Kiehl, Elmer R., and Rhodes, Jas. "New Techniques in Consumer Preference Research," Journal of Farm Economics, XXXVIII (December 1956).
- Knowles, D., and Johnson, P. E. "A Study of the Sensitiveness of Perspective Food Judges to the Primary Tastes," Food Research, VI, 207-216.
- Larzelere, H. E., and Dedolph, R. R. "Consumer Acceptance of Greenhouse-Grown and Southern Field-Grown Tomatoes," Quarterly Bulletin of the Michigan Agricultural Experiment Station, XXXIV (1962), 554-558.
- Larzelere, H. E., and Gibb, R. D. "Consumers' Opinions of Quality in Pork Chops," Quarterly Bulletin of the Michigan Agricultural Experiment Station, XXXIX (1956), 327-333.
- Lazo, Hector, and Corbin, Arnold. Management in Marketing. McGraw-Hill Book Company, Inc.: New York, 1961, 191.
- Makens, J. C. "The Influence of Brands Upon Consumer Preferences for Frozen Whole Turkey." Unpublished Ph.D. dissertation, Michigan State University, 1963.
- Manderscheid, Lester V. "An Introduction to Statistical Hypothesis Testing." Department of Agricultural Economics, May 1962, 867. (Mimeographed.) Michigan State University.
- McCarthy, I. "How Can We Make Sure of Pleasing the Tastes of Mrs. America and Her Family?" Hallmark Testing Service, Safeway Stores, Inc., Oakland, California.
- McCreary, Kirk. Unpublished manuscript, 4.
- Metz, J. F. Jr. "Accuracy of Response Obtained in a Milk Consumption Study," Methods of Research Marketing, Paper No. 5, Department of Agricultural Economics, Cornell University, 1956.
- Meyers, Trienah. "Predicting Market Acceptance," Journal of Farm Economics, XXXVII (1955), 1384-1394.
- Miller, P. G., Nair, J. H., and Harriman, A. J. "A Household and Laboratory Type of Panel for Testing Consumer Preference," Food Technology, IX (1955), 449.

- Mills, W. C. Jr., Dawson, L. E., and Larzelere, H. E. "Consumer Preferences for Turkeys with Various Grade Labels," Quarterly Bulletin of the Michigan Agricultural Experiment Station, Vol. 43, No. 2 (November 1960), 258.
- Morgan, James. "A Review of Recent Research on Consumer Behavior," Consumer Behavior-Research on Consumer Reactions, "ed." L. H. Clark (New York: Harper Brothers, 1958).
- Nair, John H. "Mass Taste Panels," Food Technology, III (1949), 131-136.
- O'Dell, William F. "Personal Interviews or Mail Panels?" Journal of Marketing, XXVI (October 1962), 34-39.
- Ortengren, John. "When Don't Research Panels Wear Out?" Journal of Marketing, XXI (April 1957), 442.
- Oxenfeldt, Alfred R. Pricing for Marketing Executives. Wadsworth Publishing Company: San Francisco, 1961.
- Peryam, D. R., and Girardot, N. F. "Advanced Taste-test Methods," Food Engineering, XXIV (1952), 61.
- Peryam, D. R., and Gutman, N. J. "Variations in Preference Ratings for Foods Served at Meals," Food Technology, IX (May 1955).
- Platt, Washington. "How Will Consumers Rate Your Product?" Food Industries, IX (1937).
- Platt, Washington. "Why Consumer-Preference Tests?" Food Industries, XIII (March 1941), 40-41.
- Polits, Alfred. "What Lies Behind Consumer Preferences?" Market Demand and Product Quality, A Report of the Marketing Research Workshop. Michigan State College, July 1951.
- "Product Evaluation, An Examination of Research Procedures," Bulletin Prepared by Market Facts, Incorporated, Chicago, Illinois, 1964.
- Remingea, Elmer. Mimeographed Outline for Analysis of Variance Course, Colorado State University, Winter Term, 1960.
- Robinson, R. A., and Agisim, Philip. "Making Mail Surveys More Reliable," Journal of Marketing, XV (April 1951), 415-424.

- Rockwood, Albert M. "Planning a Product-Development Program," Managerial Marketing: Perspectives and Viewpoints. "ed." William Lazer and Eugene J. Kelly, Homewood, Illinois: Richard D. Irwin, Inc., 1962, 444.
- Roessler, E. B., Baker, G. A., and Marine, M. A. "One-Tailed and Two-Tailed Tests in Organoleptic Comparisons," Food Research, XXI, 117-121.
- Roessler, E. B., Warren, J., and Guymon, J. F. "Significance in Triangular Taste Tests," Food Research, XIII, 503-505.
- Savage, I. Richard. "Nonparametric Statistics," American Statistical Association Journal (September 1957), 335.
- Schwartz, M., and Pratt, C. H. "Simultaneous vs. Successive Presentation in a Paired Comparison Situation," Food Research, XXI (1956), 103-108.
- Shaffer, J. D. "Estimating Population Characteristics by Mail Survey," Journal of Farm Economics, XLI (1959), 386.
- Shaffer, James D. "Methodological Bases for the Operation of a Consumer Purchase Panel." Unpublished Ph.D. dissertation, Michigan State University, 1952, 37.
- Siegel, Sidney. Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill Book Company, Inc.: New York, 1956.
- Snyder, L. H. "Inherited Taste Deficiency," Science, LXXIV (1931), 151-152.
- "Super Marketing, U.S.A." This Week Magazine, Eighth Biennial Grocery Study, New York (1957), 9.
- "The Big Challenge in Food Marketing." This Week Magazine, Eighth Biennial Grocery Study, New York (1959), 9.
- "The Super Market Industry Speaks." Chicago, Super Market Institute, 1958, 9.
- Valdes, Rose Marie, and Roessler, E. B. "Consumer Survey on Dessert Quality of Canned Apricots," Food Technology, X (1956), 481.
- Wald, Abraham. Sequential Analysis. John Wiley & Sons, Inc.: New York and London, 1947.

Walsh, Robert M. "What Is and What Is Not Good Consumer Research," Paper presented at the Twelfth Reciprocal Meat Conference, Washington: U.S.D.A., A.M.S., 1959.

Weckel, K. G. "Put Your Taste Buds to Work on Your Flavor Problems," International Association Milk Dealer Bulletin 34, 136-140.

Wells, William D. "Measuring Readiness to Buy," Harvard Business Review (July-August, 1961), 81-87.

Work Group I. "Techniques of Studying Consumer Behavior and Preference," A Report of the Marketing Research Workshop, Michigan State College, July 13 to 21, 1951, Prepared by the United States Department of Agriculture, Washington: Agriculture Research Administration, 1951, 94-95.