MISUSES OF STATISTICAL TECHNIQUES IN ADVERTISING CLAIM SUBSTANTIATION SURVEYS

Solomon Dutka and Lester R. Frankel, Audits & Surveys
650 Avenue of the Americas, New York, NY 10011

Introduction

In much of the national advertising of brands, products and services claims are made that stress their superiorities over competition. Prior to 1946 these claims were generally accepted at face value.

In the early 1940's, acting upon complaints from individual consumers and consumer groups concerning misleading advertising claims Congress finally in 1946 enacted legislation. Section 43(a) of the Lanham Trademark Act prohibits false description or representation, including words or other symbols in connection with the advertising of any good or service. Any person or group who feels that he has been damaged or likely to be damaged by such deception may institute a civil suit. As a result all advertising claims required substantiation.

The Federal Trade Commission (FTC) was the Agency responsible for the enactment of this legislation. Subsequently the FTC set up a set of standards that had to be adhered to in substantiating advertising claims. Such groups as the Food and Drug Administration, the National Division of the Council of Better Business Bureaus, advertising industry associations, professional associations such as AMA and ADA, the individual magazines, networks and magazines as well as their associations currently serve as watchdogs in this area. These groups as well as individuals have successfully challenged many unsubstantiated advertising claims and corrective actions were taken.

In 1970 operating under the assumption that consumers benefited from comparative advertising the FTC encouraged their use. Then, many more challenges were brought by competitors who felt that they had been injured unjustifiably by false claims. In some cases litigation occured in Federal Courts. In many cases advertising media became involved when both competitors were clients. As a consequence, many publishers and broadcasters require clearance review before an ad could be published or a TV commercial put on the air.(6)

In the process of substantiating advertising claims statistical surveys and statistical analytic techniques are often used. An examination of many of the complaints about false or misleading advertising claims all too frequently indicate that the statistical methods involved were misused or abused on many occasions.

This paper is concerned with survey methodology used in many of the advertising claims including the design, the execution and the statistical analysis of the findings. Although many misuses have been brought about by question wording and question sequencing these aspects are not discussed here.

1. Misuse - Employing A Partial Frame

In 1976 Litton Industries advertised the superiority of its microwave oven by making the following claim: "76 percent of independent microwave oven technicians surveyed recommended Litton."

In 1980 the FTC ruled that the claim was defective because the list from which the sample was selected contained only Litton authorized technicians who serviced Litton and at least one other brand. Those service technicians authorized by competition who also serviced Litton were not included.(7)

This type of questionable claim has been used by magazines who select a sample of their subscribers and ask them to compare the magazine with others they subscribe to.

2. Misuse - Projecting to a Finite Population Using a Non-Probability Sample

In order to prove that a claim by a cigarette company was misleading a competitive company conducted a communications test among smokers in a shopping mall.(8)

"The main finding of the communication test was that the statement 'Triumph Beats Merit' means that 'Triumph was Better Tasting than Merit' to 37 percent of the smokers."

This projected percentage was not accepted by the Court. The comment of Judge Sweet of the U. S. District Court, Southern District was:

"Further, I find that a mall intercept study as employed in this instance fails to produce a nationally projectable statistical percentage. The suburban character of the sample and the demonstrated effect here of an error in the weighting of the sample does not provide the statistical reliability that would be obtained by random sampling, for example. That is not to say, however, that I find a mall intercept study as such to be an unreasonable basis for comparative advertising. However, because of their "quick and dirty" nature, such studies—and ads based upon them—must receive particularly close scrutiny."

3. Misuse - Using the Sampling Unit As The Analytical Unit

Brand "X" Ski Boots claimed that, based upon samples of skiers selected at random at several ski slopes, more skiers preferred their brand over all competing brands.

In this case the sampling unit was the ski day and not the skier. As a consequence, frequent skiers have a higher probability of being included in the sample than the infrequent skier. Unless the observations have been weighted to take into account frequency of skiing this generalization cannot be made.

This type of misuse is quite common when site sampling is used to conduct studies among passengers such as at airports or shoppers at shopping malls. When findings based on this type of sample have been challenged the results have been rejected.

4. Misuse – Misrepresenting Type of Test Used and Reporting Partial Results

"Body on Tap" shampoo claimed that among tests with 900 women it received higher ratings than Prell on "body," higher than Flex for "conditioning" and higher than Sassoon for "strong, healthy looking" hair."
This claim was challenged and brought to Court by Vidal Sassoon, Inc.
It would appear that each woman compared four different brands. Such was not the case. The sample was divided into four groups of 225 each. Each group considered a different shampoo and after trying the shampoo rated it on a monadic scale for each of 27 characteristics. The claim reflected differences only for selected averages, i.e. those that favored the advertiser.(3)

5. Misuse – Juggling the "No Preference" "Don't Know" and "About the Same" Groups

The following set of two tables which appeared in a Booklet by TRIUMPH cigarettes was pointed out by Jon N. Zoler. (8)

Results of the Blind Test

(Comparing Triumph and Merit Non-Menthol)

"53% of the respondents preferred Triumph, 42% preferred Merit and 5% had no preference. This difference is statistically significant."

	Number	Percent
Prefer Triumph Non-Menthol	173	53*
Prefer Merit Non-Menthol	138	42*
No Preference	14	5
Total Respondents	325	100

*Significant at 95% level of confidence.

Question 1

#234 has 3 mg. tar while #281 has 8 mg. of tar. Taking this into consideration, which would you prefer to smoke?

Source: Triumph Booklet

In this case the difference between the expressed preference for TRIUMPH was significantly different from the preference for MERIT even when the "No Preference" group is taken into account.

A different situation occurs in the following:

Result of the Blind Test

(Comparing Triumph and Merit Non-Menthol)

"60% find the taste of Triumph as good as or better than Merit. This is a significantly larger percentage than the percentage of people stating that Merit tastes better. The difference between those stating Triumph is better tasting and those stating Merit is better tasting is not statistically significant."

Triumph Non-Menthol vs. Merit Non-Menthol

	Number	Percent
Triumph is...		
Much better tasting	45	14 ⎫
Somewhat better tasting	73	22 ⎬ 60%
About the same in taste	77	24 ⎭
Somewhat worse tasting	93	29 40%
Much worse tasting	36	11
No answer	1	–
Total respondents	325	100

*Significant at 95% level of confidence.

Question 2

Comparing the taste of the two cigarettes, how would you say the #234 cigarette you tried compares to the #281? Is the #234 cigarette...?

Source: Triumph Booklet

In this ad, if "About the same in taste" is omitted MERIT comes out ahead. The percent reporting TRIUMPH is better tasting now becomes 36 percent while 40 percent indicate MERIT is better tasting.

6. Misuse - Paired Comparison: Using an
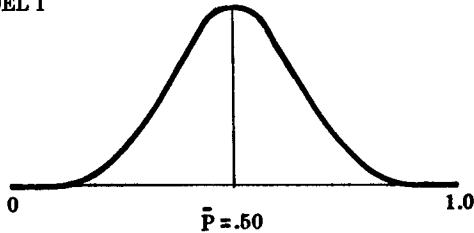   Incorrect Model

   CLAIM: People cannot tell the difference
between Brand A and Brand B.
   In a sample of 1000 people 51% preferred A
while 49% preferred B.
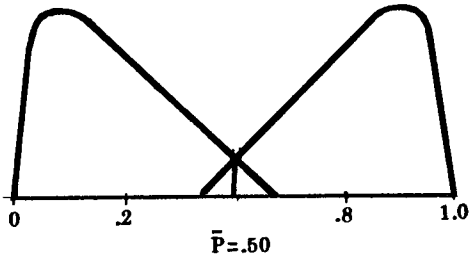   The sampling error of this 51% is 1.6
percent.

Assumed Model of Underlying Sampling Distribution.

MODEL I

0   P̄ =.50   1.0

Alternative Model Might look like ...

MODEL II

0   .2   P̄=.50   .8   1.0

To determine which model applies an Extended
Use Test is employed (4).

Procedure:
   1. Respondent is forced to select A or B.
   2. Test is repeated on two or more occasions

Premise:
   1. There is a latent probability distribution of preference,
   2. The tests are independent,
   3. The respondent's probability remains constant for each occasion.

Illustration:
   Two extended use tests for the choice between
Brand A and Brand B.
   Extended use over 3 occasions for 1,000
respondents in each example.

Selections		Model 1	Model II
Brand A 0	Brand B 3	124	358
Brand A 1	Brand B 2	351	126
Brand A 2	Brand B 1	396	143
Brand A 3	Brand B 0	129	373
	Total	1000	1000
Average Percent for A		51%	51%

7. Misuse - Using Triangular Tests to Eliminate Segments of the Population Who Cannot Discriminate Before Paired Comparison Tests Are Made.

   When forced choice paired comparison tests
are conducted there are certain subjects who
cannot discriminate between, for example, the
taste of two products. Thus, in response to
the question, "Which Tastes Better?", it can be
expected that 50 percent will name one product
and the other 50 percent of them will name the
other.
   The triangular test is used in an attempt to
eliminate the non-discriminators. Respondents
are given three samples to test, (1), (2) (5)
two are of one brand and the third the other.
Without seeing any brand identification they
are asked to select the odd one. It is expected that since the non-discriminators cannot
tell the difference two-thirds of them will
fail the test and on the basis of guessing one-
third of this group will pass. In order to
reduce this one-third a second triangular test
is given. In this way one-ninth of the non-
discriminators remain. The triangular test may
be repeated as many times as desired.
   The fallacy of this approach is that dis-
crimination is not a zero - one proposition,
there is a continunum and a probability distri-
bution. Hence as a result of a triangular test
even some discriminators tend to be eliminated.
If the process is continued indefinitely only
those with probability equal to one will remain
and the paired comparison test will be based
solely upon their tastes. This will represent
a very small segment of the real world market.

Suppose the population is divided in five
discrimination - preference groups as follows,
and a sample of 1000 is selected:

Group	(1) P(D)	(2) P(A)	(3) D	(4) C(A)	(5) F(A)
I	.0	None	.33	.500	25
II	.25	.2	.50	.425	400
III	.50	.4	.67	.450	450
IV	.75	.6	.83	.575	75
V	1.00	.8	1.00	.800	50

(1) Probability of Discrimination

(2) Underlying Preference for Brand A

(3) Probability of passing triangular test
     (Includes "Guessers")

(4) Probability of Choosing Brand A

(5) Frequency Distribution

Note: Columns (1), (2) and (5) are given;
columns (3) and (4) are derived.

The effect of triangular tests upon the size of the test group and its choice is shown below:

Number of Triangular Tests	Size Sample	Percent Choosing Brand A	"t" Value
0	1000	46.8	-2.0
1	620	48.4	-.8
2	404	50.2	.1
3	278	52.9	1.0
4	200	55.6	1.6
5	152	58.6	2.1
6	120	61.7	2.6
7	100	65.0	3.0
8	87	67.8	3.3
9	77	70.1	3.5
10	70	72.9	3.8
15	56	76.8	4.0
20	53	79.2	4.2
25	51	80.4	4.3

In the above example, the choice has shifted from a statistically significant negative value for Brand A to a statistically significant positive value. The shift has been in the direction of the most discriminating group. This group, however, represents only five percent of the sample.

Thus, by controlling the number of triangular tests, the percentage favoring a particular brand may be manipulated. This is a more subtle approach though no more correct than juggling the "Don't Know" Group.

REFERENCES

1. Buchanan, B. S. (1985), "Comparative Ads and Product Testing: What's a Victim To Do," Harvard Business Review

2. Buchanan, B. S. and Morrison, D. G. (1984), "Taste Tests Psychophysical Issues in Comparative Test Design," Psychology & Marketing, 1, 69-91.

3. Diamond, S. A. "Market Research Latest Target in Ad Claims", Advertising Age, January 25, 1982.

4. Gottlieb, M. J. (1962), A Modern Marketing Approach in Measuring Consumer Preference, Audits & Surveys, New York, N.Y.

5. Morrison, D. G. (1981), "Triangular Taste Tests: Are the Subjects Who Respond Correctly Lucky or Good?", Journal of Marketing, 45, 111-119.

6. Schneider, C. (1986), "Network Claim Substantiation: What We Expect/What We Look For," Delivered to AMA Marketing Research Conference, Orlando, Florida, September 30, 1986.

7. Turner C.E. and Martin. E. (1984), Surveying Subjective Phenomena, p. 45, Russell Sage Foundation, New York

8. Zoler, J. N. (1983), "Research Requirements for Ad Claim Substantiation," Journal of Advertising Research, 23, 9-16.