

**A DECISION FRAMEWORK FOR DETERMINING
THE LEVEL OF MONETARY INCENTIVE IN MAIL SURVEYS**

David A. Gautschi

and

Douglas L. MacLachlan

June 1995

Both authors are professors, Department of Marketing and International Business at the University of Washington, School of Business, Box 353200, Seattle, WA 98195-3200 U.S.A. E-mail: gautschi@u.washington.edu and macl@u.washington.edu.

INTRODUCTION

Widely used in marketing research, mail surveys appeal to marketing researchers for a number of reasons: they are low in cost, geographically flexible, and able to reach a widely dispersed sample simultaneously (Kanuk and Berenson, 1975). However, mail surveys suffer from a number of disadvantages, principal among these is the problem of low response rates. Acknowledging this problem of mail surveys, marketing researchers have invented a number of measures to increase response rates (Heberlein and Baumgartner, 1978). One ploy which has been used with some success is to include or promise a monetary reward for completing and returning the questionnaire (Wotruba, 1966; Armstrong, 1975; Cox, 1976; Gunn and Rhodes, 1981; Schewe and Cournoyer, 1976).

In this paper we investigate the relationship between the size of monetary inducement and the probability of response. One should expect the response probability to be a monotonically increasing function of the level of monetary inducement. In this context, the basic question of interest to the manager is how much of monetary inducement is needed to generate a response of a desired rate (or amount). The trade-off which the manager faces typically is either to vary the monetary incentive to achieve some target rate of response or to vary the size of the sample to assure a responding sample of desired size. Both of these options involve costs. Obviously, the cost of administering each questionnaire rises directly with the value of the monetary inducement. Similarly, as the sample size increases, additional costs of administration are incurred.

In the next section of the paper, we discuss the principal features of the response model called NORMIT. We then discuss an empirical application illustrating the calibration of the response model. Because the response model is similar to the well-known quantal response models, logit and probit, we compare the results of normit, probit and logit analyses. In the subsequent section of the paper, we consider the issue of integrating the response model into a practical methodology for the research manager.

THE RESPONSE MODEL: NORMIT

During the fifties, Berkson conducted pioneering research in discrete probability models and their estimation. Although his area of application was biometry and medical statistics, his contributions (in particular, logit analysis) have been adapted to problems in other fields such as economics, psychology, and marketing. We describe here one of the discrete probability (quantal response) models which is attributed to Berkson (1955, 1957), called normit analysis. The normit model has two principal attractions:

1. Its estimator, which is asymptotically equivalent to the maximum likelihood estimator, is in the class R.B.A.N. (Regular Best Asymptotically Normal) (Neyman, 1949) with a variance less than that of the maximum likelihood estimator in most cases, and
2. the estimation of the model is extremely simple, not requiring a computer.

MODEL

Let P_i denote the probability of response to some stimulus x at level i . The response function for P_i is given as:

$$P_i = (\text{equation needs to be inserted}) \quad (1)$$

This is the well-known normal frequency function describing an ogive. The normit of P_i is derived from

$$(2) \quad (x_i - \mu) / \sigma$$

by defining

$$\alpha = -\mu/\sigma \text{ and } \beta = 1/\sigma$$

so that the normit of P_i is

$$(3) \quad Y_i = \alpha + \beta x_i$$

The observed relative frequency (maximum likelihood of estimate of P_i) corresponding to x_i is denoted f_i , and the observed normit of f_i is denoted Y_i . The estimation method is based on

the realization that Y_i plotted against x_i should fall along the straight line defined by (3). The estimates of α and β are obtained by minimizing

$$(4) \quad M = \text{(equation needs to be inserted)}$$

Where $Y_i = \alpha + \beta x_i$ is the estimate of (3), n_i is the number of individuals exposed to x_i , and

$$(5) \quad w_i = \text{(equation needs to be inserted)}$$

and summation is over i .

Berkson (1954) has shown that M is asymptotically distributed as X^2 . The estimates of α and β are obtained by a weighted least squares on (3) where α and β are given directly as follows:

$$(6) \quad \beta = \frac{\sum n_i w_i Y_i x_i^2 - \sum n_i w_i Y_i \sum n_i w_i x_i / \sum n_i w_i}{\sum n_i w_i x_i^2 - (\sum n_i w_i x_i)^2 / \sum n_i w_i}$$

$$(7) \quad a = \text{(equation needs to be inserted)}$$

The estimate of the inflection point of the ogive (where $E(P_i) = .5$) is given as:

$$(8) \quad x_{50} = -\alpha/\beta$$

Formulae for standard errors of the estimates are given in the Technical Appendix. For values of $n_i > 2$, Tables are provided in Berkson (1957) for calculating w_i and $w_i Y_i$.

In the next section we describe an empirical application, comparing the normit model with the probit and logit models.

AN EMPIRICAL EXAMPLE

DATA

To evaluate the proposed response model, we developed a questionnaire pertaining to food shopping behavior of French consumers. Wording of the questionnaire and assignment of treatment levels were refined by means of a pilot test conducted on ten French housewives. The

treatment of interest for this study was level of monetary inducement. Ten treatment levels were established on an approximately logarithmic scale as follows:

0 francs	2 francs
.10 franc	5 francs
.20 franc	10 francs
.50 franc	20 francs
1 franc	50 francs

In the interest of maintaining the anonymity of respondents, we decided to affix the monetary incentive to the covering letter enclosed with the questionnaire rather than to promise a monetary reward for returning the questionnaire. The questionnaire was six pages in length, asking an assortment of background questions and involving a sorting task associated with a conjoint analysis exercise. Fifty respondents were included in each treatment cell so that the total sample size was 500.

As it is illegal to send coin and currency through the French postal system, we were obliged to deliver the questionnaires to households. We selected the Fontainebleau-Avon area, a community of approximately 50,000 residents sixty kilometers south of Paris.

The 500 questionnaires were distributed according to a systematic random sampling scheme in which all treatments were randomized to avoid any geographic, income or other biases. The questionnaires were placed in mailboxes during a single day and addresses were noted for eventual follow-up mailings. After two weeks time (the exact date being designated beforehand) returns within each treatment were tabulated as shown in Table 1. A follow-up postcard was mailed, but questionnaires returned later are not considered further in this paper.

A question of some interest to us was whether inclusion of a very small monetary incentive would actually be worse than no incentive. Ten centimes (.10 FF) is worth very little (approximately 1.5 cents in 1989 U.S. terms). However, we found no evidence that people were insulted or "turned-off" by the small values. The hypothesized monotonicity of responses was

observed over practically the entire range of results. Therefore, our quantal response modeling approach was deemed appropriate.

TABLE 1

Responses by Treatment Group

<u>Treatment Group</u>	<u>Treatment Group</u>	<u>Cell Size</u>	<u>Responses</u>	<u>Observed Response Frequency</u>
1	.00	50	8	.16
2	.10	50	9	.18
3	.20	50	11	.22
4	.50	50	12	.24
5	1.00	50	16	.32
6	2.00	50	15	.30
7	5.00	50	20	.40
8	10.00	50	21	.42
9	20.00	50	25	.50
10	50.00	<u>50</u>	<u>31</u>	<u>.62</u>
Totals		500	168	33.6

TABLE 2

Estimates Obtained from Alternative
Response Models

<u>Estimated parameters*</u>	<u>Normit</u>	<u>Probit</u>	<u>Logit</u>
α	-.491 (-8.07)	-.493 (-8.05)	-.816 (-7.82)
β	.157 (6.29)	.158 (6.23)	.167 (6.08)
x ₅₀	23.00	22.59	21.19

*The t-statistics are in parentheses.

TABLE 3

Predictive Ability of Alternative
Response Models

Stimulus Value (Francs)	Observed Relative Frequency	<u>Predicted Response Probability</u>		
		<u>Normit</u>	<u>Probit</u>	<u>Logit</u>
.00	.16	.113	.112	.115
.10	.18	.197	.196	.193
.20	.22	.229	.226	.224
.50	.24	.274	.274	.269
1.00	.32	.312	.311	.307
2.00	.30	.351	.351	.347
5.00	.40	.406	.405	.404
10.00	.42	.448	.448	.449
20.00	.50	.491	.492	.496
50.00	.62	.549	.549	.557
M.A.D.*		.0281	.0276	.0251

*Mean Absolute Deviation between observed relative frequencies and predicted response probabilities.

ANALYSIS

The inducement levels for the experiment were selected purposely so as to be spaced approximately according to a logarithmic scale, because of the Weber-Fechner Law of just-noticeable differences.¹ The latter suggests that responses to sensory stimuli are proportional to the logarithms of the stimuli. Therefore, our estimations of the response functions employed the natural logs of the inducement levels.² We estimated the functions in original inducement level units as well, but obtained considerably poorer fit and we decided to omit them from the paper in the interest of space.³

Values of w_i and $w_i Y_i$ corresponding to the treatment responses were obtained from Berkson's (1957) tables and inserted in equations (6), (7), and (8) to yield the normit function estimates shown in the first column of Table 2. (Note that the n_i dropped out of the equations, because they are equal for all treatments.) For comparison purposes we include also in Table 2 the results of probit and logit estimations obtained with an available computer program (A Q D). All coefficients are highly statistically significant. Very similar coefficient estimates were obtained using all three quantal response models, except for a slightly steeper slope and lower intercept for the logit model. The inflection point of the function occurs at x_{50} . Each of the models estimated this value to be slightly higher than 20FF.

Table 3 shows the relative predictive accuracy of the three estimated models. By this criterion as well, there is very little difference in the models, although the mean absolute deviation measure is slightly better for the logit model.

One can never guarantee that the "right" model of any particular phenomenon is selected. There is no theoretical rationale, for example, for preferring normit over one of the others. We introduced normit because it is not known in the marketing literature and because of the advantages mentioned earlier. Since each model yields approximately the same predictions for response probability, the normit approach can be recommended if one has no access to a computer or to the probit/logit programs.

In order to employ the decision models to follow, the research manager must select one response model that allows prediction of response rate as a function of incentive level. For the above reasons, we will use the normit function in the ensuing discussion.

DECISION MODELS

Once a response function has been estimated, various approaches can be suggested for deciding the level of inducement to use in a particular study. Two basic approaches are diagrammed in Figure 1. It is understood that prior to the inducement level decision various steps have been taken such as determination of the target population, development of questionnaire, and pilot test of the questionnaire with experimentally manipulated incentive levels such as was done in this paper. Response to the inducement levels will depend, of course, on the nature of the target population and the questionnaire itself. The decision models are described in greater detail and illustrated with numerical examples in the following.

Model 1 - Sampling Budget Fixed

Given a sampling budget B (which excludes fixed costs not depending on sample size or level of incentive) and costs per mailed questionnaire c (including, for example, printing, two-way postage, and labor for stuffing and sealing envelopes), it is desired to obtain the optimal level of incentive x^* which yields the largest expected return sample size n_r . That is, we want to select the x that will maximize n_r subject to B and c being given constants. We know by definition that:

$$(9) \quad n_r = n p$$

where p is the response probability for a given x .

Also, it is clear that

$$(10) \quad B = n (c + x)$$

combining (9) and (10) we have

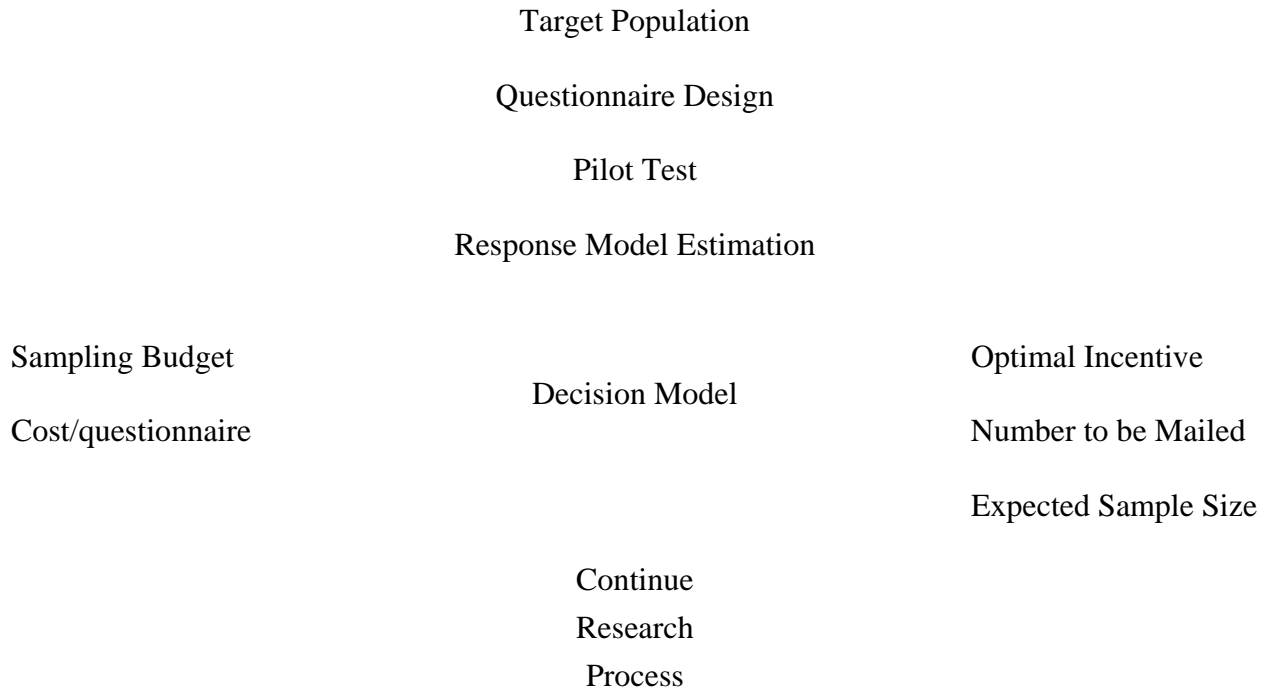
$$(11) \quad n_r = pB/(c + x)$$

FIGURE 1

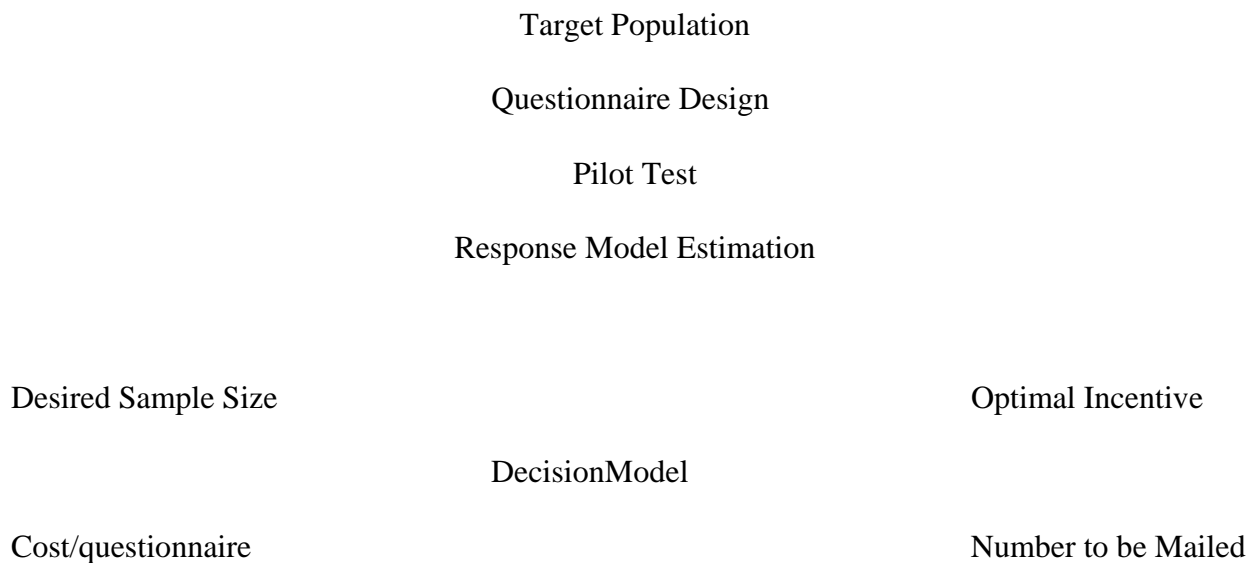
(David and Mac, Figure 1 needs some work. I would be happy to do it if you like. Eva)

Two Approaches to Use of Response del

1. Sampling Budget Fixed



2. Minimum Sample Fixed



Budget Required

Continue
Research
Process

By the earlier discussion, p is a function of x , the latter appearing in the upper limit of the integral of a density function. Standard calculus methods for determining the optimum are intractable, given the complex nature of the objective function. Fortunately, however, there are only a finite number of incentive levels which need to be explored for optimality (it would be silly, for example, to include an incentive with the questionnaire of 37FF or any other odd-sized amount). Therefore, a simple numerical solution is possible in a given practical situation.

As an example, suppose we are given a sampling budget of 10,000FF (French francs) and know the per questionnaire costs are 20FF. For each level of incentive under consideration we can compute $n = 100000 / (20 + x)$ and an expected response probability (from our selected response model). The calculations for a range of incentive levels are shown in Table 4 (using the normit function estimated earlier). We see that the optimal incentive level is 5FF, since it yields the highest expected returned sample size, 162. As secondary outputs, we expect this level of incentive to require a mailed sample of 400 and to achieve a response rate of 40.6%.

Model 2 - Sample Size Fixed

Given a minimum returned sample size n_r and sampling costs per questionnaire c , one's task would be to determine the optimal incentive level x^* which would yield the lowest budget B . That is, we want to determine the x^* that will minimize B subject to n_r and c being given constants.

We know, by combining (9) and (10) that

$$(12) \quad B = (n_r/p)(c + x)$$

Again, standard calculus methods are intractable, but we can easily obtain a simple numerical solution in any particular case. For example, suppose we desire a returned sample of 350 questionnaires and have a cost per questionnaire of 15FF. Calculations for a range of incentive levels are shown in Table 5. In this situation, the optimal incentive is 2FF, since it requires the lowest budget, 16,949FF. As secondary outputs, we determine that the mailed sample size should be 997 and will achieve an anticipated response rate of 35.1%.

TABLE 4

Calculations for Obtaining Optimal Incentive Given
Budget 10,000FF and Sampling Costs 20FF

<u>Incentive</u> <u>Level</u>	<u>p</u>	<u>n</u>	<u>n_r</u>
.00	.113	500	57
.10	.197	497	98
.20	.229	495	113
.50	.274	488	134
1.00	.312	476	149
2.00	.351	455	160
5.00	.406	400	162
10.00	.448	333	149
20.00	.491	250	123
50.00	.549	143	79

TABLE 5

Calculations for Obtaining Optimal Incentive
Given Sample Size 350 and Sampling
Costs 15FF

Incentive <u>Level</u>	<u>p</u>	<u>n</u>	<u>B</u>
.00	.113	3097	46,455
.10	.197	1777	26,833
.20	.229	1528	23,226
.50	.274	1277	19,794
1.00	.312	1122	17,952
2.00	.351	997	16,949
5.00	.406	862	17,240
10.00	.448	781	19,525
20.00	.491	713	24,955
50.00	.549	638	41,470

DISCUSSION

Limitations. This study was performed on a specific target population, so the substantive findings cannot be generalized beyond that population. However, the major purpose of the study was to illustrate a proposed methodology which should be applicable to any study where an S-shaped response to monetary incentive level can be safely assumed.

An important issue that was ignored in this study is the extent to which monetary incentives can influence the nature of responses to mail surveys. That is, monetary incentives have been postulated to cause volunteer biases (i.e., the tendency for the incentive to induce some groups to respond over others) and response bias (i.e., the tendency for the incentive to affect some questionnaire item responses). Both types of biases have been observed (Gelb, 1975; Rush et al., 1978; Nederhof, 1983). As examples of response bias, a study of elites by C,odwin (1979) demonstrated that large monetary incentives substantially improved questionnaire completion levels and caused respondents to give more correct information than when they received lesser incentives.

Potential biases caused by different levels of incentive deserve further investigation. The research manager will need to weigh the likelihood that such biases are significant against the importance of achieving a large (or optimal) returned sample of questionnaires.

Some Statistical Issues. It turned out in this study that we only used one incentive level above the estimated inflection point of the response function. Although it is never known until after the data have been collected whether sufficiently high incentive levels have been included, it is helpful to employ a large range of values so that one would avoid ever having to make out-of-range predictions of response probabilities. This was not a problem for our examples, which yielded optima well within the empirical range of incentives, but we could have constructed examples for which our probability estimates would be considerably less secure.

The quantal response functions have a theoretical range of from negative to positive infinity. However, of course there is no way to give a negative monetary incentive (unless we were to ask respondents to pay to respond). Furthermore, it is possible to postulate that there is

an absolute maximum response probability below 1.0 (Sudman, 1982). By adding two more parameters to the response functions, we could estimate both the probability of response at zero monetary incentive and an upper limit on the response probability. However, the degrees of freedom would be reduced, requiring a larger number of experimental treatments to achieve the same level of statistical precision. We do not believe the additional information provided by these parameters would justify the expense or the additional complexity of estimation.

Another way to extend the model would be to add other independent variables than simply monetary incentive to the response functions. It is possible to experiment with a variety of stimuli during the pretest. Although it would take some effort to extend the normit model to accommodate multiple independent variables, the multivariate logit model would be directly applicable.

CONCLUSION

Low response rates are the critical problem of mail surveys. It has been shown that response rates are monotonically related to level of monetary incentive at least in some instances. We have demonstrated in this paper how one can use that information to decide on the level of monetary incentive to employ in a particular survey. Although we determined that any of several quantal response functions could adequately model the relationship under investigation, we introduced the normit model to the marketing literature as a model with some attractive features, perhaps the most important of which is estimation simplicity.

We believe we have described a practical procedure that might be employed in any mail survey setting. Of course the primary requirement is that one experimentally manipulate the level of monetary incentive during a pretest of the survey. The cost and time required to do this might cause the method to have its greatest appeal to those companies who conduct multiple surveys using essentially the same questionnaire on similar populations over time. On the other hand, we recommend use of the method in any mail survey where the research manager desires an appropriate balance between response and survey budget.

FOOTNOTES

1. Readers interested in the Weber-Fechner Law are referred to Mueller (1965) and Corso (1967). An important marketing reference is Monroe (1971).
2. To avoid taking the logarithm of zero, the value 0.01 was substituted for zero francs in estimating the functions.
3. Although it was not a purpose central to our research, the fact that a better fit was achieved with the incentive level measured in log units provides a partial validation of the Weber-Fechner Law for our data.

TECHNICAL APPENDIX

Standard errors for normal parameters*

(Equation needs to be inserted) (1)

(Equation needs to be inserted) (2)

(Equation needs to be inserted) (3)

where s^2 (Equation needs to be inserted)

*Derived in Berkson (1955).

REFERENCES

- Armstrong, J. Scott (1975), "Monetary Incentives in Mail Surveys," Public Opinion Quarterly, 39, 111-16.
- Berkson, Joseph (1955), "Estimate of the Integrated Normal Curve by Minimum Normit Chi-Square with Particular Reference to Bio-Assay," Journal of the American Statistical Association, 50, 529-49.
- (1957), "Tables for Use in Estimating the Normal Distribution Function by Normit Analysis," Biometrika, 44, 411-35.
- Corso, John F. (1967), The Experimental Psychology of Sensorv Behavior New York: Holt, Rinehart & Winston, Inc.
- Cox, Eli (1976), "A Cost Benefit View of Prepaid Monetary Incentives in Mail Questionnaires," Public Opinion Quarterly, 40, 101-4.
- Gelb, Betsy (1975), "Incentives to Increase Response: Social Class Considerations," Journal of Marketing Research, 12 (February), 109-9.
- Godwin, R. Kenneth (1979), "The Consequences of Large Monetary Incentives in Mail Surveys of Elites," Public Opinion Quarterly, 43, 378-87.
- Gunn, Walter J. and Isabelle N. Rhodes (1981), "Physician Response Rates to a Telephone Survey: Effects of Monetary Incentive Level," Public Opinion Quarterly, 45, 109-15.
- Heberlein, Thomas A. and Robert Baumgartner (1978), "Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature," American Sociological Review, 43, 447-62.
- Kanuk, Leslie and Conrad Berenson (1975), "Mail Surveys and Response Rates: A Literature Review," Journal of Marketing Research, 12 (November), 440-53.
- Monroe, Kent B. (1971), "The Information Content of Prices: A Preliminary Model for Estimating Buyer Response," Management Science, 17, (April), B-519-32.
- Mueller, Conrad (1965), Sensory Psychology, Englewood Cliffs, New Jersey: Prentice-Hall, Inc.
- Nederhof, Anton J. (1983), "The Effects of Material Incentives in Mail Surveys: Two Studies," Public Opinion Quarterly, 47, 103-11.
- Neyman, Jerzy (1949), "Contribution to the Theory of the Chi-Square Test," in J. Neyman, ed., Proceedings of the Berkeley Symposium on Mathematical Statistics and Probabilities, Berkeley and Los Angeles: University of California Press, 23973.

Rush, M.C., J.S. Phillips, and P.E. Panek (1978), "Subject Recruitment Bias: The Paid Volunteer Subject," *Perceptual and Motor Skills*, 47, 443-49.

Schewe, Charles D. and Norman G. Cournoyer (1976), "Prepaid vs. Promised Monetary Incentives to Questionnaire Response: Further Evidence," *Public Opinion Quarterly*, 40, 105-7.

Sudman, Seymour (1982), "Estimating Response to Follow-Ups in Mail Surveys," *Public Opinion Quarterly*, 46, 583-4.

Wotruba, Thomas R. (1966), "Monetary Inducements and Mail Questionnaire Response," *Journal of Marketing Research*, 3 (November), 398-400.