CHAIM M. EHRMAN
KLAUS J. MIESCKE

Structured Decision Rules for Ranking and Selecting Mailing Lists and Creative Packages for Direct Marketing

CHAIM M. EHRMAN received his PhD in marketing from the Wharton School of Business, University of Pennsylvania. He also earned an MBA degree with a marketing concentration from Baruch College, City University of New York, and an MS degree in statistics from Temple University. He is an assistant professor in the marketing department at Loyola University of Chicago. His research interests are in quantitative modelling in marketing and consumer behavior. He has published in several scholarly journals. KLAUS J. MIESCKE is a professor of mathematics and statistics at the University of Illinois at Chicago. He earned his degree of *Dr. Habil.* from the University of Mainz, and a PhD and diploma of mathematics from the University of Heidelberg. His research interests include multiple decision theory, Bayesian analysis, and applied stochastic models. He has published widely and lectured at Purdue University, the University of Mainz, and Academia Sinica in Taipei. Acknowledgement is given to Professor Irwin Gross of Pennsylvania State University, Bernie Silverman of The Signature Group, Professor Charlie King and Paul Ragas of University of Illinois at Chicago, and Roland Priebe of Optimal Services, Chicago, IL. This research was supported by the Air Force Office of Scientific Research, AFOSR Grant #85-0347

ABSTRACT

The field of direct mail advertising is becoming increasingly important. Many selection decisions must be made by direct marketers, such as those concerning package testing and list and segment within list selection. These decisions can be quite complex, especially when sample sizes and average order size per package and list are not equal. In this article, Bayesian and non-Bayesian statistics are applied to these problems to generate optimal decision rules for package testing and list evaluation and selection. An example is given using real data from test results.

CCC 0892-0591/89/01047-13$04.00

INTRODUCTION

Direct mail advertising is increasing in usage and in importance in the area of media selection for advertising campaigns. For example, advertising expenditures (in billions of dollars) were 17.15 for direct mail advertising in 1986. This is an increase of over 124 percent from the expenditures for direct mail in 1980 and compares very favorably to total advertising expenditures, which increased less than 88 percent over the same period (3, 11). Direct mail is clearly increasing in prominence for advertisers. It has experienced the highest relative percentage increase in dollar expenditure over the 1980–1986 period compared to all other media advertising expenditures.

In the following table, we indicate the relative increase in media advertising expenditures for 1970, 1980, 1986.

Medium	1960–1970 % Increase in Spending	1970–1980 % Increase in Spending	1980–1986 % Increase in Spending
Newspaper	154	274	176
Magazines	142	250	175
Television	221	315	203
Radio	190	282	189
Direct mail	151	276	224
Outdoor	115	265	169

From: *Advertising Age*, (3, 11).

A recent article in the Wall Street Journal (17) gives further evidence of the increasingly important role direct marketing has in advertising:

> Many advertisers are shifting their advertising budgets to the direct medium. The structure of advertising is starting to change. Segmented advertising is the way of the future . . . Advertisers also cite economic reasons for the shift. Low inflation is holding down many ad budgets but TV advertising prices have kept climbing even though audience share has dwindled. A 30-second network TV-spot can cost more than $300,000. Since 1977, newspapers and magazines have more than doubled their charges for reaching 1,000 readers. Total ad pages in magazines were up in 1985, but most mass periodicals posted declines . . . Direct marketing is the fastest-growing related area and the one best suited for specific audiences.

We see that although total revenues for TV advertising are increasing, the trend is toward increased usage of direct marketing as a cost-effective and well segmented promotional medium.

In direct marketing, mailing lists are the primary vehicle for dissemination of promotional messages. Assuming a direct marketing effort, decision makers face evaluation and selection decisions, such as: Which mailing list should be purchased for mass mailings (we use the term "rollout" instead of mass mailing from now on) and which clusters within chosen lists should be most heavily promoted?

1. Problem Description

Decision makers for list selection typically mandate an initial test mailing for all competing lists under consideration. Many mail order houses offer list buyers the option to test a given sample size for a promotional item to determine its response rate before purchasing the entire list. The sample response rates now become the basis for the selection decision. If a test response rate is greater than an acceptable minimum response threshold (perhaps as low as .002), a larger sample from the list will be mailed again. If not, the list will be discarded. Once the "go" decision is given, a rollout is undertaken on the population defined by the entire list or a remail procedure to a larger sample takes place.

We address two important questions that direct marketers face. First, how many creative packages should be generated and tested before selecting the best one? There is a tradeoff between incremental costs of developing creative packages and incremental revenues realized from these creatives. Gross (12) addressed this question in an unpublished dissertation. He did publish his methodology in an advertising framework (13), but not in a direct marketing context. We initially review his work and subsequently present our approach. Second, how should one compare different lists when test sample sizes and revenues per response are different?

The organization of this article is as follows. In Section 2, we present a mathematical foundation based on Gross (12). This is used for solving the optimal number of creative packages to be generated prior to testing. Section 3 presents an alternative methodology for solving the number of kits to be created; Section 4 focuses on the related problem of list selection given unequal sample sizes and response revenues per list, and Section 5 presents solution methodology with an illustration using real data. Section 6 gives concluding remarks.

2. Subset Size Decision

An important question decision makers face is the subset size problem, namely, how many creative packages should be developed and tested. Clearly, as we continue to develop more and more creative packages, the likelihood of finding a "winner" increases. On the other hand, the cumulative developmental costs can become prohibitive. Below, a method of Gross and our alternative are discussed, in which the sample sizes $N_1, \ldots, N_n$ are equal, and the rollout sizes $M_1, \ldots, M_n$ are equal as well. (The subscript n represents the number of creative packages we develop and test.)

2.1 LITERATURE REVIEW. The subset size problem was addressed by Gross in a direct marketing context (12) and published in an advertising setting (13). Recent texts in advertising have devoted sections of chapters to explain his works and how to implement his technique. [See (1), for instance.] Although his framework was in the area of advertising campaign generation, we focus here on the direct marketing example (12). He assumes that each direct marketing campaign has n possible creative alternatives. He provides methodology to determine the optimal number of creative alternatives to develop and test for a direct marketing campaign before selecting the best. His testing method for direct marketing campaign effectiveness is also based on sample response data.

2.2 SYNOPSIS OF GROSS' MODEL. Gross (12) assumes the following: Each promotional item that is mailed can generate a sale (response) or non sale (no response) per respondent. This scenario has characteristics of a Bernoulli trial, where, for example, each toss of a coin can result in a win or a loss. There is a single parameter (θ) that characterizes, for all possible tosses of a given coin, the chances of a win on any given toss. A marketer using direct mail is ultimately interested in aggregate sales or total number of responses of the rollout effort. The sum of n Bernoulli trials, i.e., the number of wins for n tosses, is characterized by the binomial distribution, which, in turn, can be approximated by the normal distribution as n gets large. Each creative package has been pretested to generate a mean response rate, and we want to know how many themes should be tested. His data are given in Table 1, and

TABLE 1
Summary of Mail Response Data [Gross (12)] for Testing 12 Different Themes

Theme #	Response Data
A	.0240
B	.0233
C	.0212
D	.0211
E	.0210
F	.0207
G	.0205
H	.0190
I	.0188
J	.0185
K	.0167
L	.0156

N = 50,000, Average Response Rate = .02, Standard Deviation = .00245.

we briefly quote his basic premise and assumptions [(12), p. 36].

> One may think of the process of creating a number of different advertising campaigns as that of drawing a sample of n from a probability distribution of relative effectiveness. Any one of the n alternatives might be chosen as the reference and the relative effectiveness of the others would be defined with respect to that one. Hence, if one imagined a very large number of alternatives to have been generated, one might conceive of a probability distribution of relative effectiveness with the zero point chosen arbitrarily as the profitability of one of the alternatives. Rather than choosing the zero arbitrarily, it shall be set at the mean of the probability distribution.

We present the mathematics used by Gross to derive his decision rule. Illustration of his method is given in Section 2.3.

The variables defined by Gross are the following:

μ_o = The overall mean response that the given package should generate, based on the product being marketed, regardless of the creative used.

T_j = The deviation from the overall mean that package j generates because the creativity on the appeal (can be an asset or liability).

t_{ij} = The error associated with the ith test on the

jth creative package. The assumption is that each of the j packages can be tested i times, with $i \geq 1$.

O_{ij} = The Observed response on the ith test for the jth package.

Θ_j = The true parameter, what the true response for package j should be, assuming an error-free environment.

E_j = The Expected incremental profit that is realized by developing the next jth package.

W_j = The deviation in incremental revenues from the overall mean (μ_E) that is realized from package j.

μ_E = The overall average incremental profits to be realized by generating another creative package.

Assuming an unbiased relationship, the expected value of T_j, W_j, and t_{ij} are all zero, and their respective standard deviations are $\sigma_T, \sigma_E, \sigma_t$. Reliability of the ith test is simply $\sigma_T/(\sigma_T^2 + \sigma_t^2)^{1/2}$.

Using a normal model and the Bayes approach, Gross defined the relationships as follows.

$$O_{ij} = \mu_o + T_j + t_{ij} \quad \text{with} \quad T \sim N(0, \sigma_T^2),\ t \sim N(0, \sigma_t^2), \text{ independent,} \tag{1}$$

$$\Theta_j = \mu_o + T_j, \quad \text{and}$$

$$E_j = \mu_E + W_j \quad \text{and} \quad W \sim N(0, \sigma_E^2).$$

The T_j variable is the true deviation from μ_o for the jth campaign, and t_{ij} is the measurement error for the ith test on the jth campaign.

The variables Θ and E are bivariate normal, $(\Theta, E) \sim \text{BVN}\ (\mu_o, \mu_E, \sigma_T^2, \sigma_E^2, \rho)$.

The reliability of a test is measured by the proportion of total variance explained by the test. Gross defines R to measure this reliability,

$$R = [1 - \sigma_t^2/(\sigma_T^2 + \sigma_t^2)]^{1/2},$$

which can be written $(1 + \sigma_t^2/\sigma_T^2)^{-1/2}$. (2)

The decision maker will ultimately select that one of n alternatives which yields $O_{[n]}$, say, the largest of the *observed* sample means O_j, $j = 1, \ldots, n$. Let $\Theta_{(n)}$ and $E_{(n)}$ denote the response parameter and profitability, respectively, of the selected campaign (which corresponds to the largest observed mean response). A larger n results in a larger expected $E_{(n)}$, but the costs are also increasing in n. Gross' solution for the optimum number of creative themes to be generated, or n, is derived as follows. Since

$$E(\Theta_{(n)} \mid O_{[n]}) = (O_{[n]}\sigma_T^2 + \mu_o\sigma_t^2)/(\sigma_T^2 + \sigma_t^2) \tag{3}$$

is linear in $O_{[n]}$, and

$$E(O_{[n]}) = \mu_o + e_n(\sigma_T^2 + \sigma_t^2)^{1/2}, \tag{4}$$

where e_n denotes the expected maximum of n independent standard normally distributed random variables, substitution of eq. 4 into 3 leads to

$$E(\Theta_{(n)}) = \mu_o + e_n\sigma_T R. \tag{5}$$

In the same way, it is seen that

$$E(E_{(n)}) \mid \Theta_{(n)}) = \mu_E + \rho\sigma_E(\Theta_{(n)} - \mu_o)/\sigma_T \tag{6}$$

is linear in $\Theta_{(n)}$. Thus, substitution of eq. 5 into eq. 6, and assuming μ_E to be zero, the expected profitability of the best of n tested campaigns turns out to be

$$E(E_{(n)}) = e_n\sigma_E\rho R. \tag{7}$$

In a last step, this has to be adjusted for costs. Let C_a be the costs of generating a new alternative, let C_b be the costs of testing (including sampling), and c be the costs of mailing. Then, if C_n denotes total aggregate costs of n alternatives, the profit function turns out to be

$$E(E_{(n)}) - C_n = e_n\sigma_E\rho R - n[C_a + N(c + C_b)]. \tag{8}$$

In estimating the reliability of the test, R, Gross estimates σ_t^2 as $E(\Theta_{(n)})[1 - E(\Theta_{(n)})]/N$, and $\mu_o + e_n\sigma_T$ is an estimate for $E(\Theta_{(n)})$. By substituting $Y^2 = E(\Theta_{(n)})[1 - E(\Theta_{(n)})]/\sigma_T^2$, he rewrites $R = (1 + Y^2/n)^{-1/2}$. The profit function (eq. 8) is now both increasing in N, through R, and decreasing in N, through costs. Gross takes partial derivatives with respect to N, sets it equal to zero and solves for the optimal N (for a given fixed n). He then solves for n once an optimal N has been found. Optimal N is used in eq. 8. The value for e_n and e_{n+1} are given in Table 2.

The *decision rule* is to increase n to $n + 1$ as long as the expected profitability increases by a larger amount than the total aggregate costs or, equivalently, as long as the following holds:

$$e_{n+1} - e_n > [C_a + N(c + C_b)]/\sigma_E \rho R. \qquad (9)$$

2.3 ILLUSTRATION ON USING GROSS' TECHNIQUE. The following example shows how the optimum n is found for Gross' data.

$N = 50{,}000$, actual sample size
$\mu_o = .02$
$\sigma_t^2 = (\mu_O)(1 - \mu_O)/N = 3.9 \times 10^{-7}$
$\sigma_T^2 = 60.3 \times 10^{-7}$
$\sigma_E = M \cdot g \cdot \sigma_T = 16.6 \times 10^4$
Optimal $N = 72{,}000$
$R = .972$
$M = 3.5 \times 10^7$, rollout size
g = earnings per response = 2.00
$C_a = 5{,}000$.
$c = .10$ (1961 data)
$C_b = .02$
$\rho = 1$ (mail response rates generate both test data and profits, so correlation is 1.)

Substituting these values into the decision rule (eq. 9), the subset size should be expanded from n to $n + 1$ as long as

$$e_{n+1} - e_n > 0.0845.$$

Table 2 shows that the highest value for n that satisfies eq. 9, is $n = 7$. This is Gross' method of solving for optimal n in a direct marketing context.

TABLE 2
Incremental Gains in Expectation for Maximum Standard Normals as Subset Size Changes from $n \rightarrow n + 1$

n	e_n	$e_{n+1} - e_n$
2	0.564	0.564
3	0.846	0.282
4	1.029	0.183
5	1.163	0.134
6	1.267	0.104
7	1.352	0.085
8	1.424	0.072
9	1.485	0.061
10	1.539	0.054
11	1.586	0.047
12	1.629	0.043
13	1.668	0.039
14	1.703	0.035
15	1.736	0.033
16	1.766	0.030
17	1.794	0.028
18	1.820	0.026
19	1.844	0.024
20	1.867	0.023

Source: (16).

2.4 SUBSET SIZE PROBLEMS IN CREATING PACKAGES FOR DIRECT MARKETING. Gross' (13) work addressed the subset size problem in an advertising framework. Clearly, n independent themes can be used to produce n separate advertising campaigns about a specific product. Each campaign is compared to drawing, with replacement, balls from an urn. There is a common response parameter that affects the success likelihood for each campaign. The campaigns are independent and identically distributed. His direct marketing data (12), which tested different promotional themes for a magazine subscription, are also characterized by one common response parameter. His approach has a drawback in that he fits initially a binomial distribution to a normal distribution, which can result in a loss in accuracy.

In the following section, we introduce an alternative methodology to address the subset size problem. We also employ the Bayesian approach. However, our model is different. Rather than approximating the binomial distributions from the very *beginning* by normal distributions and using a prior distribution that is normal, we *combine* the binomial distribution with a beta distribution as our prior. Only at the end, we use a normal approxi-

TABLE 3
Probability that a Rollout Response Rate Will Exceed a Cutoff Value c, Given Sample Size n and Response Rate p

	—c = .001—				—c = .002—				—c = .003—				—c = .004			
p	n = 5	10	20	40	n = 5	10	20	40	n = 5	10	20	40	n = 5	10	20	40
.0010	.50	.50	.50	.50												
.0012	.65	.71	.79	.87												
.0014	.77	.85	.93	.98												
.0016	.86	.93	.98	.99												
.0018	.90	.97	.99	.99												
.0020	.94	.98	.99	.99	.50	.50	.50	.50								
.0022	.96	.99	.99	.99	.61	.66	.72	.80								
.0024	.97	.99	.99	.99	.71	.79	.87	.94								
.0026	.98	.99	.99	.99	.79	.88	95	.99								
.0028	99	.99	99	.99	.85	.93	.98	.99								
.0030	.99	.99	.99	.99	.90	96	.99	.99	.50	.50	.50	.50				
.0032	.99	.99	.99	99	93	98	.99	.99	.59	.63	.69	.76				
.0034	.99	.99	.99	.99	.95	.99	99	.99	.68	.75	.83	.91				
.0036	.99	.99	99	.99	.97	.99	99	.99	.76	.84	.92	.97				
.0038	.99	.99	.99	.99	.98	99	99	.99	.82	.90	.96	.99				
.0040	.99	99	99	.99	.98	99	99	.99	.86	.94	.98	.99	.50	.50	.50	.50
.0042	99	.99	.99	.99	.99	99	.99	.99	.90	.96	.99	99	.58	.62	.66	.73
.0044	.99	.99	.99	.99	.99	99	.99	.99	.93	.98	.99	.99	.66	.72	.80	.88
.0046	.99	99	.99	.99	.99	99	.99	.99	.95	.99	.99	99	.73	.81	.89	96
.0048	.99	99	.99	99	.99	99	.99	.99	.96	.99	.99	.99	.79	.87	.94	.98
.0050	99	.99	99	.99	.99	99	99	.99	.97	.99	99	.99	.84	92	.97	99

mation to the beta posterior to arrive at a decision criterion, which is easy to use and reliable in its performance. The latter is also due to the fact that our rule requires only the determination of the prior *mean* and *variance,* which can be achieved satisfactorily by using past data. We also keep N as fixed, without solving for optimal N. A comparison of our decision criterion (eq. 15) with the one by Gross (eq. 9) is made at the end of Section 3.

3. Using Bayesian Statistics to Solve the Subset Size Problem

3.1 INTRODUCTION TO EMPIRICAL BAYES STATISTICS. In the classical framework of statistical decisions, observations follow distributions which are known, except for the values of specific parameters. The Bayesian approach considers these parameters to be chosen randomly, according to some known prior distribution, before observations are drawn. Combining the conditional distribution of the observations, given the parameters, with the prior distribution, allows for determining the posterior distribution—i.e. the conditional distribution of the parameters, given the observations, from which Bayesian decisions can be derived.

In our context, we are faced with a loss function. Determination of the loss function, whose posterior expectation has to be minimized, is usually, as in the present situation, not a difficult task. More challenging is the proper choice of a prior, which is often made by first narrowing it down to a large and flexible parametric family of distributions and then by choosing the appropriate family parameters. If the latter can be estimated from past data, which were generated by the same model, then it is called the empirical Bayes approach, which is described in detail in DeGroot (6) and in Berger (4). Application of Bayes Statistics in a direct marketing framework is given by Ehrman (10). This approach will be applied to binomial distributions with a beta prior.

3.2 SOLUTION METHODOLOGY FOR OPTIMAL SUBSET SIZE—GIVEN EQUAL SAMPLE SIZE AND ROLLOUT SIZE PER LIST. Before we derive the criterion for solving the subset size problem, we address, in a Bayes context, the distributions of random variables. In the prior distribution, let the response rate parameter for the ith creative package, Θ_i be independent, identically beta-distributed, $Be(\alpha, \beta)$, $i = 1, \ldots, n$. Let W_i represent the *observed* aggregate responses, and be independent binomial, $W_i \sim B(N, \theta_i)$, and the *observed response rate* be X_i, $X_i = W_i/N$, at $\Theta_i = \theta_i$, $i = 1, \ldots, n$. We assume the decision rule: Select the jth list if and only if $X_j = \max\{X_1, \ldots, X_n\}$. The rollout total response results, V_i, are also binomial, $V_i \sim B(M, \theta_i)$ with $N \ll M$, and the *rollout* response rates are Y_i, $Y_i = V_i/M$. Define $E_n(\Theta^*)$ as the expected value of the rollout response rate associated with creative package that generates the highest observed response rate $X_{[n]}$. The term Θ^* is the response rate parameter that is associated with creative package with the highest observed response rate. The overall net expected revenue from mass mailing is:

Expected Profit

$$= M[E_n(\Theta^*)g - c] - n[C_a + N(C_b + c)]. \quad (10)$$

First we have to find $E_n(\Theta^*)$, where Θ^* is the randomly selected Θ_j with j determined by $X_j = X_{[n]}$, the highest X value of a subset size n, and Θ^* is the true response rate for the list selected, i.e., the list with the maximum response rate or highest X-value.

We know that the conditional distribution of Θ_i is: $\Theta_i \mid (X = x) \sim Be(\alpha + w_i, \beta + N - w_i)$, $i = 1, \ldots, n$, independent, and that the conditional distribution of Θ^* is $\Theta^* \mid (X = x) \sim Be(\alpha + w_{[n]}, \beta + N - w_{[n]})$, with $E\{\Theta^* \mid X = x\} = (\alpha + w_{[n]})/(\alpha + \beta + N)$. When we take expectations from both sides, we get,

$$E_n(\Theta^*) = (\alpha + E(NX_{[n]}))/(\alpha + \beta + N). \quad (11)$$

Since $E(NX_{[n]}) = NE(E\{X_{[n]} \mid \Theta\})$, and N is large, this is approximately $NE(\Theta_{[n]})$. The expected maximum of a sample from a beta distribution $Be(\alpha, \beta)$ is very cumbersome to compute. However, Johnson and Kotz (14) point out that when α/β is fixed and α and β are large, a normal approximation exists. Since our mean is fixed at about 0.02, and the standard deviation is small, we use this normal approximation for the beta-prior. Recall that $e_n = E(Z_{[n]})$, where $Z_{[n]}$ is the maximum of $Z_1, \ldots, Z_n$, a sample of standard normals with mean equal zero and variance equal 1, $N(0, 1)$. Let us approximate eq. 11 by replacing $E(X_{[n]})$ with $m + \tau e_n$, where m is the mean, and τ is the standard deviation for the prior beta distribution. We rewrite eq. 11 as $E_n(\Theta^*) \approx \{\alpha + N(m + \tau \cdot e_n)\}/\{\alpha + \beta + N\}$. Recall, a beta distribution $Be(\alpha, \beta)$ has the following summary measures: mean $m = \alpha/(\alpha + \beta)$ and variance $\tau^2 = \alpha\beta/[(\alpha + \beta)^2(\alpha + \beta + 1)]$ and by conversion substitution, $\alpha = m\{[m(1 - m)/\tau^2] - 1\}$, $\beta = (1 - m)\{[m(1 - m)/\tau^2] - 1\}$. The sum $\alpha + \beta$ is seen to be $[m(1 - m)/\tau^2] - 1$. We substitute α and β and get:

$$\begin{aligned} E_n(\Theta^*) &\approx [m\{[m(1 - m)/\tau^2] - 1\} + N(m + \tau \cdot e_n)]/[\{m(1 - m)/\tau^2\} - 1 + N] \\ &= [m^2(1 - m) - m\tau^2 + N\tau^2(m + \tau e_n)]/[m(1 - m) - \tau^2 + N\tau^2] \\ &= m + \{N \cdot \tau^3/[m(1 - m) + (N - 1)\tau^2]\}e_n. \end{aligned}$$

We have shown that for subset size n, using the beta-binomial conjugate model, we have approximately

$$E_n(\Theta^*) \approx m + \{N \cdot \tau^3/[m(1 - m) + (N - 1)\tau^2]\}e_n, \quad n = 1, 2, \cdots \quad (12)$$

Now we can rewrite eq. 10 approximately as follows:

$$M[m \cdot g - c + \{N \cdot \tau^3 g/[m(1 - m) + (N - 1)\tau^2]\} \cdot e_n] - n[C_a + N(C_b + c)]. \quad (13)$$

The incremental net gain in our campaign, as we expand the subset size from n to $n + 1$, is predicted to be positive as long as the following holds:

$$e_{n+1} - e_n > [C_a + N(C_b + c)][m(1 - m) + (N - 1)\tau^2]/MN\tau^3 g. \quad (14)$$

We can simplify eq. 14 even further, based on the following fact.

$$[C_a + N(C_b + c)]\cdot[m(1 - m) + (N - 1)\tau^2]/MN\tau^3 g = (N/M\cdot\tau)\{[C_a/N + C_b + c]/g\} \times [1 + m(1 - m)/N\cdot\tau^2 - 1/N].$$

If we drop m^2 and $1/N$ in the second bracket, we get the simple *decision rule* for subset size as follows. Increase the subset size from n to $n + 1$ as long as

$$e_{n+1} - e_n > (N/M\cdot\tau)\{[C_a/N + C_b + c]/g\} \times \{1 + m/(N\cdot\tau^2)]\} \quad (15)$$

otherwise, keep the subset size at n.

Thus, if the expected incremental revenues to be realized by producing another creative package *exceed* the developmental cost, the direct marketer should produce another package. If not, then testing and selection should begin. The formula used to assess expected incremental values are given in eq. 15. We present an illustration in the next subsection.

3.3 ILLUSTRATION USING OUR ALTERNATIVE TO GROSS. We present an illustration to show how the Bayes decision rule can be used with actual empirical data. For the purpose of comparison we use the same data of Gross (12), presented in Table 1.

Gross assumes the following:

$M = 3.5 \times 10^7$
$m = .02$
$\tau = .00245$
$g = 2.00$
$N = 5 \times 10^4$
$c = .10$
$C_a = 5 \times 10^3$
$C_b = .02$

We use the values in Table 1 to determine α and β for the prior $Be(\alpha, \beta)$. As mentioned earlier, this approach is a version of Empirical Bayes, where the prior is characterized by empirical observation. The average score and variance are used to estimate the prior mean and variance, respectively.

We can solve for the optimum subset size by applying selection criterion 15. When we substitute the given values for the variables in criterion 15, we find that the subset size should be expanded from n to $n + 1$, as long as

$$e_{n+1} - e_n > (0.58309)(0.11)(1.06664) = .0684.$$

We find the solution for n by referring to Table 2. The largest value for n that allows $e_{n+1} - e_n$ to exceed .0684 is 8. Hence, in the given situation, the optimum subset size, i.e., the number of creative packages to develop and test is 8. It is interesting to note that Gross, using the bivariate normal model, found $n = 7$ in this particular case.[1]

3.4 COMPARISON OF TECHNIQUES. Using our technique these data indicate that the direct marketer would be well advised to generate 8 creative packages before testing and selecting the best package. Gross claims that 7 is optimal.

A comparison of our decision rule with the one by Gross can now be made by comparing eq. 15 with eq. 9. What is needed for determining eq. 15 is to find m and τ^2, i.e., the mean and variance of the beta prior. As we have shown, this is usually not a difficult task since past data can be used for this purpose. Gross' rule requires *more parameters* to determine optimal subset size which may be harder to estimate. Moreover, Gross employs a normal approximation to the binomial model from the very beginning. We believe that our decision rule is a suitable and possibly more accurate alternative. We require *fewer parameters* and maintain the beta-binomial distribution. Furthermore, Gross assumes a tradeoff between alternatives n and sample size N. By increasing N, he is limiting n, the number of alternatives for testing. Our methodology keeps N

[1] If Gross would omit solving for optimal N, our methods converge on 7. However, this phenomenon is a result of the particular data presented and different data would produce divergence.

fixed at a large number, and allows the decision maker to use more resources for testing more creative packages. We now turn our attention to list selection. We address the list selection problem relaxing the assumption of equal response revenue and equal sample size across competing lists.

4. Mathematical Foundation for the List Selection Problem—Unequal Response Revenues and Sample Sizes and Rollout Size per List

Prior to a list purchase, sample response rates are compared against a fixed minimum response threshold for a list or cluster purchase decision. We use the word "list" in this discussion but we also mean clusters within lists. Many direct marketers have a "bench mark" figure that a list must meet and surpass to be considered good. However, this decision rule to use a *fixed* minimum response rate may be inappropriate. This is especially true when the revenue per response is a *variable* dollar amount, such as magazine subscriptions, premiums for insurance, or purchases of apparel. For instance, if a given list targets a segment that has a higher-than-average dollar return per response, a "go" decision may be appropriate even when the response rate is below the fixed minimum threshold. A higher-than-average contribution to profits per response may make a list profitable even though the response rate is below the minimum threshold. A simple solution is to vary the minimum response rate with dollar revenues per order, order based on average order size. However, this solution works only if sample size across clusters or lists is identical.

Test sample sizes may not be equal across different lists or clusters. Therefore, sample response rates can be misleading when they alone are used to evaluate performance of different mailing lists. A list that generates a high sample response rate from a small sample does *not* indicate necessarily greater rollout response compared to other lists that generate lower sample response rates, but were tested with a much larger sample size. A *small* sample induces a *high* sampling error that may reduce the chance that the rollout response will be high. An alternative list that was tested on a larger sample size has lower sampling error and can have a *higher* chance that the rollout response rate will be adequate.

The issue of different sample sizes and their effects on mailing lists' response rates has been addressed in (8). He suggested that for each list, the decision maker should compute the *likelihood* of exceeding the threshold value given the sample size and response rate. The ranking and selection of lists should be based on the highest probability of exceeding the minimum threshold. We present the probabilities of rollout success, based on given response rates, cut off values, and sample size in Table 3.

This technique works when two prerequisites are met: The cutoff values are given and the revenue per response is equal. In the real world of list testing, many selection decisions entail unequal sample size and unequal revenues per response. Therefore, we present new methodology to address the list selection problem, relaxing the assumptions of equal revenues per response, equal rollout population, and a predetermined cut off value.

Assume that there are n possible lists available for purchase. We now use the letter n to represent the number of lists available for purchase. Every ith list has N_i subscribers and a response parameter θ_i representing the *true* rollout response rate. Let

- X_i be the sample response rate for the ith list,
- Y_i be the rollout response rate for the ith list,
- G be the minimum required dollar gain to make a list profitable,
- D_i be the minimum response rate that makes the ith list profitable,
- g_i be the average revenue per response for the ith list,
- N_i be the sample size for the ith list,
- M_i be the rollout size for the ith list,
- c be the cost per item mailed (postage, printing, etc.).

The variable Y_i has $E(Y_i) = \theta_i$ and Var $(Y_i) = \theta_i(1 - \theta_i)/M_i$. The decision rule for any list purchase would be the following: If expected revenues exceed cost, purchase the list. Mathematically, it is as follows.

If θ_i were known:

$$\text{If} \quad M_i[\theta_i g_i - c] \geq G, \quad \text{purchase list, otherwise, do not purchase.} \qquad (16)$$

Equivalently, in terms of θ_i,

$$\text{If} \quad \theta_i \geq (G/M_i + c)/g_i = D_i, \quad \text{purchase list, otherwise, do not purchase.}$$

We note that Y_i will be the realized response rate for the ith list, and θ_i is only the parameter, whose value may never be realized. Therefore, we modify eq. 16 in terms of Y_i, the response rate to be realized once a mass mailing takes place.

$$\text{If} \quad Y_i \geq (G/M_i + c)/g_i = D_i, \quad \text{purchase list, otherwise, do not purchase.} \quad (17)$$

We compute the probabilities that Y_i will be less than D_i using a normal approximation. Suppressing the i subscript for easier notation, we have as an estimate

$$Pr(Y < D) \approx \Phi[(D - \theta)\{\theta(1 - \theta)/M\}^{-1/2}], \quad \text{where } \Phi \text{ denotes the } N(0, 1) \text{ c.d.f.} \quad (18)$$

We may replace θ by X, our observed response rate for the ith item, to get

$$Pr(Y < D) \approx \Phi[(D - X)\{X(1 - X)/M\}^{-1/2}] \quad (19)$$

It is known that if we define $q(\theta) = \Phi[(D - \theta)\{\theta(1 - \theta)/M\}^{-1/2}]$, and if we estimate $q(\theta)$ by $q(X)$, then $q(X)$ is asymptotically normal with mean and variance

$$E(q(X)) \approx q(\theta), \quad \text{and}$$
$$\mathrm{Var}\,(q(X)) \approx \{\theta(1 - \theta)/N\}[q'(\theta)]^2.$$

[The derivation for the asymptotic mean and variance is based on the Delta method (5).] We can further approximate the variance by substituting X for θ. Let $\varphi(z) = (2\pi)^{-1/2} \exp -(z^2/2)$ denote the $N(0, 1)$ density. Then

$$\mathrm{Var}\,(q(X)) \approx [X(1 - X)/N][q'(X)]^2, \quad \text{where}$$
$$q'(y) = \varphi([D - y][y(1 - y)/M]^{-1/2})[M]^{1/2} \times d\{(D - y)[y(1 - y)]^{-1/2}\}/dy. \quad (20)$$

Our confidence interval for $q(\theta)$ is therefore the following:

$$q(\theta) \approx q(X) \pm z_{\alpha/2}[X(1 - X)/N]^{1/2} q'(X), \quad \text{where} \quad z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2) \quad (21)$$

We can use eq. 21 to estimate the probability of failure, i.e., the chances that the rollout response rate is below the required threshold, or simply that $Y < D$. For one-sided confidence statements, $\alpha/2$ can be replaced by α in eq. 21. Any ith list which has a large probability of rollout response less than D, is rejected from a mass mailing. This approach can give us a decision rule whether to go ahead or not. However, these equations can be cumbersome and are difficult to solve. In the next section we present a *simplified* version to address the list selection problem.

4.1 SOLVING FOR CUTOFF VALUES. Our goal is to derive a *simpler test* for go/no-go decisions. We suppress the subscript i and focus on a decision rule for selection. Define α as the maximum allowable error margin for the decision maker. In this context we choose both $\alpha = .05$ and $\alpha = .01$, with z_α being the corresponding z-value for a given α, derived from the standard normal distribution.

Recall from eq. 17 that we are using $D = (G/M + c)/g$ as a threshold for θ. Since the expected return is evaluated by D, the minimum accepted rate of return, we compute from our sample data a minimum Z score using D as the expected value, $Z = (X - D)/[D(1 - D)/N]^{1/2}$. If Z is below the stipulated value z_α that can be readily found in the standard normal table, then we reject the list from the rollout mailing. We can write the Z score as $Z = (X - D)/[D(1 - D)/N]^{1/2}$. The decision rule is defined in the following way. If:

$$Z < z_\alpha, \quad \text{reject the list. Equivalently, in terms of } X, \quad \text{if } X < z_\alpha[D(1 - D)/N]^{1/2} + D, \quad \text{reject the list,} \quad (22)$$

with $D = (G + M \cdot c)/(g \cdot M)$. Otherwise we should purchase the list. We present in the next section an illustration using this rule to determine cutoff values for list selection.

4.2 ILLUSTRATION FOR THE LIST SELECTION DECISION RULE. We present an illustration to show how to use this decision rule. The purpose of this illustration is to facilitate comprehension of the decision rule to determine the go/no-go decision. Assume the following values: Let $g = 100$, $c = 0.10$, $G = 10^5$, $M = 10^5$, $D = 0.011$. The decision rule is now based on the sample response and size.

Whenever $N = 100$, the cutoff value is .028 for $\alpha = .05$, and .035 for $\alpha = .01$. We assume a 95 percent and 99 percent confidence limit that our rollout response rate will exceed a cutoff score, and the α values correspond to the respective error terms. If $N = 1000$, then the cutoff value is .016 for $\alpha = .05$ and .019 for $\alpha = .01$. As $N \rightarrow \infty$, the cutoff value is .011. In Section 5 we demonstrate a list selection problem using *actual* current data.

5. List Selection Problem Illustrated with Actual Data

There are many firms who provide mailing lists and the ability to select names by cluster for a modest fee. In this example, we use clusters generated by the Claritas Company. They use the Prizm system to cluster individuals based on lifestyle segments. Each cluster consists of mutually exclusive clusters of individuals and do not overlap.

A particular firm, prominent in direct marketing had to decide, for selling a given service, which clusters across a number of lists to select for rollout. A total of 9 clusters were tested. In Table 4 we present the response rates, average revenues and rollout size per cluster. Please note that average dollar per response, sample size, and rollout population are not constant. The costs of mailing are also given.

Table 4 contains some interesting phenomena. First, the cluster with the highest dollar revenue per response, #13, is not very likely to succeed. Second, although clusters #14 and #37 have identical response rates, cluster #14 has a higher chance of success, since the sample size N is over three times that of cluster #37. We see that sample size, average dollar per response and response rate all contribute to the Z score per cluster that determines *the probability of success.* It is clearly inadequate to base the selection criterion on sample response rates alone, as is currently done by many direct marketing firms. The clusters from which names should be purchased are #15 and #31, which by coincidence have here the largest two response rates, too. However, #31, the winner in terms of z-values, turns out to be the cluster with the second-largest response rate.

The managerial implications of this methodology

TABLE 4
Summary Test Response Data

List #	M	N	g ($)	X	D^*	Z^{**}
1. 15	809,743	20,403	183	.0023	.00150	2.95
2. 31	1,041,234	44,156	188	.0020	.00146	2.97
3. 14	1,777,568	52,225	193	.0016	.00142	1.09
4. 37	669,601	15,991	187	.0016	.00147	0.43
5. 9	1,282,705	32,697	184	.0015	.00149	0.05
6. 13	2,405,997	25,104	230	.0012	.00120	0.00
7. 6	828,848	17,143	202	.0013	.00136	negative
8. 32	1,990,129	49,497	181	.0014	.00152	negative
9. 11	1,937,988	29,035	200	.0011	.00138	negative

* We assume a conservative lower bound for c and G. c = $225 per (000) printing, bulk mail postage + $25 per (000) for list purchase.
Total: $250 per (000).
G = $25 per (000), 10% above cost.
$D^* = (G/M + c)/g$.
$Z^{**} = (X - D)/[D(1 - D)/N]^{1/2}$.

are self-evident. First, list and cluster select decisions should *not* be based on sample response rates alone, which are prone to error. In Table 4, we note that the error resulting from a high response on a small sample size *more* than offsets the attractiveness of a high response rate. Second, the decision maker can easily convert all three variables: sample size (N), sample response rate (X), and average revenue per response (g) into a Z value. The Z statistic can be compared to a z_α for $\alpha = .01$ or $\alpha = .05$. The data in Table 4 show that only clusters #15 and #31 generate a z-value greater than z_α. Cluster #31 has a slightly *higher* z-value due to the greater revenue per response, in spite of a *lower* sample response rate.

The direct marketer is *well* advised to compute Z-scores to base the list selection decision. Our illustration demonstrates that lists with *higher* response rates can have a *lower* chance of success. In the next section we summarize our results and give some concluding remarks toward future research.

6. Conclusion

In this article, we addressed some important issues facing a direct marketing campaign. First, we addressed the subset size issue, i.e., how many creative packages or kits should be generated and tested before selecting the best. We referenced earlier important work by Gross (12, 13) in this direction, and proposed an alternative methodology which seems more appropriate in a direct marketing setting. Given the increasing interest in direct marketing promotion, these decision rules seem to be very useful for this growing field. Second, we presented methodology to derive cutoff values that can be used as a prerequisite for list purchase or the selection of a cluster within lists.

The decision maker needs to be aware that chances of success or failure in a rollout mailing are based on sample response rates, and that the error factor *is* an important issue, especially for small sample response rates. In addition, a decision rule was given which allows the decision maker to compute the required minimum cutoff values of a sample response rate which makes a rollout effort worthwhile. We also presented an illustration of how to use recent data in the "go/no-go" decision problem.

Managerial implications are clear. First, the "standard" rule for list selection, "check if test response rates are above a given threshold," is suboptimal. We have given a more refined rule in eq. 22. This rule can be applied to competing lists with different revenues per response, different rollout sizes, and different sample sizes. We also answered the subset size question: How many creative packages should be tested before we select the best one? The decision rule is given in eq. 15, and represents an alternative method to Gross (12, 13) where the subset size issue was addressed both in an advertising development setting and a direct marketing setting, respectively.

Directions of future research include list evaluation where there is an overlapping of names between lists. This problem is difficult to solve, because we lose independence across lists. It will be necessary to develop dependency measures across lists before it is possible to apply the methodologies presented here. In terms of managerial perspectives, applications of the suggested methodology using real databases will be helpful and cost-effective for decision makers in the field. Many firms that sell mailing lists avoid overlapping of names by offering a "merge-purge" program that automatically drops individuals from multiple listings and generates a new list of names with multiple listings.

An important extension of this article is the search for optimum subset size determination for situations where parameters such as α, β, N, and M are not constant, but vary from alternative to alternative. Recent results of Abughalous and Miescke (2) show that list selection in terms of the largest observed response rate $X_i = W_i/N_i$ may not always be the best decision rule. Preferable decision rules are derived, which are to be considered in the area of direct marketing. ■

REFERENCES

1. Aaker, David A., and John G. Myers (1982), *Advertising Management*, Englewood Cliffs, N.J.: Prentice Hall.

2. Abughalous, Mansour M., and Klaus J. Miescke (1989), "On Selecting the Largest Success Probability Under Unequal Sample Sizes," *Journal of Statistical Planning and Inference*, 21, 53-68.

3. *Advertising Age* (February 16, 1981).

4. Berger, James O. (1984), *Statistical Decision Theory and Bayesian Analysis,* 2nd ed., NY: Springer-Verlag.

5. Bickel, Peter, and Kjell A. Doksum (1977), *Mathematical Statistics,* NY: Holden Day.

6. DeGroot, Morris (1970), *Optimal Statistical Decisions,* NY: McGraw/Hill.

7. Ehrman, Chaim Meyer (1984), "Subset Selection Techniques: Development and Use in Marketing Applications," Doctoral dissertation, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104.

8. Ehrman, Chaim Meyer (1987), "Estimating the Probability of Lists' Rollout Success from Test Results," *Journal of Direct Marketing Research,* 1; 1 (Summer), 95–104.

9. Ehrman, Chaim Meyer, Abba M. Krieger, and Klaus J. Miescke (1987), "Subset Selection Towards Optimizing the Best Performance at a Second Stage," *Journal of Business and Economic Statistics,* 5; 2, 295–303.

10. Ehrman, Chaim Meyer (1988), "Using Bayes Statistics in a Direct Marketing Framework," *Journal of Direct Marketing,* 2; 1 (Spring).

11. Evans, Joel and Barry Berman (1987), *Marketing,* MacMillan.

12. Gross, Irwin (1966), "Creative Aspects of Advertising," Ph.D. Dissertation, Case Western Reserve, Cleveland, OH.

13. Gross, Irwin (1972), "The Creative Aspects of Advertising," *Sloan Management Review* (Fall), 83–109.

14. Johnson, Norman L. and Samuel Kotz (1970), *Continuous Univariate Distributions-*2, Boston: Houghton Mifflin.

15. Kotler, Philip (1984), *Marketing Management,* 5th edition, Englewood Cliffs, NJ: Prentice-Hall.

16. Owen, Donald B. (1962), *Handbook of Statistical Tables,* Reading Mass: Addison/Wesley.

17. *Wall Street Journal* (August 21, 1986), NY: Dow-Jones & Company, Inc.