THE ROLE OF WITHIN GROUP VARIANCE IN THE DESIGN AND ANALYSIS OF MARKET TESTS

David C. Schmittlein
Vijay Mahajan
Yoram Wind

David C. Schmittlein is Associate Professor and Yoram Wind is the Lauder Professor, Marketing Department, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104. Vijay Mahajan is the Herman W. Lay Chair Professor of Marketing, Cox School of Business, Southern Methodist University, Dallas, Texas 75275.

The authors thank Jerry Olson and the Bureau of Business Research at the University of Texas, Austin for providing the data on Texas Planning Regions. They also thank the Editor and two anonymous JMR reviewers for their helpful comments on the manuscript. This research was supported by a grant from the Center for Marketing Strategy Research, The Wharton School, University of Pennsylvania.

THE ROLE OF WITHIN GROUP VARIANCE IN THE DESIGN AND ANALYSIS OF MARKET TESTS

This paper is concerned with the variance within treatment groups for each of the three types of variables (treatment, control and response) present in market experiments and quasi-experiments. The effect of this within-group variance on the interpretation of experimental results is outlined conceptually, and illustrated through a set of realistic numerical examples. A useful approach for incorporating within-group variance into the design and analysis of market tests, based on Matusita's affinity between probability distributions, is also introduced.

1. Introduction

The last decade has seen a marked increase in the use of experiments and quasi-experiments to test the effectiveness of actual or proposed marketing programs. This increase has been caused, in part, by the increased availability of standardized data (including scanner data), technological developments (e.g., split cable and two way cable) and numerous methodological developments. Examples include concept testing (Wind 1982, Chapter 10), product tests (e.g., taste tests; Buchanan 1983; Morrison 1981), advertising experiments (Wind and Denny 1974), test markets (Blattberg and Golanty 1978; Pringle, Wilson and Brody 1982; Urban 1970) and simulated test markets (Silk and Urban 1978).

Some of these marketing tests are quasi-experiments (Cook and Campbell 1979, p. 6) using nonequivalent group designs, since individuals are not randomly assigned to treatment groups. Most test markets fall in this category, though the system developed by Information Resources, Inc. is a notable exception.[1] On the other hand, many concept and product tests and simulated test markets are true experiments which include random assignment. In this paper the term "market tests" will refer both to the experimental and quasi-experimental designs above.

Marketing experiments and quasi-experiments have been utilized primarily for two purposes: (a) selection of the best of a set of administered treatments (i.e., products, concepts or marketing programs) and (b) the use of the test results to define and develop an "optimal" treatment, which may not have been tested. Two examples of the latter are conjoint analysis based product optimization approaches

such as the POSSE System (Green et al. 1981); and the estimation of response functions for one or more of the marketing mix variables. In all of these cases the research task can be viewed as the administration of one or more treatments (e.g., ad campaigns) to one or more groups of individuals (e.g., cities).

Common to these experiments and quasi-experiments is the use of average measures to summarize the three major components of all experiments:

(a) "Control" variables such as income or age of the respondents in the test. The values of these variables are often summarized as averages for each group of individuals. In most of our examples we will use respondent characteristics as the control variables, although in practice marketing mix variables are also sometimes used as controls in quasi-experiments.

(b) Marketing mix treatment variables such as advertising or price which are often defined in terms of average level of advertising exposures or average retail price.

(c) Response measures such as sales, share or intention to buy which are often defined in terms of average level of sales or share or intention to buy in each experimental cell.

Ignored in virtually all marketing tests are the other distributional characteristics (and especially the variance) of the three sets of variables. Although this limitation has been recognized, little has been done to overcome it. Consequently the objectives of this paper are to:

1. Illustrate the importance of within cell variance comparisons in design and interpretation of market experiments and quasi-experiments, as they relate to the 3 sets of experimental variables (control, treatment and response variables).

2. Describe the conditions under which reliance on average measures can be misleading and the use of within cell variance comparisons should be undertaken.

3. Outline some approaches for dealing explicitly with within-cell variance comparisons.

4. Present two illustrations of the proposed approach as it relates to:

   (a) design of market tests, and

   (b) analysis of the results of such tests.

5. Discuss extensions of the proposed approaches to other distibutional properties.

2. On the Importance of Within Cell Variance Comparisons

As mentioned above, in designing and evaluating any market tests the researcher is interested in control variables, treatment variables (i.e., marketing mix variables) and the response variables. These three sets of variables and the typical procedures for analyzing them are illustrated in Table 1. These practices clearly ignore possible differences across experimental cells with respect to the variances of the variables. Therefore, incorporating the variance can lead, within each of the sets of variables, to four conditions:

1. means and variances are the same across the experimental cells (with respect to the control variables, treatment variables, or response variables),

2. the means are the same but the variances are different,

3. the means are different but the variances are the same, and

4. both the means and variances are different.

For each of the three types of variables in Table 1, it will be helpful to briefly describe the role that within-group variance plays in market experiments.

INSERT TABLE 1 ABOUT HERE

2.1 Control Variables

As Table 1 illustrates, these are often demographic, psychographic or product usage characteristics of respondents which can have their own effect on response measures like sales or intention to buy, or can moderate the effect of marketing variables on response. The marketer has essentially two strategies for coping with these effects. First, test and control groups can be matched on these control variables, which usually involves checking that the means of the control variables are approximately equal across groups. An example here would be the use of cluster analysis to choose test market regions (Green, Frank and Robinson 1967). Second, the control variables can be used as covariates, in effect adjusting the responses in each group to legitimize comparisons across treatment groups. In principle, this adjustment should be performed for each respondent. However the more common practice is to take an individual - level response function - relating household sales to head of household's age, for example - and use it to adjust the average response for a group, based on the group's average on the control variable.

In focusing solely on the means of the control variables, the two procedures above raise several questions, namely:

(a) from a theoretical standpoint, when can the mean be used?

(b) from a practical standpoint, how serious are the consequences in using only the averages? and

(c) when means alone are not sufficient for representing the control variables, how can a market test be designed to incorporate other characteristics of the distribution?

Questions a and c are answered in Sections 3 and 4, respectively, and while no definitive answer can be given to question b, a systematic set of numerical examples are constructed in the next section to shed some light on this issue.

In answering these questions we will presume that the researcher is able to identify the appropriate control variables -- those which have an effect on responses. This identification may be based on published research, secondary data services and past research conducted by the researcher's organization. However, in practice there will always be some omitted control variables. Also, the ones included will often be measured imperfectly. Both of these problems can lead to biased treatment effect estimates when the measured controls are used as covariates, even if our procedures below answering questions (a) - (c) are followed (Cook and Campbell 1979, pp. 159-202; Lord 1960). So this investigation of within group variances does not resolve all of the issues that arise in experimental and quasi-experimental designs. It simply describes (and suggests a solution for) one such problem that typically confronts a marketing researcher.

2.2 Treatment Variables

Reliance solely on the mean has been somewhat less of a problem in summarizing marketing mix variables, largely because of the way those treatment variables are applied. For most concept tests, product tests and simulated test markets a group of individuals assigned to a certain treatment do receive identical levels of the stimuli. So there is no within-group variance in the marketing mix variables. For example, simulated test markets rigidly control the number of advertising exposures, shelf facings, and the retail prices to which a respondent is exposed.

However, in situations where actual retail sales in a region constitutes the response measure, as is the case for test markets and some advertising quasi-experiments, the situation is much the same as for control variables described earlier. In test markets these treatment variables are usually described by total advertising expenditures or GRP's (equivalent to the average exposures per household) in a region, average retail price, average number of shelf facings, etc. So, with the exception of studies which consider both reach and frequency of advertising exposures, these test markets and advertising quasi-experiments again focus only on the variables' averages. Consequently, the same three questions raised for control variables are also relevant here (and, as it happens, are answered in the same way).

2.3 Response Variables

For treatment and control variables the importance of within-group variance stemmed from its effect on the distribution of responses in each group. The situation is different for response variables, where any interest in the within-group variance stems directly from the managerial significance of that variance. One example where the response variables' variance is important concerns behavior-based market segmentation. Consider a firm that runs its usual ad campaign in region A and an experimental/new campaign in region B. Assume that, prior to the test, the distribution across consumers of sales per capita in region A is (approximately) equal to the distribution in region B. Assume also that the brand's mean sales per capita in a suitable period after the campaigns have run is still equal in both regions, but the variance in sales across households is now greater in region B than region A. So region B has more people buying very large amounts of the brand, and also more people buying very small amounts. In this sense, region B is more

"polarized" about the brand than region A. If this polarization reflects higher loyalty toward the brand by its heavier users, the experimental campaign may be considered successful even though the average sales per capita has not changed. And even if the polarization is not indicative of loyalty, it suggests that the population in region B is more segmentable in terms of brand usage. Efficiencies resulting from market segmentation in region B again can mean that the new campaign is preferable to the old, without changing average sales.

The opposite situation, where the variance is smaller with the new campaign in region B, may also be of interest. A new advertising or promotion strategy intended to increase sales among nonusers or light users would be expected to increase the average per capita sales, and decrease the variance in sales across households (since nonusers are acting more like users). To evaluate the effectiveness of such a campaign it is helpful to check that both these effects have occurred. If average sales increase but the variance remains constant, it is possible that the sales increase simply reflects stocking up by heavy users, and is borrowing against future sales. This suspicion can be investigated with a longitudinal study of individual households' purchases.

2.4 A Numerical Illustration

To demonstrate the importance of the variance we can consider the following example in test market design. Different marketing programs are run in three regions, and the results are to be evaluated using sales per capita in each region. Since sales are known to vary with age, the three test regions have been matched on this control variable, with average age = 38 years.

Resulting sales per capita are given in the top of Table 2, which imply that the program used in region C is preferred to those in A and B.

INSERT TABLE 2 ABOUT HERE

However, a breakdown of sales per capita by age category, also in Table 2, suggests a different conclusion. For each region, the "Sales per Capita" column in the table gives the response function linking age to sales. This function is the same for regions A and C. However, in region B the sales per capita outperformed those in A and C among consumers aged <25 and 50-65, and equalled those in A and C among persons aged 25-50 and >65. So there is evidence that the program in B is superior to those in A and C. If the distribution of age were the same in all three regions then this response function for the program in region B would have generated the highest sales per capita. Clearly, the reason for the deceptive results in the top of Table 2 is that the age distribution in the three regions is different, even though average age is the same. The observed sales per capita in each region results from the interaction of two different factors: the function linking sales to age for individuals, and the distribtuion of age across individuals in that region. The next section gives the general conditions where matching groups only on means can lead to these deceptive findings, and also investigates the typical size of these effects.

3. When Can An Average Comparison be Used and When Should it be Supplemented with the Within Cell Variance Comparison?

3.1 The Effect of Variance on Expected Response

Returning to the previous example, one can develop the general relation between the distribution of a control variable such as age, and the expected

response. For a given value of age A and response measure S, let $S = F(A) + e$ represent the relation for individuals, where $E[e] = 0$. It will be convenient to let $g(A)$ denote the probability density function for the distribution of age. Then the expected individual response in a group or cell is $E[S] = E[F(A)]$ where E indicates the expected value over the individuals in the experimental cell.

While it is not always easy to express the expectation $E[F(A)]$ in a closed form, an approximation can be obtained using a Taylor series expansion for F. Expanding F around the point $A = E[A]$ yields:

$$S = F(A)+e = F(E[A]) + (A-E[A])\, F^{(1)}(E[A]) + \ldots + \frac{1}{j!}(A-E[A])^j F^{(j)}(E[A]) + \ldots + e \qquad (1)$$

where $F^{(j)}(E[A])$ is $\frac{\partial^j F(A)}{\partial A^j}$ evaluated at $A=E[A]$.

Using the first three terms as an approximation to F(A) and taking the expected value of both sides in (1) leads to the desired formula:

$$E[S]=E[F(A)] \doteq F(E[A]) + \frac{1}{2}\, Var[A] F^{(2)}(E[A]). \qquad (2)$$

Relation (2) illustrates the impact of the variance of age on the expected response measure S. Clearly, if F is linear the variance has no effect and $E[F(A)] = F(E[A])$. Alternatively, when the second derivative of F is negative at E[A] (e.g. F is concave) we will expect $E[F(A)]$ to be less than $F(E[A])$.

INSERT FIGURE 1 ABOUT HERE

This effect of the variance when F is concave can easily be shown graphically. The situation is depicted in Figure 1 where for ease of illlustration the response function--F(A) and the distribution of the control variable age--g(A) are shown on the same graph. We are interested in the expected value of F(A). The height of the density function g(A) indicates the likelihood of occurrence of both A and the corresponding value F(A). Since F's values are represented on the vertical axis the distribution h() for F(A) which results from g(A) can also be drawn on that axis. For example, in Figure 1 the likelihood that $A=A_1$ is given by the length of line $0_1 0_2$. Consequently, the likelihood that $F(A) = F(A_1)$ is equal to the length of $0_3 0_4$, which also equals the length of $0_1 0_2$. Since F is concave, the values of F(A) above F(E[A]) (when compared to the ones below) are concentrated closely around F(E[A]). Hence the distance $0_5 0_6$ on the response axis is considerably smaller than the distance $0_7 0_8$ on the age axis. The expected value of the response S, E[S] = E[F(A)] is seen to be less than F(E[A]). So in a test market setting if average age is the same in two regions, F is concave, and the variances of age are unequal, the region with the larger variance is expected to have a lower response.

The approximation (2) for the expected response can be easily extended to the case of multiple control variables $X_1, \ldots, X_k$. The response S as a function of the control variables is again $S = F(X_1, \ldots, X_k) + e = F(\vec{X}) + e$. Expanding F in a multivariate Taylor series around the point $(E[X_1], \ldots, E[X_k]) = E[\vec{X}]$ (Fleming 1977, p. 94) and taking the expected value of terms up to the second derivative leads to

$$E[S] = E[F(X_1, \ldots, X_k)$$

$$\doteq F(E[\vec{X}]) + \frac{1}{2} \sum_{i=1}^{k} \mathrm{Var}[X_i]\, F_{ii}(E[\vec{X}])$$

$$+ \sum_{i>j} \mathrm{Cov}[X_i, X_j]\, F_{ij}(E[\vec{X}]) \qquad (3)$$

where $F_{ij}(E[\vec{X}])$ is $\frac{\partial^2 F(\vec{X})}{\partial X_i \partial X_j}$ evaluated at $\vec{X} = E[\vec{X}]$.

So as one would expect when more than one control variable is relevant, the expected response S depends on both the variances and covariances of those controls.

INSERT TABLE 3 AND FIGURES 2 AND 3 ABOUT HERE

3.2 Effectiveness of Predictions Using the Mean vs. Mean and Variance

This discussion has given an informal argument for the benefits obtained when the two-term approximation (2) to E[S] is used. To more formally investigate the use of the expectation and variance of a control variable (A) to predict response, we must specify the shape of the response function F(A) and the distribution of the control variable A over the population. Table 3 reports the results for the combination of three response functions with four distributions for A. The response functions used are sketched in figure 2 and the distributions for A are given in Figure 3

As can be seen from the diagram, response F_1 indicates that S increases less than linearly with A, and for F_2 S's increase with A is more than linear. The function F_3 exemplifies a unimodal response where S first increases, then decreases with A. The general distribution chosen to

represent A in cases 1-4 (Figure 3) is the gamma distribution, with the probability density function

$$g(A) = \frac{\alpha}{\Gamma(r)} (\alpha A)^{r-1} e^{-\alpha A} \quad , A>0. \tag{4}$$

With this two-parameter distribution the expected value and variance of A are:

$$E[A] = r/\alpha \ ,$$
$$Var[A] = r/\alpha^2 \ .$$

The gamma distribution is always unimodal, but otherwise is quite flexible. This model is widely used to represent variables which are greater than zero and have unimodal, right-skewed distributions (as do many demographic and product-use characteristics). The particular values of the shape parameter r chosen in cases 1-4 have been used previously for the distribution of income (McDonald and Jensen 1979) while α is simply a scale parameter which adjusts the mean of the distribution. The four cases to be examined are based on two dimensions: the density of the distribution, represented by the coefficient of variation (small: $r = 1$; large: $r = 2$) and the mean (small: $\alpha = 2$; large: $\alpha = 1$).

The formulas for the expected response $E[S] = E[F(A)]$ when the three functions F_1, F_2, F_3 are combined with the gamma distribution for A are derived in the Appendix. So for each response function F_i, $E[F_i(A)]$ in Table 3 is the exact expected response using these formulas. The estimated response S using information on both the mean and variance of A, denoted the M-V estimate in Table 3, was computed using equation (2).

Finally, the estimate of the response using only the mean - the M estimate in the exhibit - uses only the first term on the right in equation (2). To indicate the extent to which the M-V estimate outperforms the estimate using the mean alone, the percent improvement for the former over the latter is also reported. This is defined as

$$\text{Percent Improvement} = 100\left(1 - \frac{\text{ABS}(E[F_i(A)] - \text{M-V estimate})}{\text{ABS}(E[F_i(A)] - \text{M estimate})}\right)$$

where ABS( ) indicates the absolute value. In addition to the percent improvement, we are also interested in the absolute difference between the exact expected response and the M estimate. That is, if the M estimate is already accurate in an absolute sense, the percent improvement with the M-V estimate loses its importance.

For example, the combination of case 4 and response function F_1 leads to an 87% improvement in the error of M-V over that of M. The absolute difference between the expected response $E[F_1(A)]$ and the M estimate is 1.41 - 1.33 = .08. Whether this discrepancy is "large" (and hence the 87% improvement is noteworthy) depends on the managerial significance of the response measure $F_1(A)$. For some applications this bias may be negligible. On the other hand management might well be interested in a difference in projected sales of 1.41 million units versus 1.33 million units, especially if the breakeven point was 1.25 million units.

With the exception of the F_3 - case 4 combination the percent improvement for the M-V estimate ranges from 85% to 98%. In fact, the combination of case 4 with F_3 is unlikely to occur in practice since this would imply that for almost half of the values of the control variables the response is equal to zero. So the results in Table 3 show that market tests which examine only the means of the

design variables can lead the researcher's expected results to be far from the actual results both in absolute terms (i.e., the size of $E[F_1(A)]$- M-estimate) and relative to tests which also account for the variance in the design variables.

3.3 When Can Means Be Used?

As was noted above, if the response function F is linear the expected value of S depends only on the average of the control variable(s) and not on the variances. This result, with the addition of two more observations, will give us the conditions where averages can be used to summarize the the impact of control variables and treatment variables on responses.

First, it is easy to show that the analysis of equations (1) - (4) for control variables can be extended in the same way to the effect of marketing treatment variables on the expected response. If we let A now represent exposures to an advertisement, F(A) be the response of a household to A exposures, and g(A) be the distribution of ad exposure across households

in a region, equation (2) again gives the expected response per household in this region. This says that from the standpoint of within group variance, all of the experiment's design variables (i.e. the control and marketing treatment variables) behave according to the same relation (2).

The second observation of interest concerns the other properties of the response S, besides its expectation. Returning to equation (1), if S is linearly related to A the expression for $S = F(A) + e$ can be written exactly as

$$S = F(A) + e = F(E[A]) + (A-E[A])F^{(1)}(E[A]) + e \ . \qquad (5)$$

Taking the variance of both sides in (5) and assuming that A and e are uncorrelated yields

$$Var[S] = (F^{(1)}(E[A]))^2 Var[A] + Var[e]. \qquad (6)$$

So if F is linear, the variance of A does not affect the expected response E[S], but it <u>does</u> affect the variance of the response.

Using these observations, it is clear that the averages of the design variables can be effectively used for comparing results across treatment groups if:

1. the variances of all design variables are equal across cells of the experiment or quasi-experiment, or

2. for all design variables with unequal variances across cells,

   (a) the relation between the response measure(s) and those design variables is linear, <u>and</u>

   (b) the researcher is only interested in the average of the response variable(s). In particular, differences across cells in the variance of the response are not of interest.

Both conditions 2a and 2b may fail to hold in practical market tests. In 2a, note that the response must be linear over the entire range of the design

variable. If we again take age as an example, consumption of some products increases almost linearly with age - over a certain age range (e.g., 25-45 years). But that linearity typically breaks down in both the higher (over 55) and lower (under 18) age categories, where purchases are less than a linear response function would predict.

Condition 2b can also fail to hold. Strictly speaking, in situations where cells are compared using analysis of variance the validity of the test statistics depends on the assumption of equal variances in the response measure across cells. However, most studies have found that ANOVA and MANOVA are robust to violations of this homogeneity-of-variance assumption (Wildt and Ahtola 1978, pp. 89-90). So unequal variances in the responses does not generally lead to a failure of condition 2b. (In fairness, though, Glass et al. (1972, pp. 274-275) found that this robustness for ANCOVA depends on the values of the covariates.)

But condition 2b can fail to hold for another reason. Namely, as noted in Section 2.3 the within-cell variances in the response variables can be of interest in their own right. Such an inherent interest in a characteristic of the responses which is not captured in the mean has two implications. First, within-region variance in the design variables cannot be ignored even when the responses are linked linearly with the control variables and marketing variables. In other words, to ensure that the cells of a market test are comparable it is necessary that the within-cell variances (as well as the means) of the design variables are equal across cells.

The second implication of an interest in within-cell variance for the responses concerns the analysis of results from a market test. Usually, an index of the difference between results in one cell and results in another is used to summarize the findings. Such an index here should not depend solely

on the average response in a cell, but should (at least) include the variance in response as well. So, from a methodological standpoint this development of an index of between-cell differences in results from a market test and the matching of cells in the design involve the same issue: measurement of differences between groups of multivariate observations, where the measure reflects differences across groups in both the means and variances of the observations. The next section is concerned with designing such a measure. Subsequently, Sections 5 and 6 apply the measure to the matching of regions in designing a test market and analyzing test market results, respectively.

4. Comparing Distributions Based on Both Means and Variances

To develop a distance between distributions based on the mean and variance, it will be helpful to introduce some notation. For M distributions, each of which are K-variate let the mean vector and covariance matrix for distribution i be μ_i and Σ_i, respectively. Similarly, let the K-vector of variable variances for distribution i be σ_i^2. Then, intuitively, a distance between distributions i and j should take into account the difference between μ_i and μ_j and also the difference between σ_i^2 and σ_j^2.

One possible approach would be to let $\gamma_i = (\mu_i , \sigma_i^2)'$ be the vector of both means and variances for distribution i. Then a measure of the distance between distributions i and j is

$$D_{ij}^2 = (\gamma_i - \gamma_j)' \Lambda^{-1} (\gamma_i - \gamma_j)$$

Where the 2K x 2K matrix Λ^{-1} weights the differences in means and variances for the K variables. When Λ is the covariance matrix of γ_i over the M distributions, D_{ij}^2 is a Mahalanobis distance between the means and variances.

Morrison (1967) has recommended the use of Mahalanobis distance as an input to cluster analysis. Clearly, when Λ is the identity matrix then D_{ij}^2 is just the squared Euclidean distance between the means and variances in one group with the corresponding means and variances in another.

The measure D_{ij}^2 is easy to calculate, and is a very natural extension of the Euclidean and Mahalanobis distances which are based only on means. However, it does not take into account differences between distributions in the covariances as well as the variances. As equation (4) demonstrated, the expected response in an experiment's cell also depends on covariances of the control variables. Matusita (1966, 1977) has introduced a measure of distance between distributions which overcomes this shortcoming and has several other attractive properties. Let $g_i \equiv g_i(x_1, \ldots, x_K)$ represent the probability density function for distribution i. Then Matusita's distance is

$$d_{ij}^2 = \int ([g_i(x)]^{1/2} - [g_j(x)]^{1/2})^2 \, dx \quad . \tag{8}$$

An important characteristic of d_{ij}^2 is the fact that $d_{ij}^2 = 2(1-\rho_{ij})$, where

$$\rho_{ij} = \int [g_i(x)]^{1/2} [g_j(x)]^{1/2} \, dx \tag{9}$$

and $0 \leq \rho_{ij} \leq 1$. ρ_{ij} is known as the <u>affinity</u> between two distributions, and is a measure of their similarity. Matusita (1977) shows that affinity is related to the error rate in discrimination between distributions. As is generally true with discriminant analysis, given distributions $g_i(x_1, \ldots, x_K)$ and $g_j(x_1, \ldots, x_K)$ one is trying to be able to classify future observations as having come from g_i or g_j. Some proportion w_i of future sample values will actually come from g_i, and the rest from g_j. The discriminant function partitions the K-dimensional space into two regions Ω_i and Ω_j, where future points lying in region Ω_i will be said to come from

g_i. With this decision rule the error rate is the expected number of future observations misassigned; and an optimal partition (Ω_i^*, Ω_j^*) makes this error rate as small as possible. Matusita (1977, p. 217) notes that this error rate with an optimal partition will be less than or equal to $w_i(1-w_i)\ \rho_{ij}$ and must be greater than or equal to $1/2\ w_i(1-w_i)\ (\rho_{ij})^2$. So, as expected, as the affinity increases from 0 to 1 (and so the distance decreases from 2 to 0) the minimum error rate in discrimination also increases.

Matusita's affinity ρ_{ij} and distance d_{ij}^2 have two other features which are useful. First, as is evident in (8) and (9), ρ_{ij} and d_{ij}^2 depend in general on the entire distribution $g_i(x)$. So they include, but are not limited to, comparison based on means, variances and covariances. This advantage will be taken up again in section 7. Second, when it is desirable to compare distributions based only on the first two moments, a convenient special case of equation (9) is available. If the $g_i(x)$, $i = 1, \ldots, M$ represent normal distributions with mean vector μ_i and covariance matrix Σ_i then Matusita's affinity ρ_{ij} becomes

$$\rho_{ij} = 2^{K/2} \frac{|\Sigma_i\ \Sigma_j|^{-1/4}}{|\Sigma_i^{-1} + \Sigma_j^{-1}|^{1/2}} e^T \tag{10}$$

where

$$T = 1/4\ [(\Sigma_i^{-1}\ \mu_i + \Sigma_j^{-1}\ \mu_j)'(\Sigma_i^{-1} + \Sigma_j^{-1})(\Sigma_i^{-1}\ \mu_i + \Sigma_j^{-1}\ \mu_j) - \mu_i'\ \Sigma_i^{-1}\ \mu_i - \mu_j'\ \Sigma_j^{-1}\ \mu_j]$$

and $|\Sigma|$ indicates the determinant of the matrix Σ.

So, since $d_{ij}^2 = 2(1-\rho_{ij})$, for normal distributions both Matusita's affinity and distance are easy to compute knowing only the means, variances

and covariances. When the covariance matrices Σ_i and Σ_j are equal ρ_{ij} is ordinally equivalent to the Mahalanobis distance between two distributions. In fact, when $\Sigma_i = \Sigma_j = \Sigma$ equation (10) becomes

$$\rho_{ij} = e^{-1/8\ T_1} \tag{11}$$

where $T_1 = (\mu_i - \mu_j)' \Sigma^{-1} (\mu_i - \mu_j)$. So comparing (11) with (7), in the normal case when $\Sigma_i = \Sigma_j$ Matusita's affinity is related to the Mahalanobis distance D_{ij}^2 by

$$\rho_{ij} = e^{-1/8\ D_{ij}^2} . \tag{12}$$

Finally, we note that when the means are equal ($\mu_i = \mu_j$) but the covariance matrices are different, T=0 and equation (10) becomes

$$\rho_{ij} = 2^{K/2} \frac{|\Sigma_i \Sigma_j|^{-1/4}}{|\Sigma_i^{-1} + \Sigma_j^{-1}|^{1/2}} . \tag{13}$$

An axiomatic rationale for use of Matusita's distance d_{ij}^2 has been described by Kaufman and Mathai (1973). It is ordinally equivalent to Renyi's entropy function for discriminating between distributions (Renyi 1961; Aczel and Daroczy 1975) and has been used by Stein (1965) for measuring the dissimilarity of two posterior distributions. It is also closely related (though not equivalent) to the Kullback-Leibler information for discriminating between two distributions (Kullback 1968; Goel and DeGroot 1981). For example, Matusita's distance is symmetric, $d_{ij}^2 = d_{ji}^2$, while Kullback-Liebler is not. Finally, d_{ij}^2, (like all of the other measures mentioned in this paragraph) is a special case of the so-called

"f-divergence" measure between distributions (Ali and Silvey 1966; Goel 1983).

The next two sections provide applications of Matusita's affinity to the selection of test market regions and the analysis of test market results, respectively. Equation (10) will be used to compute the affinity ρ_{ij} between distributions. So implicitly it is assumed that the control and response variables can be approximated, for these purposes, by the normal distribution.

5. Illustrative Application: Selection of Test Market Regions

In designing a test market to compare different marketing programs it is usually desirable that the regions used for comparison be as similar as possible on some demographic (control) characteristics. This can be accomplished by clustering possible test market regions, and choosing for comparison sites that fall in the same cluster (Green, Frank and Robinson 1967). However, as the results of sections 2 and 3 indicate, it is important to have regions which are comparable on variances, as well as means of the control variables.

This section describes the use of Matusita's affinity ρ_{ij} to cluster 24 planning regions in Texas[2]. These results are compared with the more standard clustering of regions based only on the means. Our main finding is that the two clustering solutions bear only a slight resemblance to each other. So the empirical evidence suggests that clustering based on means alone is not an effective surrogate for clustering on means, variances and covariances. We expect that the same result would hold in most cases where test market sites are being selected. However, since the

specific characteristics of this application will not match those in many test market situations the generalizability of the conclusions here is an open question.

Features defining the scope (and generalizability) of this study are:

1. Three control variables are used to assess the similarity of regions: age, education and household income. All are as measured by the U.S. Bureau of the Census. Excluded are other variables which can be important in matching sites, such as usage characteristics for the product being tested.
2. All 24 regions are in Texas. Given the wide variation in the three demographic characteristics across Texas this is probably not a serious restriction.
3. In many practical cases boundaries of test market sites are defined by the media markets (e.g. areas of dominant influence for television), not planning regions. Although the planning regions are non-overlapping, cover the entire state, and are generally centered around metropolitan areas; they tend to be smaller than ADI's. So the regions used here are more similar to those in minitest market experiments (Wind 1982, pp 427-429).
4. For each of the 24 planning regions the means and variances of the control variables were computed across the subregions (usually counties) lying within it. So the subregion is the unit of analysis, and the data used are the mean income, age and education for each subregion provided by the census bureau. For example,

the computed within-region variance of income in a region is the variance, over subregions, of the average income in each subregion.

Conceptually, for a planning region it would be preferable to evaluate the variance in income over individual households, rather than the variance in average subregion income. However, since census data on individual respondent households are not generally available the approach chosen here is more likely to be a feasible option for a test marketer. For the state of Texas there are between 10 and 71 subregions per planning region.

First, the 24 regions were clustered based only on the means of the three controls (income, age and education). For clustering, the average income, age and education values were standardized across planning regions, and the sum of squared deviations on these three values was used as the distance measure. The regions were clustered hierarchically using the BMDP2M program (Dixon, 1981) with average linkage as the criterion for forming groups. For the resulting 10-cluster solution regions that are within the same group are clearly very similar on average income, age and education. However, there is not a great deal of similarity in variances for regions in the same cluster. Regions with small and large variances are frequently grouped together.

One example of this is regions 19 and 21, which were placed in the same cluster. The means for the demographic variables in these 2 regions were, respectively: income (10.86, 11.12), age (29.53, 30.48), and education (8.13, 8.00), indicating close agreement. On the other hand the variances of

the demographic variables in regions 19 and 21 were, respectively: income (3.59, 12.97), age (1.95, 19.33), education (1.20, 1.08). So the profile of variances in region 19 is not at all similar to that in region 21. Looking only at the variances, regions 19 and 21 would not have been clustered together. This example is typical of the results obtained.[3]

The ability of clustering on means to (inadvertently) cluster on variances as well was investigated further; by correlating each variable's mean with its within-region variance, over the 24 regions. If these correlations were high in absolute value, then regions which were similar on means would also be similar on variances. These correlations were .50, .24 and -.43 for income, age and education, respectively. From the evidence here, it seems unrealistic to assume that regions clustered on means will be roughly equivalent on variances as well.

As an alternative to this clustering approach, the regions were grouped using Matusita's affinity ρ_{ij} (equation (10)) as an index of similarity. The information required for each region consisted of the means, variances and covariances of income, age and education. The hierarchical clustering routine BMDP1M was employed with linkage based on the group centroids. Ten clusters were chosen because, subjectively, this seemed to maximize the similarity between the two clustering solutions (excluding the trivial cases of 1 cluster and 24 clusters). The correspondence between this clustering solution and the one above based simply on means is only slightly greater than would be expected by chance. From a decision making standpoint they are not acceptable substitutes for one another. So this empirical application highlights the

shortcomings of test market site selection based only on means, and illustrates the use of an alternate method (Matusita's affinity) which is conceptually preferable, uncomplicated, and uses data which are generally available.

The results here do not prove that Matusita's affinity is the best similarity measure for comparing distributions, and competitors which also incorporate means and variances can be imagined. (One such distance measure D^2_{ij} was described in Section 4). But any such alternatives must implicitly or explicitly weight the relative contribution of means, variances and covariances to such a measure. It is not obvious what these weights should be. Consequently the affinity, being related to the errors in discriminating between distributions, provides at least a reasonable solution to this problem.

6. An Illustrative Application: Analysis of Results from a Market Test

In contrast to the last section's application which used Matusita's affinity in designing market tests, this section uses the affinity to help analyze results from such a test. The objective of the test market conducted here was to discover the effect of both advertising expenditures and the method of distribution on sales for a consumer nondurable. Three levels of advertising expenditure and two distribution methods ("Regular" and "Special") were chosen, so each test market region was assigned to one of the 6 possible experimental cells.

INSERT TABLE 4 ABOUT HERE

A separate index of sales was available from two different data sources (i.e. research suppliers). These two data sources used different samples of retail outlets within each region. The two resulting sales measures will be denoted "Sales from Source A" and "Sales from Source B" below. From each source the dependent variable of interest was sales in the region after the test minus sales in the region prior to the test. (The sites tested here are completely unrelated to the Texas regions discussed in the last section.) There were 9 observations per experimental cell.

The usual procedure for analyzing this type of data would involve comparing the average change in sales (i.e., posttest sales - pretest sales) across the treatment cells via ANOVA or MANOVA. If ANOVA is used the changes for sales from source A are analyzed separately from the changes in sales from source B. Actually in this case it was felt that some additional characteristics of the test sites could also affect the change in sales. So those variables were added as covariates and a multivariate analysis of covariance was performed. The MANCOVA results are listed in Table 4. They suggest that the covariates ("within cells regression" in the table) are not significant, nor is the main effect of advertising. On the other hand, the interaction of advertising and distribution, the constant term, and possibly the main effect of distribution are significant. The constant term's significance says that, over all 6 cells, there is a trend in sales from the pre to post experimental period.

While analyzing the differences in average sales change across cells, MANCOVA ignores any differences across cells in within-cell variance for the 2 response measures. As suggested in Section 2.3 we may also want to know, in each cell, how the (bivariate) distribution of (Source A sales, Source B sales) has changed from the pretest period to posttest period. But given the large number of features (e.g., means, variances, covariances) which the distributions

possess it can be difficult to know which observed changes are most important. This is where the affinity ρ_{ij} can help. It can measure, for each experimental cell, the degree to which the distribution of sales (across sites in that cell) has changed over the treatment period. Then, for those cells where the greatest change has occurred, the distribution's characteristics (e.g., means, variances) can be examined to see which ones account for the change.

The statistical properties of the affinity are also available (Matusita 1966) to indicate whether the overall observed change in the distribution is statistically significant.[4] If the researcher had some hypotheses about variances or correlations a priori, he or she would probably want to examine these directly. However if there are no a priori hypotheses, statistical tests on the affinity measure can alert the researcher to potential differences in these other distributional characteristics. Subsequently, specific tests for these characteristics could be carried out.

INSERT TABLE 5 ABOUT HERE

Table 5 gives the results when this approach is applied to our test market data. For each of the 6 cells, the affinity between the actual distribution of responses and the expected distribution of responses is reported. This expected distribution is the pretest distribution plus the overall trend in each of the sales measures over time. Hence if the affinity = 1 for some experimental cell then the treatment has had no effect on the bivariate distribution of sales. For example, this affinity is .797 for the cell with medium advertising expenditures and the special distribution method. The expected and actual means, variances and correlation for the two sales measures are also given in Table 5. These are the quantities from which the affinity was computed.

The experimental groups in the exhibit are listed in order of the size of the change in the distribution, i.e., the size of the marketing program's effect. So cell I showed the greatest effect (lowest affinity) and cell VI showed the least (i.e., highest affinity). That cells I and II show a large treatment effect is consistent with the MANCOVA results for the means alone. The mean sales were especially low in cell I (low advertising, regular distribution) and particularly high in cell II (medium advertising, special distribution).

However, another factor besides the means seems to account for the low affinity between actual and expected results. The correlation between the two sales measures has decreased. Among other possibilities, this would occur if one of the sales measures responds more quickly or dramatically to the treatment than the other measure. Cells I and II having the lowest affinity values also highlights an important property of the measure ρ_{ij}. It indicates the extent of change, but not the desirability of that change from a managerial perspective. For that, the experimenter must return to the specific characteristics of the distribution.

In summary, use of the affinity measure can help identify which treatments have had the greatest effect on the distribution of the dependent variables. In doing so, it shows whether those effects are primarily due to differences in the mean of the response measure, or to other characteristics of its distribution. Consequently, it draws the experimenter's attention to that characteristic through which the treatment has had the greatest effect on the distribution of response.

7. Discussion

The main emphasis in the last three sections has been the description of a measure of similarity (ρ_{ij}) and distance (d_{ij}^2) between distributions, and their application to problems in design and analysis of market experiments. Here, we discuss briefly the extension of these measures to other characteristics of the distribution, and recommend specific steps for designing and evaluating market experiments which stem from the results in this paper.

7.1 Extension to Other Distributional Characteristics

The applications in sections 5 and 6, using equation (10), focus comparisons on the first two moments of the distributions. However, when percentiles of the distributions are known one can return to equations (8) or (9) to approximate d_{ij}^2 and ρ_{ij}. For matching test market regions this is usually easy if only one control variable is being considered, since histograms giving the distribution of demographic characteristics (e.g. age, income) or product usage characteristics (e.g. purchase rate) for each region are often available.

When matching regions on multiple control variables which are intercorrelated the procedure is slightly more involved. The entire multivariate distribution of the control variables (8) and (9) is generally not known. Instead, the histogram giving the marginal distribution of each control variable, and the correlation of each pair of control variables, can often be obtained. Then the percentage points of the multivariate distribution can be approximated using these marginals and correlations. So a variety of methods for evaluating d_{ij}^2 and ρ_{ij} are possible, depending on the data at hand.

7.2 Recommendations for Market Experiments

The analysis in this paper of distributional properties, especially within-group variance, for control, treatment and response variables leads to the following recommended steps for designing and analyzing market experiments:

1. Select the relevant control, treatment and response variables.
2. Decide on the sample size and sampling frame for each of the treatment and control groups.
3. Assess whether conditions 2a and 2b of section 3.3 are satisfied. If not, when assigning individuals to treatments, make certain that the different treatment groups (or pairs of treatment and control groups) have similar distributions for the control variables. Matusita's affinity ρ_{ij} can be used to help accomplish this matching.
4. In analyzing the results, see if the variances and covariances of the dependent variable(s) are equal across treatment cells. If so, a procedure such as analysis of variance can be used to examine differences in means. If not, the affinity can be used to identify which treatments have the greatest effect on the distribution of response.

To summarize, the analysis here provides some insight into the consequences of using averages alone to summarize variables in market experiments, contrasted with the use of both means and variances. A method based on the affinity of distributions is described, which allows an individual to incorporate these distributional characteristics. As the applications in sections 5 and 6 demonstrate, this index could help improve current analyses of average response to marketing variables, through better controls. It could also help the researcher understand how marketing programs affect the entire distribution of response measures such as household sales and market share. The discussion and empirical examples in this paper all relate to market experiments. However it is clear that the same conclusions and methods apply more generally. They are relevant for all experiments and quasi-experiments which compare responses for groups which have received different treatments.

FOOTNOTES

[1]Information Resources, Inc. has individually - targetable cable television capability for its panels of participating households. So, within a geographic region, households can be randomly assigned to the particular advertising campaigns (i.e., treatments) being tested. All purchases made by those households at stores equipped with universal product code scanners are recorded automatically by computer.

[2]For a more detailed description of these planning regions see Pluta, Wright and Anderson (1982, pp. 9-14).

[3]The complete cluster analysis results are available from the first author in a longer version of this paper.

[4]In the special case where the distributions being compared are multivariate normal the sampling properties of Matusita's affinity have a familiar and simple form. For example, let F represent the normal distribution $N(\mu, \Sigma_1)$ and S_n represent the normal distribution estimated using an i.i.d. sample of size n from the distribution $N(\delta, \Sigma_2)$. Here δ need not equal μ. It is known that $\Sigma_1 = \Sigma_2 \equiv \Sigma$, though Σ itself is unknown. Then, denoting the affinity between S_n and F by $\rho(S_n, F)$, $-8 \log \rho(S_n, F)$ is the usual generalized T^2 statistic, which has the noncentral F distribution (Matusita 1966, pp. 192-193). With these results one can test, in particular, whether the affinity among distributions is significantly different from 1. Used longitudinally, it tests whether the distribution in one experimental cell has changed over time. Used cross sectionally, it tests whether the distribution (pretest or posttest) is the same across experimental cells.

APPENDIX

In general, the expected response $E[S]=E[F_i(A)]$ when A has the gamma distribution (3) can be written

$$E[F_i(A)] = \int_0^\infty F_i(A) \frac{\alpha}{\Gamma(r)} (\alpha A)^{r-1} e^{-\alpha A} dA \quad . \tag{A1}$$

Substituting $F_i(A) = A^\beta$ in (A1) leads to the result

$$E[F_i(A)] = E[A^\beta] = \alpha^{-\beta} \frac{\Gamma(r+\beta)}{\Gamma(r)} \quad . \tag{A2}$$

So the expected response for functions F_1 and F_2 are obtained by substituting $\beta = .5$ and $\beta = 1.5$, respectively, in (A2).

Similarly, the expectation of $F_3(A)$ can be written

$$E[F_3(A)] = \int_0^4 \frac{\alpha^r}{\Gamma(r)} A^r e^{-\alpha A} dA - .25 \int_0^4 \frac{\alpha^r}{\Gamma(r)} A^{r+1} e^{-\alpha A} dA$$

$$= \frac{r}{\alpha} \int_0^4 \frac{\alpha}{\Gamma(r+1)} (\alpha A)^r e^{-\alpha A} dA - \frac{r(r+1)}{4\alpha^2} \int_0^4 \frac{\alpha}{\Gamma(r+2)} (\alpha A)^{r+1} e^{-\alpha A} dA$$

$$= \frac{r}{\alpha} F_G(4|r+1,\alpha) - \frac{r(r+1)}{4\alpha^2} F_G(4|r+2,\alpha) \tag{A3}$$

where $F_G(a|x,y)$ is the c.d.F. of the gamma distribution with shape parameter x and scale parameter y. Equation (A3) is easily evaluated using the relationship.

$$F_G(a|x,y) = 1 - \sum_{k=0}^{x-1} e^{-ya} \frac{(ya)^k}{k!}$$

for integer values of x.

Table 1

COMMON PRACTICE REGARDING WITHIN-REGION VARIANCE
FOR TEST MARKET VARIABLES

VARIABLE TYPE	EXAMPLES	TYPICAL PRACTICE
1. Control Variables	Respondents' - Income - Age - Education	a) Matching on Means b) Random selection followed by examination of means c) Means as covariates
2. Treatment (Marketing Mix) Variables	- Advertising - Distribution - Pricing	Selection of a number of treatments (e.g. advertising exposures, retail prices, etc.) which vary with respect to their average levels.
3. Response Variables	- Sales - Share - Intention to buy	Choose one or more response measures and examine differences in means (ANOVA, MANOVA).

Table 2

THE IMPACT OF VARIANCE IN CONTROL VARIABLES ON RESPONSE

Evaluation Using Means

Region	A	B	C
Control Variable:			
Average Age	38	38	38
Response Variable:			
Sales per Capaita	2.98	3.11	3.18

Conclusions: The 3 regions are comparable with respect to age. Region C outperforms A and B in terms of sales per capita.

The Impact of Within-Region Variance

SALES BY AGE CATEGORY

Region	A		B		C	
Age	% in Category	Sales per Capita	% in Category	Sales per Capita	% in Category	Sales per Capita
<25	22	2	23	2.5	18	2
25 - 34	18	4.5	20	4.5	21	4.5
35 - 49	18	5	16	5	21	5
50-65	20	3	17	3.5	21	3
>65	22	1	24	1	19	1

Conclusions: The 3 regions are not directly comparable with respect to age. Region B outperforms A and C on a sales per capita basis.

Table 3

PREDICTION OF EXPECTED RESPONSE USING
THE MEAN ONLY, vs. USING THE MEAN AND VARIANCE

Case		1	2	3	4
Distribution of A	r	1	1	2	2
	α	2	1	2	1
	E[A]	.500	1.00	1.00	2.00
	Var[A]	.250	1.00	.500	2.00
	$E[F_1(A)]$	.627	.886	.940	1.33
$F_1(A)=A^{.5}$	M-V estimate	.619	.875	.937	1.32
A>0	M estimate	.707	1.00	1.00	1.41
Percent Improvement		90	90	95	87
	$E[F_2(A)]$	.470	1.33	1.17	3.32
$F_2(A)=A^{1.5}$	M-V estimate	.487	1.37	1.19	3.36
A>0	M estimate	.354	1.00	1.00	2.83
Percent Improvement		85	88	88	92
	$E[F_3(A)]$	.375	.527	.627	.674
$F_3(A)=A-.25A^2$	M-V estimate	.374	.500	.625	.500
0<A<4	M estimate	.437	.750	.750	1.00
Percent Improvement		98	88	98	47

Table 4

ANALYSIS OF THREE ADVERTISING LEVELS AND TWO DISTRIBUTION STRATEGIES: TEST MARKET MANCOVA RESULTS

Effect	Wilks' Approximate F-Statistic	Degrees of Freedom	Significance Level
Within Cells Regression	1.333	4,90	.264
Advertising X Distribution	2.955	4,90	.024
Advertising	.704	4,90	.591
Distribution	2.444	2,45	.098
Constant	3.492	2,45	.039

Table 5

INCORPORATING WITHIN-REGION VARIANCE
IN TEST MARKET RESULTS

	Experimental Cell*	Affinity	Sales From Source A		Sales From Source B		Correlation
			Mean	Variance	Mean	Variance	
I	Ad. L	.793	.057	.00070	.026	.00036	.996
	Dist. R		.050	.00061	.025	.00041	.862
II	Ad. M	.797	.054	.00037	.057	.00097	.963
	Dist. S		.056	.00015	.064	.00087	.808
III	Ad. M	.818	.066	.00021	.052	.00030	-.270
	Dist. R		.063	.0006	.044	.00079	.199
IV	Ad. L	.890	.071	.00090	.063	.00064	.787
	Dist. S		.084	.00174	.072	.00046	.602
V	Ad. H	.902	.048	.00043	.060	.00053	.973
	Dist. R		.044	.00013	.055	.00016	.890
VI	Ad. H	.929	.054	.00020	.043	.00079	.846
	Dist. S		.051	.00015	.038	.00089	.934

*The experimental cell is indicated by the level of advertising expenditures (Low = L, Medium = M, High = H) and type of distribution (Regular = R, Special = S) used. For each cell the first row indicates the expected value, and the second row gives the actual value.

Figure 1

DISTRIBUTION OF RESPONSE F (A), WHERE
F IS A CONCAVE FUNCTION OF AGE A

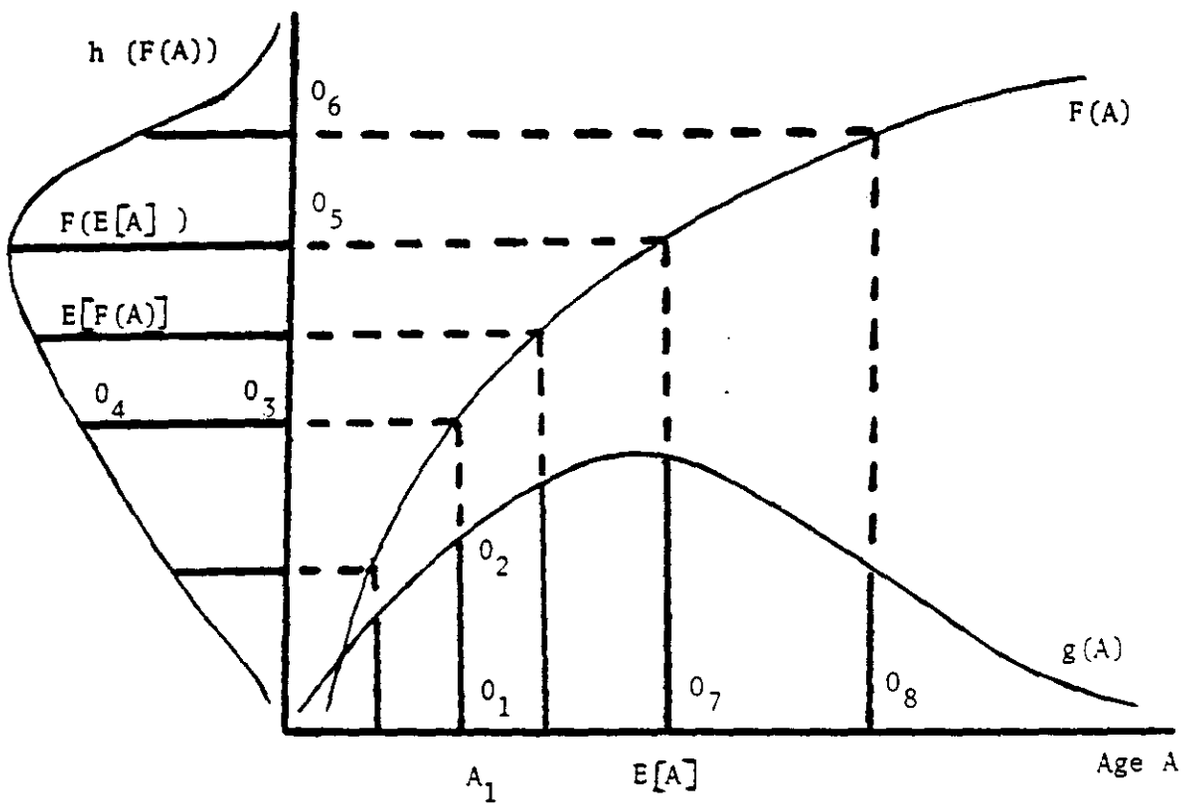


Figure 2

RESPONSE FUNCTIONS FOR S=F(A) + e

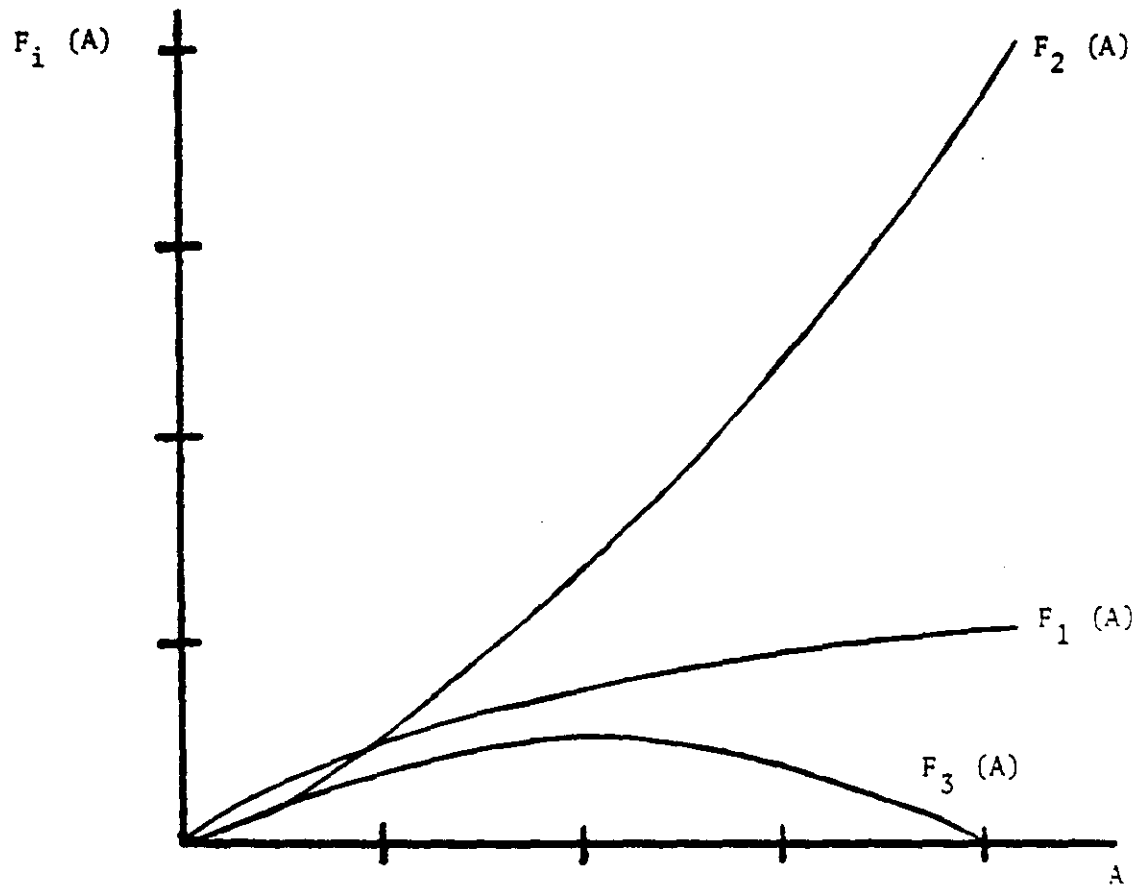
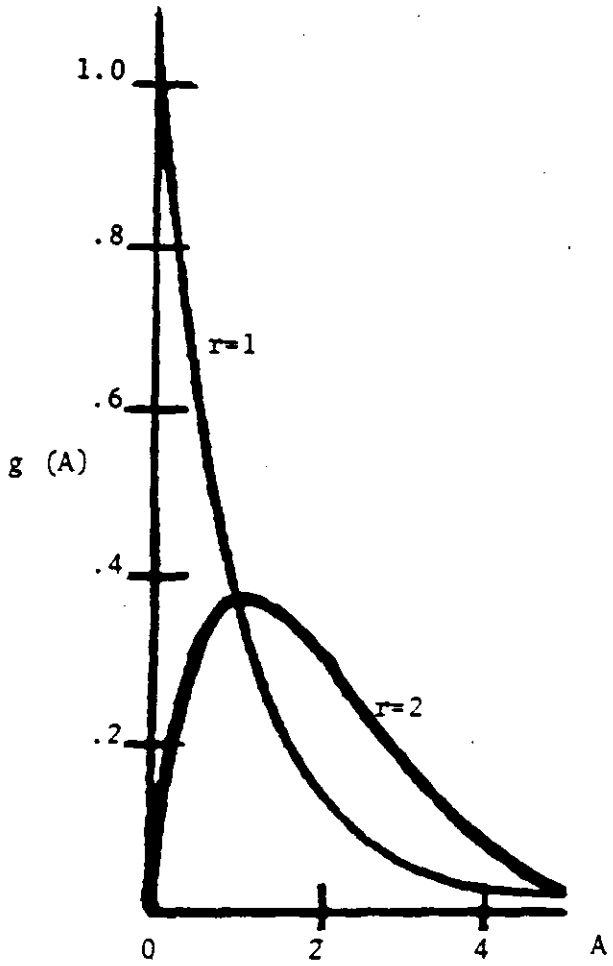


Figure 3

GAMMA DISTRIBUTION FOR THE CONTROL VARIABLE



References

Aczel, J. and Z. Daroczy (1975) On Measures of Information and Their Characterizations, New York: Academic Press.

Ali, S. M. and S. D. Silvey (1966), "A General Class of Coefficients of Divergence of One Distribution from Another," Journal of the Royal Statistical Society, Ser. B, 28, 131-142.

Blattberg, Robert and John Golanty (1978), "TRACKER: An Early Test Market Forecasting and Diagnostic Model for New Product Planning," Journal of Marketing Research, 15, 192-202.

Buchanan, Bruce (1983) "Issues in Comparative Test Design," Doctoral dissertation, Graduate School of Business, Columbia University.

Cook, Thomas D. and Donald T. Campbell (1979) Quasi-Experimentation: Design and Analysis Issues for Field Settings. Boston: Houghton Mifflin.

Dixon, W. J., ed. (1981) BMDP Statistical Software. Berkeley, California: University of California Press.

Fleming, Wendell (1977) Functions of Several Variables, 2nd edition. New York: Springer-Verlag.

Glass, G.V., P.D. Peckham and J.R. Sanders (1972) "Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analysis of Variance and Covariance," Review of Educational Research, 43, 237-288.

Goel, Prem K. (1983), "Information Measures and Bayesian Hierarchical Models," Journal of the American Statistical Association, 78, 408-410.

Goel, Prem K. and Morris H. DeGroot (1981), "Information About Hyperparameters in Hierarchical Models," Journal of the American Statistical Associaton, 76, 140-147.

Green, Paul E., Ronald E. Frank, and Patrick J. Robinson (1967), "Cluster Analysis in Test Market Selection," Management Science, 13, B387-400.

Green, Paul E., J. Douglas Carroll and Stephen M. Goldberg (1981), "A General Approach to Product Design Optimization Via Conjoint Analysis," Journal of Marketing, 45, 17-37.

Kaufman, H. and A. M. Mathai (1973), "An Axiomatic Foundation for a Multivariate Measure of Affinity Among a Number of Distributions," Journal of Multivariate Analysis, 3, 236-242.

Kullback, Solomon (1968), Information Theory and Statistics, New York: Dover Publications.

Lord, Frederic M. (1960), "Large-Sample Covariance Analysis When the Control Variable is Fallible," Journal of the American Statistical Association, 55, 307-321.

Matusita, Kameo (1966), "A Distance and Related Statistics in Multivariate Analysis," in P. R. Krishnaiah (ed) Multivariate Analysis. New York: Academic Press.

Matusita, Kameo (1977), "Discrimination and the Affinity of Distributions," in P. R. Krishnaiah (ed.) Multivariate Analysis - IV. New York: North-Holland.

McDonald, James B. and Bartell Jensen (1979), "An Analysis of Some Properties of Alternative Measures of Income Inequality Based on the Gamma Distribution Function," Journal of the American Statistical Association, 74, 856-860.

Morrison, Donald G. (1967), "Measurement Problems in Cluster Analysis," Management Science, 13, B775-780.

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

Pluta, Joseph E., Wright, Rita J. and Mildred C. Anderson (1982) Texas Factbook. Austin, TX: Bureau of Business Research, University of Texas at Austin.

Pringle, Lewis G., R. Dale Wilson and Edward I. Brody (1982), "NEWS: A Decision-Oriented Model for New Product Analysis and Forecasting," Marketing Science, 1, 1-29.

Renyi, Alfred (1961), "On Measures of Entropy and Information," Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, (Vol. 1). Berkeley, CA: University of California Press, 547-561.

Silk, Alvin J. and Glen L. Urban (1978), "Pre-Test Market Evaluation of New Packaged Goods: A Model and Measurement Methodology," Journal of Marketing Research, 15, 171-191.

Stein, Charles (1965), "Approximations of Improper Prior Measures by Prior Probability Measures," in J. Neyman and L. LeCarm (eds.) Bernoulli, Bayes, Laplace Anniversary Volume. New York: Springer-Verlag.

Urban, Glen L. (1970), "SPRINTER Mod III: A Model for the Analysis of New Frequently Purchased Products," Operations Research, 18, 805-854.

Wildt, Albert R. and Olli T. Ahtola (1978) Analysis of Covariance. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-012. Beverly Hills, CA.: Sage Publications.

Wind, Yoram (1982) Product Policy. Reading, Mass.: Addison-Wesley.

Wind, Yoram and Joseph Denny (1974), "Multivariate Analysis of Variance in Research on the Effectiveness of TV Commercials," Journal of Marketing Research, 11, 136-142.