ARE DONOR DOLLARS RELATED TO HOW LONG THEIR NAMES ARE?

Peter B. Wylie

When it comes to the world of fundraising and charitable giving, opinions abound as to what works and what doesn't in terms of getting individuals and corporations to fork over money. What is not at issue, of course, is that there are billions and billions of dollars (and equivalent foreign currencies) in this huge arena. Mulville (2000) is only one of many experts who proffer compelling evidence of this fact.

In spite of the staggering amounts of money at stake, a review of the consumer marketing literature suggests that relatively little solid scientific research has been done on the antecedents and causes of charitable giving. Some exceptions include studies by Smith (1980), Ziegler Sojka (1986), and Smith and Berger (1995).

While the work of these researchers is eminently defendable and respectable, it is firmly grounded in the hypothetico-deductive model of the social sciences. And it ignores a whole new field whose members are reluctant to call themselves scientists (Berry and Linoff, 2000), but who (in this investigator's opinion) practice a rigorous, disciplined form of applied science called data mining and statistical modeling.

But here's the rub. In spite of the huge amount of work that is done by data miners to find predictors of buying and giving in customer and donor databases, this investigator has found no published research on predictors of giving that appear to work across donor databases. There is simply no body of research that points to generalizations that can be made about donor databases. The purpose of this investigation was to put forward a small amount of evidence that augurs well for the potential of such generalizations.

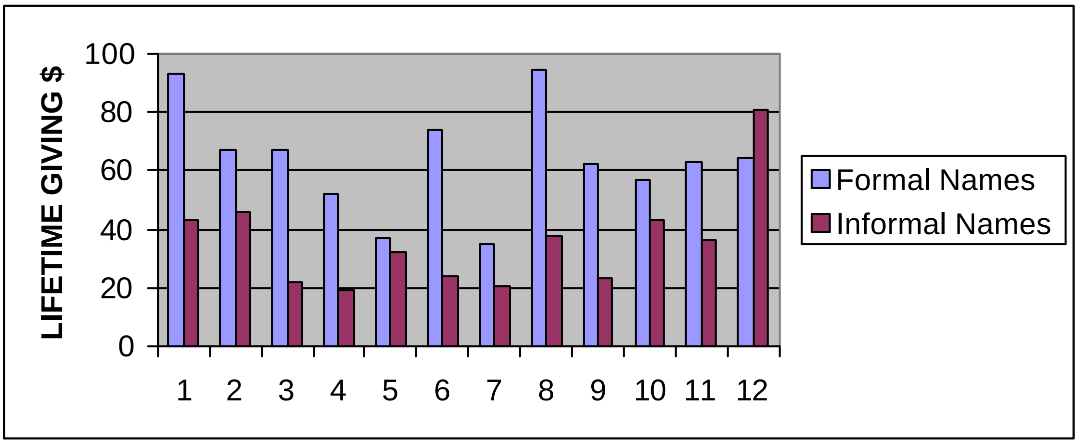
BACKGROUND

The idea for the study was born one day after the investigator had been manipulating some data for a fundraiser whose donors were largely male. In looking at the field in the database called FIRSTNAME he discovered that a number of donors had either an informal or a formal version listed of the same name. Specifically, he was able to look at small groups of the following pairs of such names:

1. WILLIAM's versus BILL's
2. ROBERT's versus BOB's
3. RICHARD's versus DICK's
4. EDWARD's versus ED's
5. KENNETH's versus KEN's
6. MICHAEL's versus MIKE's
7. RONALD's versus RON's
8. THOMAS's versus TOM's
9. DONALD's versus DON's
10. JOHN's versus JACK's
11. JAMES's versus JIM's
12. RAYMOND's versus RAY's

Would the lifetime giving amounts associated with each pair differ consistently between the formal and the informal? Figure 1 shows how they compared.

Figure 1



In 11 out of the 12 pairs (RAYMOND's versus RAY's were the only exception), the formal names gave considerably more than the informal names. The probability of this having occurred by chance (if there were no true relationship between first name length and giving for this population of donors) is less than one in four thousand.

Unfortunately, these results couldn't be replicated on any of the other donor databases the investigator was studying because most of them had only formal first names listed in the "first name" field.

Several months later, it occurred to the investigator that what he was really seeing in the formal versus informal comparison had to do with the amount of information listed about donor names. Could it be that the more information (in general) listed about a donor's name, the more he or she would give? That question provided the focus for this study.

METHODOLOGY

Ten fundraising organizations that include a variety of missions from academic development to conservation to the rights of minority groups were used to provide the databases for this investigation. From each database a probability sample, either random or systematic (every kth record), of 2,000 records but no more than 10,000 was drawn.

For each of the ten samples an outcome variable was specified. In most cases this variable was the total amount of dollars given by each person (all corporate donations were excluded) in the sample for the period (often many years) he or she had been in the database. If the total amount given was not available, either "highest gift amount" or "amount for last year" was used as the outcome variable.

These are the steps that were followed for each of the ten samples:

Step One. A field called TOTAL NUMBER OF NAME CHARACTERS was created. (It was simply a count of all the characters in any of the fields having anything to do with a donor's name including prefixes, first names, middle names, and suffixes.

Step Two. Each sample database was divided into three groups:

BOTTOM THIRD - 33% of the donors who had the fewest number of name characters in the sample

MIDDLE THIRD - 33% of the donors who had the next highest number of name characters in the sample

TOP THIRD – 33% of the donors who had the highest number of name characters in the sample

Step Three. The mean giving level and range of giving were computed for each of the three groups for all ten samples.

Step Four. A one-way analysis of variance was computed for differences among the means for each of the ten samples.

RESULTS

This study, like all those involved with comparing means of dollar amounts, was subject to the problem of the relationship between means and variances (Murphy, 1982) and extreme positive skewness. In the donor databases of almost all fundraisers, the distributions of giving dollars are non normal and asymmetrical. Most people (perhaps as many as 80%) have given nothing at all. About ten percent have given modest amounts (often less than fifty dollars), and the remaining ten percent have given amounts than can vary from less than a hundred dollars to well over ten thousand.

These distributional departures from normality and symmetry can make it difficult to achieve significant results with analysis of variance because of the assumptions of normality and homoscedasticity on which this technique is based. However, since a number of authors (Winer, 1971) have pointed to the robustness of analysis of variance with even extreme violations of these assumptions, the investigator chose to use this classical technique before resorting to transformation of the outcome variables to ranks or logarithms.

Six of the ten samples analyzed yielded significant differences (well beyond the conservative .01 alpha level) using analysis of variance of actual dollar amounts. These results are displayed in Tables 1-12. Notice that for each table in which group means are shown, the mean for the bottom third for name characters is less than the middle third which, in turn, is less than the top third.

This same patterns exists in Tables 13 –16. However, with these last two institutions (G and H), the classical one way analysis of variance did not yield a significant F value. The investigator then converted the dollar amounts given for each record in these two samples to ranks and performed an analysis of variance on the rank mean differences among the three name length groups. As Tables 14 and 16 show, the obtained F values for both these samples were highly significant.

In two of the ten samples the pattern of mean differences was the same as the other eight samples, but neither a one way analysis of the actual dollar amounts nor dollar amounts converted to ranks yielded a significant F value.

Table 1

Mean Differences Among The Three Name Character Groups for Institution A

Groups	Mean $	Range	Number of Records
BOT 3RD	44.0	1000	941
MID 3RD	51.8	2300	717
TOP 3RD	69.2	3645	785

Table 2

One Way Analysis of Variance of Mean Differences for Institution A

Source	df	Sums of Squares	Mean Square	F-ratio
Groups	2	278017	139008	8.0697 ($p<0.0003$)
Error	2440	42.0312e6	17225.9	
Total	2442	42.3093e6		

Table 3

Mean Differences Among The Three Name Character Groups for Institution B

Groups	Mean $	Range	Number of Records
BOT 3RD	133.3	550	1104
MID 3RD	147.7	870	890
TOP 3RD	156.6	870	986

Table 4

One Way Analysis of Variance of Mean Differences for Institution B

Source	df	Sums of Squares	Mean Square	F-ratio
Groups	2	290702	145351	13.072 ($p<0.0001$)
Error	2977	33.1018e6	11119.2	
Total	2979	33.3925e6		

Table 5

Mean Differences Among The Three Name Character Groups for Institution C

Groups	Mean $	Range	Number of Records
BOT 3RD	70.4	919	1450
MID 3RD	78.5	841	1758
TOP 3RD	91.1	915	1584

Table 6

One Way Analysis of Variance of Mean Differences for Institution C

Source	df	Sums of Squares	Mean Square	F-ratio
Groups	2	331602	165801	13.166 ($p<0.0001$)
Error	4789	60.3078e6	12593	
Total	4791	60.6394e6		

Table 7

Mean Differences Among The Three Name Character Groups for Institution D

Groups	Mean $	Range	Number of Records
BOT 3RD	10.8	362	1570
MID 3RD	15.1	525	2060
TOP 3RD	27.7	479	1370

Table 8

One Way Analysis of Variance of Mean Differences for Institution D

Source	df	Sums of Squares	Mean Square	F-ratio
Groups	2	223277	111638	88.98 (p<0.0001)
Error	4997	6.26948e6	1254.65	
Total	4999	6.49276e6		

Table 9

Mean Differences Among The Three Name Character Groups for Institution E

Groups	Mean $	Range	Number of Records
BOT 3RD	133.1	12300	1815
MID 3RD	145.2	12965	1965
TOP 3RD	192.3	24550	1220

Table 10

One Way Analysis of Variance of Mean Differences for Institution E

Source	df	Sums of Squares	Mean Square	F-ratio
Groups	2	2.71548e6	1.35774e6	3.3181 (p<.04)
Error	4997	2.04475e9	409196	
Total	4999	2.04747e9		

Table 11

Mean Differences Among The Three Name Character Groups for Institution F

Groups	Mean $	Range	Number of Records
BOT 3RD	56.2	1562	1159
MID 3RD	62.3	1633	999
TOP 3RD	70.7	1543	999

Table 12

One Way Analysis of Variance of Mean Differences for Institution F

Source	df	Sums of Squares	Mean Square	F-ratio
Groups	2	164082	82041.1	9.6805 (p<0.0001)
Error	4735	40.1286e6	8474.88	
Total	4737	40.2927e6		

Table 13

Mean Differences Among The Three Name Character Groups for Institution G

Groups	Mean $	Range	Number of Records
BOT 3RD	39.3	6001	1239
MID 3RD	291.7	325501	1845
TOP 3RD	1239	135101	1523

Table 14

One Way Analysis of Variance of Mean Rank Differences for Institution G

Source	df	Sums of Squares	Mean Square	F-ratio
Groups	2	143.973e6	71.9863e6	40.423 ($p<0.0001$)
Error	4604	8.199e9	1.78084e6	
Total	4606	8.34297e9		

Table 15

Mean Differences Among The Three Name Character Groups for Institution H

Groups	Mean $	Range	Number of Records
BOT 3RD	48.5	3000	1457
MID 3RD	69.2	25000	2233
TOP 3RD	71.9	5500	1692

Table 16

One Way Analysis of Variance of Mean Rank Differences for Institution H

Source	df	Sums of Squares	Mean Square	F-ratio
Groups	2	91.3368e6	45.6684e6	19.17 ($p< 0.0001$)
Error	5379	12.8089e9	2.38129e6	
Total	5381	12.9003e9		

DISCUSSION

What is not in doubt as a result of this study is whether or not there is a systematic relationship between donor name length (as recorded in fundraising databases) and amount of giving. The evidence provided here is too compelling to dispute that fact.

However, what may well be in doubt, or at least a reasonable question in the mind of the reader is “So what?” And as part of the “so what” question, couldn’t the reader also challenge these results as simply being spurious? That is, couldn't the reader offer an argument that goes something like this:

"The amount of information contained in name fields may simply be a function of the information that can be gathered from donors when they give. The more people give, the more information a fundraiser can collect on their names, titles, suffixes, etc. So the findings of this study could simply be attributed to the identification of a 'surrogate' for giving, not the identification of a useful predictor of giving."

Let's deal with the latter question first. And the most reasonable answer would seem to be "possibly." The only way to know for sure is to test whether name length predicts how much people will give in some future campaign.

Here the investigator can offer only very limited evidence. For Institution D (see Tables 7 and 8), there was an opportunity to test the predictive power of name length on a sample of 5,000 donors for a four month period of giving subsequent to the construction of the three name length groups. Specifically, the percentage of donors who gave anything at all for this period was computed for the three name length groups. These were the results:

BOT 3^{RD} 3.4%
MID 3^{RD} 3.1%
TOP 3^{RD} 5.3%
Chi-square =12.78 with 2 df, p = 0.0017

Is this unequivocal proof of the power of name length to predict donor giving behavior? Clearly, not. On the other hand, it seems encouraging, especially in a field where small percentage differences can result in enormous differences in revenue generated if hundreds of thousands of potential donors are mailed to.

But let's return to the original intent of this study and the preparation of this article. The goal of this effort was not primarily to underline the relationship between donor name length and giving behavior or even the possible predictive usefulness of this relationship. Reporting these findings was a means to a larger end. And that end was to prod researchers in the field of fundraising to start paying much more attention than they currently do to the huge potential value of information contained in donor databases. This investigator is convinced that there are consistent, lawful patterns that exist across very different types of such databases. Exploring and confirming these patterns will only help a field whose missions (for the most part) are enormously worthwhile.

This is stimulating, rewarding applied science. Let's do more of it. Lots more!

REFERENCES

Berry, M. J. A. & Linoff, G. S. (2000). *Mastering data mining: The art and science of customer relationship management*. New York: John Wiley & Sons, Inc.

Mulville, B. (2000). Corporations expect charities to show results, *Advancing Philanthropy*, 8(3), 6.

Murphy, E. A. (1982). *Biostatistics in medicine*. Baltimore: Johns Hopkins University Press.

Smith, G. E. & Berger, P. D. (1995). The impact of framing, anchorpoints, and frames of reference on direct mail charitable contributions. *Advances in Consumer Research*, 22, 705-712.

Smith, S. M. (1980). Giving to charitable organizations: a behavioral review and a framework for increasing commitment. *Advances in Consumer Research*, 7, 753-756.

Winer, B. J. (1971) *Statistical principles in experimental design*. New York: McGraw-Hill.

Ziegler Sojka, J. R. (1986). Understanding donor behavior: a classification paradigm. *Advances in Consumer Research*, 13, 240-245.

(Peter Wylie is a Washington, DC based psychologist who specializes in data mining and statistical modeling for non-profit, fundraising organizations. His e-mail address is PBRADWYLIE@AOL.COM)