A SOLUTION TO DEFINE LATENT SEGMENTS OF MEDIA AUDIENCES

Purnell H. Benson, Rutgers Graduate School of Business

Procedure

In media analysis, data are supplied for the proportions of the population reached by different vehicles of communication within a limited number of exposure opportunities. From these data, projections are made of the exposures which result with combinations of vehicles whose opportunities for exposure extend over substantial periods of time. These projections are based upon various probabilistic models as to how the frequencies of exposure are numerically related from vehicle to vehicle and exposure to exposure. The Beta function (Greene and Stock, 1967; Hyett, 1958) and the bivariate normal distribution (Benson, 1969) are two such approaches.

Let us assume that the population which has exposure opportunities is made up of segments whose sizes are $X_1, \ldots X_n$ with independent probabilities $P_1, \ldots P_n$ of these segments being exposed to an advertising vehicle in a single exposure opportunity. We then consider what takes place when opportunities for exposure to pairs of vehicles (or for two successive exposures of the same vehicle) exist. The four possibilities are for exposure to both vehicles, the first but not the second, the second but not the first, and exposure to both vehicle opportunities.

Since we assume that probabilities of exposure within each segment are independent from opportunity to opportunity, each of the four combinations of exposure and non-exposure has a frequency of occurrence which is equal to the sum of the binomially combined probabilities within segments. A system of equations relates each of the four cell frequencies for each pair of vehicles to corresponding algebraic terms for the X's and the P's. From the system of equations, the corresponding X's and P's can be calculated, if sufficient combinations of vehicles and exposure opportunities are included.

If refinement of the solution is sought through using a large number of segments, the iterative solution for both X's and P's is not rapid. Alternatively, a large number of segments of equal size can be assumed. Then the system of equations is of second degree only. For M segments, the equations take the following form:

$$MC_{00.ij} = (1 - P_{k.i})(1 - P_{k.j}) + \ldots + (1 - P_{m.i})(1 - P_{m.j}),$$

$$MC_{10.ij} = P_{k.i}(1 - P_{k.j}) + \ldots + P_{m.i}(1 - P_{m.j}),$$

$$MC_{01.ij} = (1 - P_{k.i})P_{k.j} + \ldots + (1 - P_{m.i})P_{m.j},$$

$$MC_{11.ij} = P_{k.i}P_{k.j} + \ldots + P_{m.i}P_{m.j}$$

C_{00}, C_{10}, C_{01} and C_{11} refer to cell frequencies for non-exposure to either vehicle I or J, exposure to I but not J, exposure to J but not I, and exposure to both I and J. Subscripts i and j refer to the pair of vehicles, and subscripts <u>k</u> and <u>m</u> refer to audience segments.

The solution uses a full set of trial values of variables to solve for the same variables as linear multipliers. The matrix of equations is repeated a second time in the system for solution in order to permit solving for the same variables as linear multipliers in one-half of their involvement in the equations. Approximation proceeds by averaging the trial values with the corresponding solution values. Faster convergence occurs if the average is obtained by taking the square root of the product of the trial and solution values.

Application to Media Exposure Data

Through the courtesy of W. R. Simmons, syndicated data for exposure of adults to six magazines were made available for analysis. These data give the proportions of consumers exposed to each magazine in a single opportunity, and the proportions exposed to one or both of the exposure opportunities for pairs of magazines, including a magazine paired with itself for two successive exposure opportunities. From this information frequencies in each of the 4 cells in the two-exposure tabulations are determined.

Six vehicles paired in all combinations provide 57 independent observations, 3 for each of the 15 pairs of vehicles and 2 for each of the 6 vehicles paired with itself. Before using these cell frequencies as entries for the dependent variable, residual constants on the right hand side of the equal sign were transposed and combined with the entry for the dependent variable. For instance, cell C_{00} which is set equal to the products of ones minus probabilities has combined with it those terms which do not provide an entry for the multipliers of the independent variables.

If we use 8 segments for which probabilities of exposure for each of 6 vehicles are to be found, the number of unknowns is then 48, within the 57 degrees of freedom. In addition, there are segments of the population who do not see any of the 6 magazines in a single exposure. This residual group comprises approximately 40% of the population. These may be represented by an additional 5 segments for whom the exposure probabilities are set equal to zero throughout the solution. The probability entries for these are also transposed and combined with entries for the dependent variable.

To assemble the input for multiple regression analysis to define the probabilities of exposure of each segment to each vehicle, a machine program was prepared. The output from this program is the matrix of entries for multiple regression analysis. Values are bounded by 0.01 before averaging and by 1.0 after averaging.

The matrix of trial probabilities for exposure of the 8 active segments and the 5 inactive segments is given in Table 1. Inspection of Table 1 indicates which segments of the population share common exposure to different magazines.

A practical application of the definition of latent segments is to use the associated exposure probabilities to make projections for exposures achieved by media schedules with multiple vehicles and multiple insertions. Existing methods based upon the Beta distribution are inadequate when exposure patterns for the audiences of different vehicles are heterogeneous. The use of probabilities for latent segments may prove to be a useful and inexpensive way out.

As a check upon the efficiency with which exposure projections can be made, the frequency pattern for 8 exposure opportunities was calculated for each of the 6 magazines. The results are compared with the usual Beta projections in Table 2. Since Beta analysis is

Table 1

INITIAL TRIAL VALUES AND PROBABILITY VALUES AFTER FIVE CYCLES OF ITERATION

		Segment												
Vehicle	Cycle	1	2	3	4	5	6	7	8	9*	10*	11*	12*	13*
1	1	.90	.20	.20	.20	.20	.20	.20	.20	.00	.00	.00	.00	.00
	5	.33	.00	.96	.63	.62	.30	.00	.00	.00	.00	.00	.00	.00
2	1	.20	1.00	.20	.20	.20	.20	.80	.90	.00	.00	.00	.00	.00
	5	.22	.22	.46	.35	.66	.24	1.00	.48	.00	.00	.00	.00	.00
3	1	.20	.20	1.00	.20	.20	.20	.90	.80	.00	.00	.00	.00	.00
	5	.15	.01	.57	.22	.19	.22	.96	.72	.00	.00	.00	.00	.00
4	1	.20	.20	.20	.90	.20	.20	.20	.20	.00	.00	.00	.00	.00
	5	.38	.02	.24	.90	.00	.44	.02	.16	.00	.00	.00	.00	.00
5	1	.30	.30	.30	.30	1.00	.50	.30	.90	.00	.00	.00	.00	.00
	5	.47	.10	.24	.72	.90	.62	.04	.99	.00	.00	.00	.00	.00
6	1	.20	.20	.20	.20	.50	.90	.40	.40	.00	.00	.00	.00	.00
	5	.00	.02	.25	.30	.61	.82	.45	.02	.00	.00	.00	.00	.00

* These values are set as fixed during the iterations.

Table 2

ESTIMATION OF PERCENTAGES REACHED BY EACH OF SIX ADVERTISING VEHICLES AFTER EIGHT EXPOSURE OPPORTUNITIES, USING INITIAL TRIAL PROBABILITIES AND EXPOSURE PROBABILITIES ESTIMATED AFTER FIVE CYCLES OF ITERATION

Vehicle	Beta Calculation	Initial Values		Values after 5 Cycles	
		Estimate	Error	Estimate	Error
1	34	53	+19	38	+4
2	63	55	- 8	58	-5
3	51	55	+ 4	49	-2
4	32	53	+21	38	+6
5	55	59	+ 4	52	-3
6	32	56	+24	39	+7

considered acceptable for making projections for a single magazine, this provides a test of latent segment projection. The results look encouraging for what may be accomplished with larger numbers of segments and vehicles.

Relation to Lazarsfeld's Work

The historical debt to and the inspiration from the work of Lazarsfeld for the analysis reported here is large. His creation of latent structure analysis two decades ago is one of methodological landmarks of our century. (Stouffer et al, 1950). I have returned to this type of work with somewhat different reasons and emphasis than his, being first preoccupied with media scheduling.

Some of the procedural differences between Lazarsfeld's development of latent structure analysis and the lines I pursue may be kept in mind.

1. The latent segment framework introduced here for analysis of media exposure assumes a multiplicity of latent segments of equal size, some of which may be regarded as duplicates. The degree of the system of equations is correspondingly reduced.

2. A paired matrix of probability cross-products of no higher than order two is used, leading to a quadratic system of simultaneous equations. The solution to these is relatively straightforward, involving a progressive averaging of trial and obtained values. The multiplicity of solutions appears to be equivalent to factorial combinations of segments, all of which yield sets of roots which are identical except for factor designation.

3. All four cells of the 2-by-2 cross tabulation of probabilities P_i and P_j for two classes of response and non-response are introduced, including $(1 - P_i)(1 - P_j)$, $(1 - P_i) P_j$, $P_i (1 - P_j)$, and $P_i P_j$. Thus far the fit of probabilities is limited to dichotomous response categories.

4. The model is fitted by multiple regression of the observed cell frequencies, instead of an exact solution with the number of unknowns equal to the number of degrees of freedom in the system. This may help to cope with problems of fallible data.

5. No conclusions are yet drawn concerning the latent structure of factors to account for the segment probabilities of media exposure. This is work for the future. Doubtless an endless variety of multiple factor structures can be investigated in relation to the segment probabilities which are calculated. All types of behavioral response, including brand behavior, seem amenable to latent structure analysis.

Conclusion

The present inquiry was undertaken to establish the probabilities of exposure to magazines for population segments.

The segments are latent in that their exposure probabilities are computationally inferred, not actually observed. These segments are useful as a means of representing the population of magazine readers or any population exhibiting dichotomous responses of various kinds.

The conceptual representation of latent segments in terms of probabilities of response is a useful starting point in analyzing the relationship of segment response to other population variables.

One form of such analysis is to infer the structure of factors which accounts for the segment exposure probabilities. Another line of analysis is to relate exposure probabilities to observed personality and population variables. One practical application is to use the exposure probabilities for segments to project reach and frequency of exposure for media schedules with multiple insertions over a period of time. This is also a test of the validity of latent structure analysis. Another application is to establish more precise trends in brand buying behavior for evaluating market sales. The machine program is available to others on a shared cost basis. There are many opportunities to be explored.

REFERENCES

Benson, P. H. Bivariate normal distribution to calculate media exposure. Journal of Advertising Research, 1969, 9, 41-47.

Greene, J. and Stock, J. S. Advertising Reach and Frequency in Magazines. New York: Readers Digest Association, 1967.

Hyett, G. P. The measurement of Readership. Paper read to the Statistics Seminar, London School of Economics, February, 1958.

Stouffer, S. A., Guttman, L., Suchman, E. A., Lazarsfeld, P. F., Star, S. A., and Clausen, J. A. Studies in Social Psychology in World War II, vol. 4, Measurement and Prediction. Princeton: Princeton University Press, 1950; Lazarsfeld, P. F. and Henry, N. W. Latent Structure Analysis. New York: Houghton-Mifflin Co., 1968.