An Index Method for Selecting Representative Cities of a Target Market

Sabry A. Shehata

An index was developed to rank U.S. metropolitan areas according to overall similarities in their socio-economic characteristics to the U.S. average. This index is given as a solution to matching a representative city or market area to a target market as well as matching a group of cities which are similar to each other. The index shows stability in ranking these cities with different consumption data for similar products and proves to be a promising approach whenever the researcher must find a group of cities which are similar to each other and to the target market.

Market research sometimes requires selection of cities or market areas for testing new product performance, effects of changes in different market variables on product sales, effects of changes in different economic variables on product consumption, and so on. This type of research uses representative markets as proxies for a target market area in order to get meaningful information [Shoraka, R.]. A relatively large sample size and repetitive market research are needed to obtain precise results; however, the larger the sample size, the greater are research costs [Ladik, et al.]. In many cases, the cost of increasing the sample size (number of cities) is not justifiable or may be prohibitive to the small firm undertaking the analyses [Lipstien, B.]. That is why a representative market area or city is still a practical approach in marketing research [Scott, F., Jr.].

It is difficult to find a group of markets or cities which are similar to each other and to the target market areas in all socio-economic characteristics [Scott, F., Jr.]. The conventional method of selecting representative market areas is to examine the demographic characteristics of different cities and to select the one that most closely resembles the target market [Scott, F., Jr.]. Since the effect of demographic variables on the consumption of a commodity may differ for each commodity, the impact of various demographic variables on the consumption of a commodity may differ from one variable to another [Kohls and Downey]. Therefore, different weights must be assigned for various demographic variables in the process of selecting representative areas. Then the total index is estimated by summing the weight scores of individual demographic variables. The index proposed here offers a means for ranking cities or individual market areas according to their overall similarities to the target market areas in the total index.

The Index

Like other indices, this index is developed for comparative analysis, and to rank various United States cities relative to each other and to the national market [Freund, J. and F. J. Williams]. The index procedure relates the demographic characteristics of cities to the sales or the consumption of various commodities, and different weights are assigned to these characteristics. The main objective of this index is to select a representative city to the target market. Therefore, the optimum condition exists if a city has demographic

S. A. Shehata is assistant professor of Agricultural Economics University of Hawaii at Hilo.

Special acknowledgments are given to Frank Scott, Jr., Heinz Spielmann, Gary Vieth and an anonymous reviewer of WJAE, for their valuable comments.

characteristics identical to those that exist in the target market. The optimum is defined as the value of these variables in the target market. Since these characteristics are measured in different ways, it is left to researchers to decide on the type of measurement to be used.

Assumptions

Utilization of the city index requires the following assumptions. The first assumption is that the market area selected as representative of the target market has the highest score in the index. The second assumption is that the values of socio-economic characteristics distribute normally around the target market parameters. Given this assumption equal scores are given to cities with equal absolute deviations.

The third assumption is that statistical coefficients give different weights to the socio-economic variables according to the importance of their effects on sale or consumption of the commodities. The larger a weight's absolute value the higher the degree of association a particular variable has with the consumption of the product. This assumption is needed in order to assign different weights for different socio-economic variables. Therefore, researchers can rank cities according to the similarity of key variables to the target market. If one city deviates more than another city in important variables, its score is lower. Given this assumption, it is not necessary to discriminate against such variables.

Various statistical coefficients satisfy the third assumption. These are: the absolute value of simple correlation coefficients, the absolute value of partial correlation coefficients, principal components coefficient, and the coefficient of determination. All these coefficients measure the degree of association between two or more variables, and the larger the absolute value, the higher the degree of association and, therefore, the importance of this variable relative to the other variables. In order to weight these variables relative to each other, the absolute value of the coefficient is divided by the summation of the statistical coefficients. In this way, the third assumption is met.

The fourth assumption is that the selected statistical coefficients are stable during the projection period. This assumption assures the applicability of the index. If this assumption does not hold, the estimated index will change over time and not offer an improvement to market research.

The fifth assumption is a behavioral one, which states that a researcher selects a city with minimum deviation, subject to the researcher constraints. These constraints may be availability of secondary data, funds, location, population, economic activity, weather, and so on. This assumption explains why some representative cities may be overlooked. If a sufficiently large number of cities are examined, however, one can expect characteristics of some to resemble those of the target area.

The Index in Mathematical Form

The City Index (CI) is composed of the Weight Index (W) and the Score Index (S). The weight index attaches a weight to various demographic variables. The absolute values of the correlation coefficients are utilized to arrive at such weight. The larger the absolute value, the larger the degree of association of these particular variables with the consumption of the goods whose market is being analyzed.

Then the absolute summation of these coefficients is used to weight these variables in relation to each other. The simple correlation coefficient meets assumptions three and four which imply the stability of the index; the larger the coefficient, the larger the degree of association and the more important is this variable's effect on the consumption of the product.

The weight index is developed to give different weights for different variables, and is defined as follows:

$$(1) \qquad W_{i,c} = \frac{r_{i,c}}{\Sigma r_{i,c}}$$

$$W_{i,c} \leq 1.0 \text{ and } W_{i,c} \geq 0$$
$$i = 1,2 \text{ -- } n$$
$$c = 1,2 \text{ -- } m$$

Where: $W_{i,c}$ = weight index for demographic variable "i" at a particular city, for particular commodity "c"; and $r_{i,c}$ = the correlation coefficient of variable "i" with the per capita consumption of a particular commodity "c".

The Score Index (S) presents the relationship between the value assigned to a city for a particular variable and its deviation from the optimal value where the optimum value of a variable is defined as its value at the target market area.

Suppose that a researcher selects a city with a median level of income of $11,000 per capita; a city with no deviation receives a maximum score. A city whose median income deviates positively or negatively from $11,000 is assigned a score less than the maximum. The positive and negative deviations affect the decision outcome equally. This relationship is presented as follows:

$$(2) \qquad S_i = f[\,|\,d_i - d\,|\,/sd_i]$$

Where S_i = city score in a particular demographic variable "i"; d_i = the value of particular variable "i" in the city; d = the value of the same demographic variable at the optimum or target market; sd = the estimated standard deviation of demographic variable "i"; | | = absolute value sign; and i = 1, 2, ... n

In equation 2), the absolute value of the deviation of a city is divided by the standard deviation to standardize the units of measurements, so that the scores from various demographic variables can be added together (income in dollars, education and age in years, ethnic as a percentage, and so on).

Since marketing researchers seek information about product sales, any market information obtained from a city about the sale or consumption of a product is better than no information at all. Therefore, the lower limit (S) of a particular variable will approach zero, and the upper limit (S) is one. In order to satisfy this condition an exponential function is appropriate. The score index is specified as follows:

$$(3) \qquad S_i = e^{-|\,d_i - \bar{d}\,|/sd_i}$$

In order to find the city index value, the score of each demographic variable is weighted. The weighted scores are then summed to form the city index value, CI, as in equation 4:

$$(4) \qquad CI = \Sigma W_i \cdot S_i$$

Index Application

The index was used to select cities representative of the U.S. market for fresh pineapple. Data were gathered in a visitor survey which assessed consumer preferences for fresh pineapple [Shehata].

The following conditions were required: a) *medium-size city*: large cities may better represent the national market in various economic activities, but research costs of studying these cities would be high. Small cities were eliminated from the study because their economic activities are less representative of national market; b) *Western region of the United States*: a city in the western region of the United States was selected to test if consumer preferences for fresh pineapple are the same as for Hawaii visitors from the same region; c) *East-Central city:* a city in the east-central part of the United States was selected to test: (1) if consumer preferences for fresh pineapple are the same as Hawaii visitors from the same region, (2) the consumer awareness of the different types of fresh pineapple supplied by producing areas other than Hawaii, and (3) the effect of a

Hawaii visit on the consumption of the fruit; and d) *availability of funds*, availability of funds for research influences the number of cities chosen.

Data on demographic characteristics of 125 standard metropolitan areas of the United States were obtained from the 1970 census [Bureau of Census]. Since fresh pineapple is relatively new to the United States mainland markets and data concerning consumption is not reliable; alternate products for which data were more reliable were utilized to test the ranking procedures. Products selected were apples, bananas and oranges. Estimates of per capita consumption were made from fresh fruits and vegetable unloads data for 40 cities (USDA, 1970).

Results

The results indicated that different demographic variables had different influences on consumption. Moreover the consumption of commodities correlates differently with the same demographic variable (Table 1). Income, white, black and age variables are more important than education and other ethnic variables in the consumption of these commodities. Therefore those variables were weighted heavier than education and other ethnic variables. In order to get the weight index (W), the individual correlation-coefficients were divided by the summation of correlation coefficients as indicated in equation 1).

In order to estimate the score index (S), the standard deviations (sd) of education, age, income and ethnic distribution were estimated. The difference between the actual value of a particular variable in a city, d_i, and optimal value of the same variable at the national level (d) was estimated, and the results were divided by (sd) as indicated in equation 3). The results were multiplied by the individual weights given for each demographic variable (W_i). The sum of the products gives the CI.

The Kansas City metropolitan area (Table 2) appears to be the most representative of the U.S. national market for consumption of apples, bananas and oranges, followed by Indianapolis, Indiana. The ranking of the 125 cities was similar for all commodities (Table 3).

Two representative U.S. mainland cities were chosen for testing the uses of the visitors' survey in determining consumer preferences for fresh pineapple, using the previous index [Shehata]. Those conditions stated before eliminated group "A" cities. Sacramento, California and Cleveland, Ohio, both of which were in the "B" group were chosen. Sacramento is in the western region and has a warm climate. Cleveland has a colder climate, and is supplied by different

TABLE 1. Correlation Coefficients Between Per Capita Consumption of Fresh Fruits and Demographic Variables and the Weight Assigned to Each Variable, 1976

	Median Level of			Ethnic Distribution			
	Education	Age	Income	White	Black	Others	Total
			Correlation Coefficients*				
Fresh Fruits							
Apples	.09	.19	.42	.42	.38	.06	1.59
Bananas	.03	.36	.28	.22	.22	.03	1.14
Oranges	.12	.18	.26	.22	.21	.01	1.00
		Weight (W)					
Apples	.06	.12	.27	.26	.26	.04	1.0
Bananas	.03	.32	.25	.19	.19	.02	1.0
Oranges	.12	.18	.26	.22	.21	.01	1.0

*absolute value of the coefficients are presented

TABLE 2. Estimating the Score Index for Kansas City Metropolitan Area Using Various Fresh Fruits, 1976

Demographic Variables	W_i			Optimum Value of Selected Demographic Variables ($\bar{d}$)	The Value of the Variable at Kansas City (d_i)	sdi*	$di - \bar{d}$	Si	$S_iW_i \times 100$		
	Apples	Bananas	Oranges						Apples	Bananas	Oranges
Median Level of Education	.06	.03	.12	12.1	12.2	.43	.1	1.79	4.8	2.4	9.5
Median Level of Age	.12	.31	.26	28.1	28.1	2.167	0	1.0	12.0	32.0	18.0
Median Level of Income	.27	.25	.26	$11,012	$10,653	1050.8	.36	.71	19.2	17.8	19.2
Percentage of Whites	.26	.19	.22	86.59	87.4	8.57	.81	.91	23.7	17.1	19.8
Percentage of Blacks	.26	.19	.21	12.03	12.05	8.81	.02	.98	25.6	18.6	20.6
Percentage of Others	.04	.02	.01	1.48	.55	1.19	.85	.46	1.8	.9	.5
									87.0	88.8	87.6

*(estimated for 41 cities); data obtained from U.S. Census, 1970.

TABLE 3. Ranking and Grouping the Standard Metropolitan Areas of the U.S.A. Using Fresh Oranges and Simple Correlation Index Method, 1976

GROUP A			GROUP B			GROUP C			GROUP D		
CITY & STATE	SCORE	RANK	CITY & STATE	SCORE	RANK	CITY & STATE	SCORE	RANK	CITY & STATE	SCORE	RANK
Kansas City, MO-KAN	87.62	1	New York City, NY	65.22	21	Salinas, CAL	47.74	69	Albuquerque, NM	28.64	108
Indianapolis, IND	82.26	2	Rockford, ILL	63.67	22	Phoenix, ARIZ	47.52	70	El Paso, TEX	28.12	109
New Haven, CONN	81.70	3	Louisville, IND	63.64	23	Grand Rapids, MICH	47.36	71	Wilke-Barre, PA	27.56	110
Los Angeles, CAL	80.32	4	Peoria, ILL	62.97	24	Greenville, SC	47.00	72	Columbia, SC	27.31	111
Wilmington, DEL-NJ-MD	78.96	5	Philadelphia, PA	62.92	25	West Palm Bch, FLA	46.97	73	Norfolk, VA	27.21	112
Dayton, OHIO	77.23	6	San Bernardino, CAL	62.75	26	Oxnard, CAL	46.70	74	Reading, PA	26.42	113
Trenton, NJ	75.47	7	Youngstown, OHIO	62.45	27	Chattandoga, GA	46.28	75	Baton Rouge, LA	24.74	114
Toledo, OH-MICH	73.93	8	Dallas, TEX	62.35	28	Knoxville, TENN	45.97	76	Honolulu, HI	24.29	115
Las Vegas, NEV	73.81	9	Cleveland, OHIO	61.01	29	Salt Lake City, UTAH	45.85	77	Charleston, SC	24.11	116
Rochester, NY	73.53	10	Sacramento, CAL	60.89	30	Utica-Rome, NY	45.74	78	Corpus Christi, TEX	23.62	117
Columbus, OHIO	71.23	11	Spokane, WASH	60.49	31	Jersey City, NJ	45.99	79	Birmingham, ALA	22.13	118
Milwaukee, WIS	71.00	12	Seattle-Everett, WASH	60.26	32	Lansing, MICH	44.36	80	Augusta, GA-SC	19.26	119
Southbend, IND	70.48	13	Worcester, MASS	60.20	33	Pittsburgh, PA	43.86	81	New Orleans, LA	18.23	120
Akron, OHIO	70.31	14	Syracuse, NY	60.09	34	Tampa, FLA	43.81	82	Johnstown, PA	17.94	121
St. Louis, MO-ILL	70.23	15	Springfield, CONN	59.73	35	Washington, D.C.	43.81	83	Jackson, MISS	16.78	122
Flint, MICH	67.97	16	Des Moines, IOWA	59.25	36	Houston, TEX	43.74	84	Charleston, SC	16.45	123
Orlando, FLA	67.08	17	Chicago, ILL	59.23	37	Bridge Port, CONN	43.06	85	Mobile, ALA	14.48	124
Fort Worth, TEX	66.68	18	Davenport, ILL	58.84	38	Portland, ORE-WASH	42.93	86	Memphis, TENN	13.30	125
Cincinnati, OH-KY-IND	65.34	19	Buffalo, NY	58.50	39	Fresno, CAL	41.38	87			
Tulsa, OKLA	65.32	20	Fort Lauderdale, FLA	57.18	40	San Jose, CAL	41.18	88			
			Gary-Hammond, IND	56.98	41	Greensboro, NC	41.13	89			
			Albany, NY	56.82	42	San Antonio, TEX	41.11	90			
			Stockton, CAL	56.65	43	Appleton, WIS	40.35	91			
			Hartford, CONN	55.66	44	Bakersfield, CAL	40.22	92			

TABLE 3. (Continued)

Oklahoma City, OKLA	55.36	45	Paterson, NJ	40.06	93
Wichita, KAN	55.33	46	Lancaster, PA	39.34	94
Atlanta, GA	54.81	47	Richmond, VA	38.98	95
Denver, COLO	54.55	48	Minneapolis, MINN	38.38	96
Santa Barbara, CAL	53.54	49	Madison, WIS	37.57	97
San Diego, CAL	53.44	50	Beaumont, TEX	36.82	98
Little Rock, ARK	53.25	51	Shreveport, LA	36.42	99
Fort Wayne, IND	52.49	52	Charlotte, NC	35.57	100
Boston, MASS	52.35	53	Duluth, WIS	35.06	101
Tacoma, WASH	52.33	54	Jacksonville, FLA	34.01	102
Lorain-Elyrie, OHIO	52.20	55	Providence, RI-MASS	33.81	103
Canton, OHIO	52.04	56	York, PA	32.95	104
Harrisburg, PA	51.97	57	Newport News, VA	32.53	105
Omaha, NEBR-IOWA	51.60	58	Anaheim, CAL	31.60	106
San Francisco, CAL	50.96	59	Allentown, NJ	30.83	107
Austin, TEX	50.86	60			
New Ark, NJ	49.52	61			
Tucson, ARIZ	49.29	62			
Miami, FLA	48.96	63			
Baltimore, MD	48.92	64			
Nashville, TENN	48.69	65			
Binghamton, PA	48.28	66			
Detroit, MICH	48.22	67			
Erie, PA	48.14	68			

sources of fresh pineapple. Both cities are medium in size.

Within the "A" group, a researcher can select cities which are representative of each other and of the national market in the total index. Given assumption five the researcher selects from these cities within the framework of market and environmental constraints. If the researcher wants cities which are representative of the national market and of each other in the Mid-West, then Kansas City, Indianapolis, Flint, Columbus, St. Louis, Akron, and Cincinnati are preferred. However, should the researcher wish to select a western city to represent the national market, the cities of choice change, and in this case, for example, Kansas City is not the city of choice.

Summary

The index provides another way to assist researchers in selecting a representative city to the target market from which results obtained from controlled market experimentation or consumer survey about various products can be extrapolated.

As long as a researcher collects the relevant information about demographic characteristics of various cities and utilizes appropriate statistical coefficients, then the weight and the rank can be constructed and the index can be applied. This index can be applied for any commodity to select cities representative of target areas. It should help researchers to base their city selection on a mathematical model rather than the conventional method of selection. This way market researchers may obtain an accurate result at a reasonable cost.

Researchers will be able to rank those cities according to their similarities to each other. This way, differences in results obtained from the test areas will be attributed to factors other than differences in demographic characteristics.

References

Freund, E. John, and F. J. Williams, *Elementary Business Statistics, The Modern Approach*, 3rd Ed., Prentice-Hall, Inc., New Jersey, p. 59.

_____, *Elementary Business Statistics, The Modern Approach*, 3rd Ed., Prentice-Hall, Inc., New Jersey, p. 252.

Kohls, R., and W. David Downey, *Marketing of Agricultural Products*, 4th Ed., MacMillan Publishing Co., Inc., New York.

Ladik, F., et al. "Testing Marketing of New Products," *Journal of Marketing Research*, Vol. 24, No. 4, April, 1980.

Lipstien, B., "The Design of Marketing Experiments," *Journal of Advertising Research*, December, 1965.

Scott, F., Jr., "Some Use and Limitation of Control Experiments in Marketing Research." Proceedings, Market Research Committee, Western Agricultural Economic Research Council, University of California, November, 1960.

_____ and Rahim Shoraka, "Economic Analysis of the Market for Guava Nectar," HAES, College of Tropical Agriculture, University of Hawaii, Research Report 230.

Scott, F., Jr., "An Economic Analysis of the Market for Frozen Guava Nectar Base - Buying Patterns and Potential Sales," Agricultural Economics Bulletin 14, University of Hawaii, HAES, September, 1958, p. 21.

Shehata, Sabry, "Predicting the Market Potential for Fresh Pineapple Using the Survey Approach." Ph.D. dissertation, University of Hawaii, August, 1977.

Shoraka, R., "The Determination of the Market Potential for a New Product: A Case Study of Guava," Ph.D. dissertation, University of Hawaii, December 1973.

U.S. Department of Agriculture, Marketing Service, Fruit and Vegetable Division, Market News Branch, 1960-1970.

U.S. Department of Commerce, Social and Economic Statistics Administration, Bureau of the Census, 1970.

Van de Geer, J. P., *Introduction to Multivariate Analysis for the Social Sciences*, W. H. Freeman, San Francisco.