James M. Sakoda and James P. Karon, Brown University

What we have to say is not particularly novel or earthshaking, but may be of interest and use to those engaged in enterprises involving the construction of social indicators. The use of orthogonal contrasts makes it possible to construct indices which are independent of one another and free of spurious correlations. It also permits the creation of more than one index from a variable. We have applied this approach in part to an analysis of 212 census tracts in Rhode Island, using Count 4 census tract data from the 1970 census of population and housing. We constructed a variety of indices and subjected the correlation coefficients among them to a cluster analysis procedure. On the basis of the composition of the various clusters we have concluded that many of the social indicators we have constructed are meaningful and that the use of the orthogonal contrast approach to index construction helps to create a variety of indices from a single variable.

There are two alternatives to use of orthogonal contrasts. The first is to calculate an overall index such as median education, median income, median age as a single index for a continuous variable. The use of orthogonal contrast applies to variables consisting of three of more categories, but continous variables can still be broken up into several categories, and hence one can elect either approach. When several categories of a variable are set up or are available in the original data source, there is a variety of combinations of proportions or ratios which can be constructed. A standard approach is to construct k dummy variables for a k category variable. Dummy variables are traditionally calculated by scoring one category as 1 and the remaining ones as zero, which is essentially the same as calculating a proportion in each category based on the total number in all categories. With three categories, such as single, divorced or separated and married we can construct three proportions:

proportion single, proportion divorced or separated and proportion married. The difficulty with the dummy variable approach is that there is a spurious negative correlation among the proportions, since they must add up to 1.0. In multiple regression this lack of independence is taken care of by dropping one of the dummy variables and letting the multiple regression procedure adjust the effects of negative correlations. Such variables, however, are often not suitable as social indicators, since each variable reflects the combined effects of the remaining variables. In factor analysis the spurious negative correlations can result in bipolar factors, which are not meaningful.

Setting up orthogonal contrasts involves starting with or creating several categories, such as 5 or 10 year age groupings, but it is desirable to using groupings which are meaningfully different. For example, one can break up age into life cycle categories. We selected four categories, 0-5, 6-24, 25-64 and 65 and over. These groupings correspond to young children, youths, active adults, old people. Orthogonal contrasts involve setting up a set of k-1 independent comparisons, and can be checked most easily for orthogonality by setting up weights so that they add up to zero and also so that the sum of crossproducts of weights add up to zero. One standard approach to orthogonal contrast is to pick one variable and compare it with the remaining categories, drop that category and pick another category to compare with the remaining ones, and continue in this way until only a single proportion is justified.

0r	thog	onal	Table Comparisons	1. for	Three	Categories
<u>a</u>	<u>b</u>	<u>c</u>	Prop	ort	lons	
+2	-1	-1	a /	(a -	- b + d	2)
0	+1	-1	ъ /	(b +	+ c)	

In Table 1 is shown the weights and proportions for three categories, and in Table 2, for four categories. For four categories <u>a</u> is compared with a combination of <u>b</u>, <u>c</u>, and <u>d</u>; <u>b</u> with a combination of <u>c</u> and <u>d</u>; and <u>c</u> with <u>d</u>.

				Table 2	•				
Ort	hogo	nal	Comp	arisons fo	r F	our	Categ	gories:	I
<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	Propo	rti	ons			
+3	-1	-1	-1	a / (a +	ън	+ c +	d)	
0	+2	-1	-1	Ъ/(b +	сł	⊦d)		
0	0	+1	-1	c / (c +	d)			

Each proportion has a different base denominator term. There is a discussion of orthogonal contrasts in both Snedecor and Blalock. The order of picking and dropping a category is the choice of the user, and he should pick the order that he considers to be most meaningful.

				Table 3.
Ort	hogo	nal	Compa	risons for Four Categories: II
a	<u>b</u>	<u>c</u>	<u>d</u>	Proportions
+1	+1	-1	-1	(a + b) / (a + b + c + d)
	•	0	-1	(a + b) / (a + b + c + a)
τı	-T	0	0	a / (a + b)
0	0	+1	-1	c / (c + d)

For four categories or more there is another possible approach, shown in Table 3. That is to divide the categories into two groups and then take each half separately and proceed to pick orthogonal contrasts. This approach we used for our AGE variable and is shown in Table 4. With four categories there is another possible combination which involves the notion of two main effects and their interactions. The weights and proportions are shown in Table 5. In using orthogonal contrasts it is necessary to settle for a coherent set of indices, and it is the set that is most useful that has to be constructed. From a practical point of view one may choose to

	0	rtho	gonal	Table 4. Comparisons Applied to AGE
+1	+1	-1	-1	Variable 1. Young (0-5, 6-24) / Total population
+1	-1	0	0	Variable 2. Very young (0-5) / 0-24
0	0	-1	+1	Variable 3. Aged (65+) / 25 and over

Orth	ogon	al C	ompar	Table 5. risons for Four Categories:	111
<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	Proportions	
+1	+1	-1	-1	(a + b) / (a + b + c +	d)

+1	+1	-1	-1	(a + b) / (a + b + c + d)
+1	-1	+1	-1	(a + c) / (a + b + c + d)
+1	-1	-1	+1	(a + d) / (a + b + c + d)

use nonorthogonal comparisons, but care should be taken to avoid spurious correlations.

We constructed 42 indices, not all of which were based on orthogonal contrasts. A few were medians, some were based on only two categories, and some did not follow the orthogonal approach, but were based on hunches as to the significance of a variable. These variables were subjected to a cluster analysis procedure developed by the senior author. The procedure can be described briefly as involving the following steps:

1. Pick a key variable on the basis of the highest variance of rows of absolute r's in the 42 x 42 correlation matrix from among variables not already included in clusters or tested previously as a key variable. The variance does a good job of picking good key variables for distinctive clusters.

2. Add variables to the cluster on the basis of correlation with the key variable corrected for uniqueness, generally using .70 as a cutoff point. This is a loose criterion which helps to prevent formation of clusters between two independent dimensions.

3. Accept the cluster if a minimum number of variables (usually three or four) is included in the cluster. A variable ending up in more than one cluster is allotted to the cluster with which it has the highest relationship.

4. Continue in this way until all variables are included in clusters or have been tested as key variables.

The cluster analysis is followed by a multiple group factor analysis and a procedure to improve the solution. This procedure has produced six factors, all of which were interpretable. We feel that this is due to the sensitivity of the clustering procedure. In many social area studies only three interpretable factors are found. Our six factors were: socio-economic status, group quarters, migration, residential or suburban status, ethnic status and instability. The three uncommon factors are group quarters, migration, and ethnic status. The function of the cluster or factor analysis is to examine the relationships among the indices and thereby to seek meaning for them. We were particularly interested in knowing whether the indices were meaningful and whether by use of the orthogonal contrast procedure we could increase the use of a single variable to measure different dimensions. We were gratified to find that this happened in the case of AGE, FAMILY TYPE, MARITAL STATUS, NATIVITY, MIGRATION and LABOR FORCE. The three indices based on age categories, for example, ended up in three different clusters. The proportion young went into the migration cluster, the proportion very young into the group quarters factor, and the proportion aged into the residential one. This happened in spite of the fact that the correlation between proportion young and very young was -.683.

Another one of our successes was marital status. Proportion ever-married went into the group quarters factor and proportion divorced or separated into the instability factor. Similarly, nativity produced a foreign stock index which identified with the instability factor. Not all orthogonal contrasts showed this diversity of meaning. Male labor force resulted in two variables, proportion professional-managerial and proportion blue collar, both of which represented socioeconomic status. The housing indices were not organized systematically according to orthogonal contrast principles, but at least three different denominator terms were used. Housing items are represented in three clusters--residential, socioeconomic status and group quarters.

Conclusion: Our conclusion is that where we applied orthogonal contrast principles there was considerable success in creating two or more indices which represented different meaningful factors. We might have benefited from still additional orthogonal indices in some instances. We also feel that factor analysis has generally not solved the problems of determining the number of dimensions and rotation to a meaningful position, and there remains a valid need for a good cluster analysis program. Those interested in a more detailed paper should write to us. The cluster analysis program is written in DYSTAL. a set of FORTRAN subroutines allowing for dynamic storage allocation, and is one of the programs in our DYSTAL II tape. If you are interested you can write to the senior author about obtaining a copy of the tape.

Table 6. Description of 42 Indices

х. . . .

	Fact	ors	8:	S = Soci G = Grou M = Migr	o-Economic Status p Quarters ation	R = Residential (Suburban) E = Ethnic (Nonethnic) I = Instability (Stability)
Source		Fac	tor	No.	Name	Index
4.00	ĺ		М	1	Young	Population, age 0-24 / Total population ($r_{12} =137$, $r_{13} =683$, $r_{23} =011$)
Age	{		G	-2	Very young	Population, age $0-5$ / Population, age $0-24$
	l	•	R	-3	Aged	Population, age 65 and over / Population, age 25 and over
			G	4	Group quarters	Population living in group quarters / Total population
Family type		{	I	-5	H-W families	Husband-wife families / All families
		[M	20	Couples with children	Married couples with children under 6 (should have used H-W families with children under 6 to be con- sistent)
				6	Children	Population, age 0-16 / All families
Marital state	15	ſ	G	-7	Ever-married	Ever-married persons, age 14 and over / Population, age 14 and over (r ₇₈ =342)
		l	I	+8	Div-sep	Divorced or separated persons, age 14 and over / Ever- married persons, age 14 and over $(r_{9,10} =210)$
Nativita		ſ	E	+9	Foreign stock	Foreign born or native of foreign or mixed parentage / Total population
Nativity		l	I	+10	Black	Black population / Population native of native parentage (both parents)
			М	11	Recent migrant	Foreign born persons who immigrated from 1945-1970 / Total foreign born population
			М	12	Out of staters	Persons born out of state / Total population
			М	-13	Stayers	Persons, age 5 or more, living in same house five years ago / Population, age 5 or more (excluding Armed Forces)
		{	E	+14	Old residents	Persons who moved into present unit 1949 or earlier (excluding persons born into unit) / All persons who have ever changed housing unit (r _{14,15} =218)
Migration		l	М	15	Recent movers	Persons who moved into present unit from 1968-1970 (excluding persons born into unit) / Persons who moved into present unit from 1950-1970 (excluding persons born into unit)
				16	H.S. graduates	Persons, age 18-24, who have completed high school / Population, age 18-24
				17	College graduates	Persons, age 25 or more, who have completed college / Population, age 25 or more
				18	Median education	Median educational level of population, age 25 or more
				19	Fertility	Number of children ever born to females, age 15-44 / Ever-married females, age 15-44

	E	+21	Female labor force	Females, age 16-64, not inmates and not attending school but in the labor force / Females, age 16-64, not inmates and not attending school
		-22	Male unemploy- ment	Civilian males, age 16 and over, unemployed (but in labor force) / Total civilian male labor force (r _{22,23} = .340, r _{22,24} = .139, r _{23,24} = .664)
Male and female labor force	S	23	Profmanagerial	Civilian males, age 14 and over, employed in profes- sional, technical, managerial or administrative work / Total civilian male employed population, age 14 and over
	S	-24	Blue collar	Civilian males, age 14 and over, employed in blue collar occupations (not farmers or farm managers) / All civilian employed males, age 14 and over, other than professional, technical, managerial or admin- istrative
	G	28	Males	Male population / Total population
	S	25	Family income	Median family income
	I	+26	Welfare	Families or unrelated individuals, age 14 and over, receiving public assistance or welfare payments / All families and unrelated individuals, age 14 and over, with income
	I	+27	Poverty	Families in government determined poverty status / All families
	R	-29	Old housing	Occupied or vacant year-round housing units built 1939 or earlier / All occupied and vacant year-round housing units
		-40	Vacancy	Vacant housing units for rent or sale 6 months or more / All occupied and vacant year-round housing units
	R	30	One-unit housing	One unit, detached, occupied housing / All occupied housing units
	S	31	1 1/2 baths	Occupied housing units with one and a half or more baths / All occupied housing units
		-32	Air conditioning	Occupied housing units with no air conditioning / All occupied housing units
	S	33	Rent	Median gross rent of renter occupied units
		-34	Four or more	Occupied housing units in structures of four or more stories / All occupied housing units
	S	35	Dishwasher	Occupied housing units with a dishwasher / All occu- pied housing units
		36	Owner-occupied	Housing units which are owner occupied or being bought / All occupied housing units
	R	41	Young owner	Husband-wife families with head below age 45 in one unit, owner occupied housing (not mobile homes or trailers) / All owner occupied housing units
	S	42	New homes	One unit, owner occupied housing of value \$25000 or more, built from 1950-1970 / All one unit, owner occupied housing units
	R	-39	Commercial use	One family, occupied housing units which are used for commercial purposes / All one family, occupied housing units

quarters	G	37	Persons per room	Total population /	Total	number	of room	s not	in	grou
				quarters						

38 Value of housing Median value of owner occupied, one family houses which are on a place of ten acres or less and have no business or medical office on the property. (Mobile homes, trailers, cooperatives and condominiums are not included)

Bibliography

S

- Blalock, Hubert M., Jr., <u>Social Statistics</u>, 2nd edition, McGraw-Hill, 1972.
- Dent, Franklin Owen, <u>Aspects of Change in Urban</u> <u>Social Spatial Structure</u>, Doctoral Dissertation, Brown University, Department of Sociology, 1972.
- Rummel, R.J., <u>Applied Factor Analysis</u>, Evanston: Northwestern University Press, 1970.
- Shevky, Eshref and Bell, Wendell, <u>Social Area</u> <u>Analysis</u>, rev. ed., Chicago: University of Chicago Press, 1955.
- Shevky, Eshref and Williams, Marilyn, <u>The Social</u> <u>Areas of Los Angeles</u>, Berkeley: University of California Press, 1949.

- Snedecor, George W., "Nonorthogonal Sets of Comparisons," <u>Statistical Methods</u>, 4th edition, Ames, Iowa: The Collegiate Press, Inc., 1946, pp. 406-408.
- Thomson, Godfrey H., <u>The Factorial Analysis of</u> <u>Human Ability</u>, New York: Houghton Mifflin, 1951.
- Thurstone, L. L., <u>Multiple Factor Analysis</u>, Chicago: University of Chicago Press, 1947.
- Timms, Duncan W.G., <u>The Urban Mosaic</u>, London: Cambridge University Press, 1971.
- Tryon, R.C., "Cumulative Communality Cluster Analysis," <u>Educational and Psychological</u> Measurement, 18 (1958), pp. 3-35.
- Tryon, R.C., and Bailey, D.E., <u>Cluster Analysis</u>, New York: McGraw-Hill, 1970.