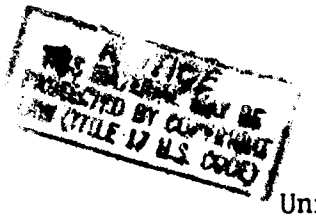

NOTES ON PERCENTAGING TABLES

by Earl R. Babbie

Survey Research Center
University of California, Berkeley

Organizing survey research data in the form of simple contingency tables ought to be an easy task; all too often it appears impossible. This short collection of notes and examples is intended as a rough guideline for preparing and reporting contingency tables.

The primary purpose of a contingency table is to demonstrate the relationship between two variables (such as sex and religious involvement). This purpose is accomplished through the comparison of percentages; the manner in which such percentages are to be computed and reported, however, is a perennial stumbling block for researchers.

At base, the logic of a contingency table is one of cause and effect. Although most researchers are reluctant--and wisely so--to declare that one attribute causes another in an absolute or mechanical sense, it is necessary to view data in these terms in preparing a contingency table.

In preparing a contingency table, the researcher must begin by labelling one of his variables as the cause and the other as the effect. Having done this, he may then proceed to organize and percentage his data in one of the two formats shown below.

		CAUSE	
		A	B
EFFECT	A	80%	30%
	B	20%	70%
		100%	100%

		EFFECT		
		A	B	
CAUSE	A	80%	20%	100%
	B	30%	70%	100%

Note that the table may be percentaged "down" or "across." The rule which must be remembered is that the percentages shown within each category of the casual variable must total 100 percent. In visual terms, the designations of "100%" appear directly opposite the labels for the causal variable.

Another simple rule may be used in determining which variable is the "cause" and which the "effect." The variable which is determined earliest in time is the cause. Thus, sex is determined earlier in time than church attendance. The college a person attended is determined earlier than his

current occupation. It is absolutely essential that the "cause" and the "effect" be identified in this manner before the relevant data are organized and percentaged in a table.

Example: The relationship between sex and church attendance.

Frequency of Church Attendance	Sex	
	Male	Female
Every week	20%	60%
Less often	80%	40%
	100%	100%

To the extent that there is an empirical relationship between sex and church attendance, it must be implicitly assumed that sex causes (predisposes, enhances, encourages, leads to, etc.) church attendance, not vice versa. The table shown above might legitimately be reported in several ways.

(1) Twenty percent of the men attend church every week, while sixty percent of the women do.

(2) Women are more likely to attend church than are men.

(3) Women are three times as likely as men to attend church every week.

Percentaging the table in the other direction leads to rather questionable conclusions, such as "People who attend church every week are more likely to become women than those who attend less frequently." [Note: There are some situations where it may be useful to percentage such tables backwards for purely descriptive purposes. In the present example, such a procedure might be useful in determining the number of men's rooms and ladies' rooms to be constructed in a new church. Normally, however, the rules which have been outlined should be followed.]

Example: The relationship between education and income.

It seems reasonable that anyone wishing to explore this relationship desires to know whether (or to what extent) increased education leads to a higher income. Implicitly, then, education is the cause (determined earlier in time) and income is the effect. The following table is an example of how the data ought to be organized and presented.

Education:	Income: Under $2,000	$2,000 to $5,000	$5,000 to $10,000	Over $10,000	
Grammar school	50%	30%	20%	0%	100%
High school	10%	20%	40%	30%	100%
College	0%	5%	45%	50%	100%

Note that this table has been percentaged "across" rather than "down" as was done with sex and church attendance. The difference is totally irrelevant. In both examples, each category of the "cause" variable (sex, education) adds up to 100%. The education-income table might be reported in several ways.

(1) People with little education earn lower salaries than those with more education.

(2) Nobody with only a grammar school education earned over ten thousand dollars; by comparison, 30 percent of those with high school educations and 50 percent of the college-educated earned over ten thousand dollars.

(3) Income increases with education.

Statements such as these are meaningful and enhance our understanding of the relationship between education and income. Percentaging the table in the opposite direction produces conclusions which lead nowhere. "The more money a man earns today, the more likely he is to have had a college education in the past."

The legitimate conclusions drawn from the proper table suggest meaningful projections. If a person gets a college education today, he can plan on a higher income tomorrow. But it is ludicrous to suggest that the more money he earns today will increase the education he received yesterday.

Miscellaneous Guides:

(1) Variables such as sex, age, and place of birth--whenever they are related to anything else--are "causes." No variables of interest to survey researchers predate these in time.

(2) With these exceptions, most variables can be either causes or effects--depending on the context. In relating education to income, for example, education is the cause; in relating sex to education, education is the effect. The determination of which variable is cause and which is effect depends solely on the two variables being considered--nothing else.

(3) Today's attitudes are caused by yesterday's conditions and experiences.

(4) In general, characteristics of parents are the causes of similar characteristics in their children. Thus, a father's education may "cause" a son's education, but not the reverse. A mother's church attendance may "cause" her child's church attendance, but not the reverse. [There are exceptions to such guidelines, of course, but the researcher must be able to present a convincing argument for making such exceptions.]

"Cause" and "Effect" in the Language of Survey Research

Most researchers are reluctant to use the terms "cause" and "effect"--and with good reasons. Very seldom is one variable the sole and absolute cause of another. Most social characteristics result from many causes. The language of survey research reflects the recognition of this fact.

Researchers speak more frequently of "independent variables" and "dependent variables" than of "causes" and "effects." The underlying logic is the same in both cases. A dependent variable is one which depends on some other variable: the independent variable. An effect depends on a cause. Church attendance depends (in part) on sex.

The following list presents some of the terms used in the language of survey research which may be viewed as more responsible ways of saying cause and effect.

CAUSE	EFFECT
Independent variable	Dependent variable
Source	Consequence
Leads toward	Reflects
Enhances	Is enhanced by
Produces	Results from

Review

Whatever language is used in reporting conclusions, the logic of contingency tables is one of cause and effect.

Two variables may be identified as cause and effect by determining their relative appearance in time. The variable which appears first is the cause.

The contingency table should be percentaged in such a way that the percentages within each category of the cause (sex: men, women) total 100%.

Conclusions are drawn by comparing the categories of the cause in terms of the percent who are characterized by a given category of the effect. Compare men and women in terms of the percents who attend church every week. Compare the educated and the uneducated in terms of the percents who earn over $10,000.