

STATISTICAL LITERACY: READING TABLES OF RATES AND PERCENTAGES

Milo Schield, Augsburg College

Department of Business 2211 Riverside Drive, Minneapolis, MN 55454

Abstract Statistical literacy depends on the ability to read and compare data presented in tables and graphs. Tables of rates and percentages are very useful in supporting arguments, but these tables are often difficult to decode. This paper presents rules for determining whether an index variable in such a table is part or whole depending on whether the associated margin value is an average, a sum or a 100% sum. Tables with missing margin values – date-indexed tables, half tables and control tables – are analyzed. Recommendations are made to improve reader understanding of any table involving rates or percentages.

Keywords: tabular data, ratio tables.

1. STATISTICAL LITERACY AND TABLES

Statistical literacy is the ability to use and evaluate arguments that use statistics as evidence. The need to base statistical literacy on reasoning that is proportional, conditional and contextual is presented in Schield (2000a).¹ A goal of statistical literacy is to construct readily understandable ratio-based *comparisons* that follow directly from data, take into account multiple factors, and can support arguments about causation. The grammar used to form comparisons of rates and percentages is presented in Schield (2000b).

Tables facilitate comparisons by separating data into groups organized in columns and rows.² Ratio tables – tables of rates and percentages – facilitate conditional reasoning because they control for multiple factors using both selection and proportions. Ratio tables are quite common. In the 1997 U.S. Statistical Abstract, about 20% of the tables contain part-whole percentages, 10% contain rates and 5% contain percentage changes.

Although there are some good books on tables³, these books say little on reading or titling tables of ratios.

¹ All three take into account (control for) the influence of related variables. Proportional reasoning uses ratios (percentages and rates); conditional reasoning uses selection (tables), and contextual reasoning uses comparisons (e.g., “percent more,” “times as much”, likely and relative risk).

² *Table* is used herein to refer to a *summary, demonstration or reference table* (data tabulated for groups of subjects) rather than a *detail list table* (data listed for each individual subject).

³ In *Plain Figures*, Chapman and Wykes (1996) give an overview of ways to present numbers effectively but say little about difficult tables: “Few readers will have any difficulty in interpreting a table of percentages (so long as it is clear what the total of percentages is based on)...”. *The Publication Manual* of the American Psychological Association (1994) and the Census Bureau *Manual of Tabular Presentation*

(1949, paragraphs 371-374) comments at length on the difficulty of titling tables of percentages and proposes “a search for a systematic presentation which focuses attention on basic problems in percentage description.” Thus, the problems involving ratio tables have been known for at least 50 years.

The purpose of this paper is to present some rules for reading tables of rates and percentages.

2. RATIO TABLE DIFFICULTIES

Ratio tables – tables of rates and percentages – can be valuable as evidence, but they can also be difficult to read. Ratio table headings can be misleading to those who apply the grammar rules for part and whole.⁴

Table 1. **Low Birth Weight and Births to Teenagers**

DIVISION AND STATE	PERCENT OF BIRTHS WITH LOW BIRTH WEIGHT \1	BIRTHS TO TEENAGE MOTHERS, % OF TOTAL
United States	7.0	12.8
New England	5.9	8.4
Middle Atlantic	7.3	9.5
East North Central	7.1	13.2
West North Central	5.9	11.1
South Atlantic & DC	7.9	14.4
East South Central	8.2	18.4
West South Central	7.3	16.3
Mountain	6.8	12.8
Pacific	5.7	11.5

\1 Less than 2,500 grams (5 pounds-8 ounces)
Table 101CD 1990, 1994 US Statistical Abstract

College students who were experienced in the grammar of percents and percentages were asked to describe the circled values in the bottom row of Table 1. They relied on the column headings, since neither title nor row headings contained any part-whole indicators. The students knew that “percent of” always introduces the *whole* as in, “X% of {Whole} are {Part}.” In describing the circled 5.7, many concluded that the *whole* was *low birth-weight births* and wrote, “In 1990, 5.7% of all US low birth-weight births occurred in the Pacific states.”

In describing the circled 11.5, they had problems with the preposition *to*. In neither *percent* nor *percentage* grammar does *to* identify part or whole. Some concluded that the *whole* was *births to teenage mothers*, and wrote, “In 1990, 11.5% of all births to teenage

(1949) focus on the design and titling of count tables. In *Presenting Your Findings*, Nicol and Pexman (1999) suggest which table to use in presenting a particular statistical test.

⁴ A ratio involves a division of two counts. In a part-whole ratio, the numerator is *part* and the denominator is *whole*.

mothers occurred in the Pacific states.” Others concluded that *births* alone was the *whole* and wrote, “In 1990, 11.5% of all births occurred to teenage mothers in the Pacific states.”

In every case, their description was wrong. Although they used the grammatical clues, the titles and headings gave inadequate guidance on what was *part* or *whole*. These problems are not exceptions. Due to space limits, misleading titles are common in ratio tables.⁵

3. TABLE TERMINOLOGY

Index variables (e.g., race and sex) are used to classify subjects into cells. **Index values** are the values an index variable can have (e.g., *male* and *female* for *sex*, *white* and *non-white* for *race*). Kelley (1996) asserts that classification values should be exclusive and exhaustive.⁶ Non-exclusive values (e.g., *white* and *Hispanic*) can lead to double counting.⁷ Non-exhaustive values (e.g., *white* and *black* for *race*) can omit data.

Table cells can be classified as either margin or body cells. Margin cells (shaded in Table 2) involve a *Total* or *All*⁸ since the values include all related subjects. A margin is a related group of margin cells (column or

⁵ Table 226: *Percent in Percent of Overweight People who think they are Overweight* should be *percentage*. Table 183: *With in Percent of Adult Persons with a Dental Visit* should be *who had*. Table 197: *Having in Percent of Women having a Mammogram* should be *who had*. Table 167: *Percent of Population Covered* should be *Percentage of Population [which are] Covered*. Table 210: *Percent Employed* might be better as *Employed as a Percent*. Table 153: *Percent Federal of Public Funds* might be better as *Federal as Percent of Public Funds*. Table 201: *Percent Distribution* is disconcerting when a group doesn't add to 100%.

⁶ Kelley presents *being essential* as another criteria for selecting variables based on Rhetoric and the Scientific Method.

⁷ Having a 100% margin value with non-exclusive parts is disconcerting. In Table 200, 1997 U.S. Statistical Abstract, *home health care* can include multiple services per patient.

⁸ Jenkinson (1949, para. 958) writes: “**Total**” vs. “**All**.”—At times the word “All” is used on total and subtotal lines. Although “All” may be used in place of “Total” at times, it is not synonymous in table design and should not be considered as such. (a) “**Total**” may be used standing by itself.—“All” should not be; it should always be followed by another word or phrase. (b) **Both may be used in indicating presentation unit.**—However, “Total,” when so used, implies nothing more than that the line entries represent the summation of the units shown in the subentries. “All” implies that the line contain figures for the entire universe comprised of units “x,” irrespective of type. Thus “Total families” means merely “Total families (of the type shown).” “All families” means “All families (of every type).” Therefore if “All” is used, the caption should specify in full the group covered; if “Total” is used, complete specification in the line caption is not necessary, although frequently desirable.

row).⁹ Margin values may be sums but this is not necessary. (See Tables 3 and 4)

In Table 2, the index variable for the columns is *Sex*; the index values are *Male* and *Female*.

Table 2. Counts Table

College Students	SEX		
Major	Male	Female	TOTAL
Business	26	24	50
Economics	4	6	10
MIS	10	30	40
TOTAL	40	60	100

4. MARGIN VALUES

Whether labeled as *All* or *Total*, margin values can be either sums or averages. Knowing whether a margin value is a sum or average is an important key to decoding part and whole in tables of rates and percentages.

It is easy to determine whether a margin value is a sum or average. If the value of any data cell is bigger than the associated margin value, then the margin value must be their average. If all the values of data cells are smaller than their associated margin value, then the margin value must be their sum.

Total can describe averages. This may seem misleading, but *Total* refers to the subjects (*All* versus *Some*)—not the arithmetic operation (*Sum* versus *Average*).

5. 100% TABLES

A **100% table** is a percentage table with at least one margin cell having a value of 100%. There are three types of 100% tables: column, row and total.

Table 3. Column Table: Each column sums to 100%.¹⁰

College Students	SEX		
MAJOR	Male	Female	TOTAL
Business	65%	40%	50%
Economics	10%	10%	10%
MIS	25%	50%	40%
TOTAL	100%	100%	100%

Table 4. Row Table: Each row sums to 100%.

College Students	SEX		
MAJOR	Male	Female	TOTAL
Business	52%	48%	100%
Economics	40%	60%	100%
MIS	25%	75%	100%
TOTAL	40%	60%	100%

⁹ **Table margins** are summaries of table data and may be located on either the outer or inner edges of a table.

¹⁰ Do not describe this bottom row as either a *Row Margin* or a *Column Margin*. *Row Margin* describes the shape. *Column(s) Margin*—margin of the columns—describes the body cells that are summarized in the margin values. Both the shape and the data source apply so both terms are ambiguous.

Table 5. Total table. Body cells sum to 100%.

College Students	SEX		
	Male	Female	TOTAL
MAJOR			
Business	26%	24%	50%
Economics	4%	6%	10%
MIS	10%	30%	40%
TOTAL	40%	60%	100%

6. 100% MARGIN VALUE RULE

This rule applies only to ratio tables with a 100% sum.

100% Margin Value Rule: If a group of cells have a 100% sum margin value, then their unique index values are *parts* and their common index value is a *whole*.¹¹

In Table 3, the margin cells of the columns are 100% sums.¹² Thus, the rows (index values) are parts; the columns are wholes. So we say: “Among these college students, 65% of the males are business majors.”

In Table 4, the margin cells of the rows are 100% sums.¹² Thus, the rows are wholes and the columns are parts. So we would say: “Among these college students, 52% of the business majors are males.”

In Table 5, the only 100% margin sum is in the lower-right corner. The right-side column and the bottom row each sum to 100%, so their pieces are parts. Thus both columns and rows are parts. So we would say: “Among these college students, 26% are male business majors.”

7. LESS STRAIGHT-FORWARD TABLES

Most ratio tables have no 100% margin values. They are harder to decode. Consider Table 6.

Table 6. Unemployment Rates (\1), by Educational Attainment, Sex, Race, and Hispanic Origin

Sex & Race	Total	--- Highest Grade Achieved ---			
		<12	12	13-15	>15
Total \2	4.4	10.4	5.1	3.8	2.0
Male	4.7	9.9	5.5	4.0	2.1
Female	4.1	11.3	4.5	3.6	2.0
White	3.9	9.4	4.6	3.4	1.8
Black	8.1	16.6	8.2	6.1	4.4
Hispanic \4	7.3	9.6	7.5	5.5	3.0

\1 Percent unemployed of the civilian labor force
 \2 Includes other races, not shown separately.
 \4 Persons of Hispanic origin may be of any race.
 Source: Adapted from Table 681, 1998 US Statistical Abstract

In Table 6, *unemployment* is a *part* because it modifies *rate* in the title, while *civilian labor force* is presumably a *whole*, since it follows *percent of* in table footnote \1.

¹¹ Arithmetically, the part and whole of a ratio are just two numbers. Conceptually, part and the whole are formed from sets by means of intersections. In tables, part and whole are formed from the intersection of index values and title terms.

¹² Note that the 100% margin value in the lower right corner is a sum in one direction but not in the other.

The part-whole status of the other index variables is grammatically indeterminate, because *by* doesn't indicate either *part* or *whole* in a table title.

8. MARGIN VALUE RULE

Margin Value Rule: If a group of cells have a margin value that is their average, then their unique index values are wholes; if they have a margin value that is their sum, then their unique index values are parts.^{13,14}

This rule applies to all ratio tables. Consider the averages in Tables 3 and 4. Table 3: the two 100% cells in the bottom row have a margin value (100%) that is their average, so their unique pieces (the columns) must be wholes. Table 4: the three 100% values in the right column have a margin value (100%) that is their average, so their unique pieces (the rows) must be wholes.

9. FULLY MARGINED RATIO TABLES

Fully margined ratio tables – ratio tables with an associated margin summary for every table index – are easily decoded using the margin value rules.

In Table 6 all the margin values are averages, so the common indexes (rows and columns) must all be wholes. For Table 6 we would say, “3.0% is the unemployment rate of Hispanics whose highest grade achieved is greater than 15.”

In Table 7 the margins (the margin summaries) are the shaded row (*Total*) and column (*All crimes*). The margin values at the top of each column are averages, so the rows must be wholes. The margin values at the left of each row are sums, so the columns must be parts. (The sums are tricky due to the subtotals.) In the title, *Victimization*¹⁵ is a part, since it modifies *rate*. To describe the circled 2.1, we would say, “In 1995, the rate of victimization involving personal thefts is 2.1 per 1,000 US females 12 or older.”

10. MISSING-MARGIN RATIO TABLES¹⁶

Most ratio tables have at least one missing margin summary.¹⁷ In such tables, the margin summary associated with the time periods (e.g., years) is most likely to be missing.¹⁷ Other tables with missing-margins include one-way half-tables and control group tables.

¹³ This rule assumes the margin value is meaningful. The death rate for men (9 per 1,000) and the death rate for women (8 per 1,000) can be summed (17 per 1,000), but this sum is not meaningful. The percentage of a group who are male (60%) and the percentage who are female (40%) can be averaged (50% if equally weighted), but this average is not meaningful.

¹⁴ In a 100% sum table, the common index value must be a whole since the sum is a 100%.

¹⁵ A victimization is a self-report by a victim..

¹⁶ Missing-margin tables lack a margin summary for at least one index variable.

Table 7. **Victimization Rates, by Type of Crime and Characteristic of the Victim: 1995**
Rate per 1,000 persons age 12 years or older. [Taken from Table 347 of the 1998 US Statistical Abstract CD]

1995 ITEM	----- CRIMES OF VIOLENCE -----							
	All crimes	All crimes of violence	Rape/ sexual assault	Robbery	Total	Assault Aggra- vated	Simple	Personal theft
Total	46.2	44.5	1.6	5.3	37.6	8.8	28.9	1.7
Male	54.4	53.2	0.3	7.5	45.3	11.9	33.5	1.3
Female	38.5	36.4	2.8	3.2	30.4	5.9	24.6	2.1
White	44.6	43.1	1.6	4.2	37.3	8.2	29.1	1.4
Black	58.4	55.4	1.6	12.5	41.3	12.6	28.7	3.1
Other	43.8	40.9	1.2	6.3	33.4	9.2	24.2	2.9

11. DATE-INDEXED RATIO TABLES¹⁷

When the date-index in a ratio table lacks a related margin, then the dates delimit table-specified wholes.

Table 8. **Percentage of Smoking Prevalence Among U.S. Adults, 18 Years of Age and Older.**

Year	Total	Male	Female	White	Black
1965	42	52	34	42	46
1974	37	43	32	36	44
1985	30	33	28	30	35
1994	26	28	23	26	27

Source: www.cdc.gov/nccdphp/osh/prevali.htm 3 Aug., 2000

Table 8 has two separate tables placed side-by-side: one indexed on *sex* (male vs. female) and one indexed on *race* (white vs. black). They share a common index (date) and a common title. While the *sex* table has exhaustive index values, the *race* table does not. Presumably, the *Total* data includes all races.

The margin values in the left data column are row averages for the *sex* and *race* tables. Thus, the column headings (e.g., *blacks*) must be *wholes*. *Smoking* is a *part* since it modifies *prevalence*¹⁸ and follows *percentage of* in the presence of *among*.¹⁹

There are no margin values for the column cells, so the date rows are each *whole delimiters*. Thus we would

¹⁷ In 126 non-percentage rate tables, 110 had at least one missing margin summary. Of the 142 missing margins, 86 were date-related (sec. 11), 34 were control group-related (sec. 14), 9 were half tables (sec. 12-13), 8 had multiple non-exclusive wholes (sec 15) and 5 had multiple exclusive parts (sec. 15). Nine tables had growth rates; three had wage rates. In 351 non-rate percentage tables, about 10% had a percentage change, 85% had part-whole percentages and 5% had both, while 85% had missing margin summaries. The part-whole percentage tables with missing margins involved dates (about 50%), half tables (25%), missing 100% sums (20%), and missing group totals (10%). 1997 US Statistical Abstract

¹⁸ *Prevalence* is a *rate* at a specific time (c.f., *unemployment*).

¹⁹ The title should use *percentage* or *rate* grammar – not both.

describe the circled 27 as, “In 1994, the smoking prevalence among blacks was 27%.”

12. SINGLE HALF-TABLES

Another common table with a missing margin summary is a single one-way half-table. **Half-tables** are ratio tables with a binary index where redundant data associated with one binary index value is omitted. By omitting redundant data, less space is required. But the cost for a reader is increased difficulty in decoding the table.

Table 9. One-Way Half-Table with a Part Header²⁰

College Students	SEX		
	Male	Female	TOTAL
MAJOR			
Business	52%	48%	100%
Economics	40%	60%	100%
MIS	25%	75%	100%
TOTAL	40%	60%	100%

The left half of Table 9 is a half-table: the right two columns are omitted.²¹ The margin value of the column (40%) is an average, so the rows must be *wholes*.²² Since there is no other *part* in the title and there must be a *part* in any ratio table, the column (*male*) must be a *part*. Thus we would say, “52% of these business majors are males.” And from this we can easily determine that “48% of these business majors are females.”

13. COMPOUND HALF-TABLES

Compound half-tables are a related series of one-way half-tables having index values placed side-by-side.²³

²⁰ The part can be in the column header, in the title or in both. A two-way half table is a single table with two index margins having a common part in the title (c.f., Table 6).

²¹ Do not describe this half table as a column table. It is a column-shaped table, but it is a ‘half’ of a 100% row table.

²² Since the 40% margin value is not a sum, the table cannot be a *column* table or a *total* table. Thus, it must be a *row* table with each row totaling 100%. Using the 100% Margin Rule, we see that the rows must be *wholes* and the columns *parts*.

²³ **Compound tables** have a shared index variable on one axis and different index variables on the other. When they have

They save space and link related items to facilitate comparison, but the cost for a reader is even more difficulty in decoding such tables.

Table 10. Percent of Women, 15 to 44, Who Received Selected Medical Services: 1995

CHARACTERISTIC	HIV Test	Pregnancy test	Pap smear
Total	17.3	16.0	61.9
AGE AT INTERVIEW			
15-19 years old	14.6	16.1	33.5
15-17	12.1	11.4	23.0
18-19	18.5	23.3	49.9
20-24 years old	23.7	27.4	68.7
25-29 years old	23.6	25.3	70.9
30-34 years old	18.5	17.4	69.5
35-39 years old	14.2	8.1	62.9
40-44 years old	10.0	4.3	62.7

Source: Taken from Table 198 in 1998 U.S. Statistical Abstract.

In Table 10, the margin value at the top of each column (*Total*) is an average, so the row index values (*ages*) are *wholes*. We still need a part. The column index values are not exclusive (a woman can receive more than one medical service in a year), so Table 10 is a compound half-table. The missing “halves” are not shown (e.g., “no HIV test”, “no pregnancy test”, etc.). Since there are no margin values for the row cells, we cannot determine the part-whole status of the columns.

However, we can use the title. *Selected medical services* would be the part if the keyword were *percentage* rather than *percent*.²⁴ If there is no other candidate for the part, we must conclude that *percent* should be read as *percentage* and the *selected medical services* are each *parts*. Each column half-table is a part. We would say, “In 1994, 4.3% of US women age 40-44 received a pregnancy test.” (Analyze Table 1 similarly.)

Table 11. Abortions by Race

	ALL RACES		WHITE		BLACK & OTHER	
	Rate per		Rate per		Rate per	
	1,000 women	1,000 births	1,000 women	1,000 births	1,000 women	1,000 births
YEAR						
1975	21.7	331	17.2	276	49.3	565
1980	29.3	428	24.3	376	56.5	642
1985	28.0	422	22.6	360	55.5	659
1990	27.4	389	21.5	318	54.4	655
1995	22.9	351	17.0	265	48.1	686

Adopted from Table 124, 1998 US Statistical Abstract.

Table 11 is a compound half table with a common part (abortions) and different wholes (per 1,000 women vs.

common subjects, the different indexes (e.g. *race, sex, ethnicity*) have non-exclusive values (e.g., *white, male, Hispanic*).

²⁴ *Percent* and *percentage* grammar are different. A relative clause following *percentage of* always indicates a *part*; a relative clause following *percent of* never indicates a *part*. If the shorter (*percent*) was used when the longer (*percentage*) was intended, then the relative clause would introduce a *part*.

per 1,000 births). Race is a whole because the associated index values are averages. The ‘per 1,000 women’ and the ‘per 1,000 births’ index values are not exclusive. Thus, the table is a compound half table. Since the column headings follow ‘per’, these headings are wholes, and the circled 686 is read as, “the abortion rate per 1,000 live births among US non-whites in 1995.”

14. CONTROL TABLES

Control tables are tables where data is separated by values of a control variable where the values are exclusive and exhaustive.²⁵ When the control groups are not summarized, the control table has a missing margin.²⁶

Table 12. Delinquency²⁷ Case Rates, by Sex

CASE RATE Per 1,000 Youth at Risk

SEX & OFFENSE	1986	1991	1995
Male, total	71.9	85.8	92.4
Person	11.5	16.6	19.7
Property	43.6	51.4	46.7
Drugs	4.5	4.3	9.5
Public order	12.3	13.5	16.6
Female, total	17.7	21.0	27.3
Person	3.0	4.4	6.7
Property	10.3	12.6	14.2
Drugs	1.0	0.6	1.6
Public order	3.5	3.4	4.8

Adapted from Table 372 in the 1998 U.S. Statistical Abstract

Control variables in control tables are typically wholes.

In Table 12, the title indicates that *delinquency case* is a part. There are three indexes: *date, sex* and *offense*. There are no margin values for the row cells, so the *dates* are *whole delimiters*. Within *sex, offenses* have margin values at the top of the columns (*Total*). These values are sums, so the rows (*Offenses*) must be parts. There is no margin summary that combines the two sexes so we can’t tell if they are *parts* or *wholes*. Since *male* and *female* are exclusive and exhaustive, they cannot be parts in compound half tables. If they were parts of a total table, the margin sums would be shown so they are likely wholes. Since *sex* is a property of *youth*, and *youth* is a whole, *sex* is a whole. On that basis we would describe the circled 4.8, by saying, “In 1995, the delinquency case rate involving a public order offense was 4.8 per 1,000 female youth at risk.”

²⁵ If the control index values are not exhaustive (e.g., white and black) and there are margin values for ALL subjects, then the missing group is identifiable (e.g., other races not shown).

²⁶ Having control groups that are exclusive and exhaustive is necessary for them to be wholes – but not sufficient. Without margin summaries, it can be difficult to have strong confidence that these control groups are wholes.

²⁷ A *delinquency offense* is an act committed by a juvenile for which an adult could be prosecuted in a criminal court.

15. OTHER MISSING MARGIN TABLES

There may be other missing-margin tables, but the three mentioned previously are the most common.

16. RATIO TABLE RECOMMENDATIONS⁵

1. Always use *percentage* (pct-age or %-age) when a related relative clause is used to indicate a *part*.²⁸
2. Include the relative clause indicators (*who*, *what*, *that*) when the relative clause introduces a *part*: “*Percentage of Population [who are] Living Alone.*”
3. Never use an adjective-noun phrase to indicate both part and whole.²⁹ E.g., “*Percentage of unemployed men*” or “*percentage of low-weight births.*”
4. If there isn’t room for a relative clause and if *percentage of* would be ambiguous, then use *as a percent of*: “*Low weight [births] as a percent of births.*”
5. Never use just *by* to introduce a *part*;³⁰ use *distributed by* or other appropriate grammar indicators.³¹
6. Never use *Total* (always use *All*) in ratio table margin headings unless *Total* is absolutely necessary.⁸
7. Include a margin summary for every index variable unless the table involves a single one-way half table or unless that margin summary is unhelpful (e.g., dates or compound one-way half tables).

A margin summary for *All* should be generated whenever meaningful because it enables a comparison of explanatory power for index variables. Schield (1999).

There are some telegraphic conventions³² in current use that readers may have to decode. These shortcuts save space by omitting part-whole keywords. These telegraphic conventions should not be used unless the author provides a proper description in a table footnote.

²⁸ Ratio table titles or data headings should clearly indicate their ratio nature unless there is a 100% in the table.

²⁹ This common usage in sports and finance is not suitable for general-purpose usage. E.g., the “2000 US black death rate” should be stated as “*Death Rate of US Blacks in 2000.*”

³⁰ In tables, *by* indicates a classification, but does not indicate either *part* or *whole*. Sometimes *by* must be repeated: “Age, *by* race, *by* sex” is different from “Age, *by* race and sex.” *By* can also indicate *by means of*. In the 1999 US Statistical Abstract Table 147, the title, “Death Rates for Injury *by* Firearms, Sex, Race, and Age: 1996,” misleads by mixing *by* [means of] *firearms*’ with ‘[classified] *by* sex, race, and age.’

³¹ Jenkinson (1949, para. 351) writes: **Using “for” and “by” in area reference.** “By’ and ‘for’ are closely related but are distinguished by the presence or absence of the totality. “Data by State(s)” should include the data for the total US. “Data for [these] States” should exclude data for the total US.

³² To avoid the *percentage of* ambiguity, *percent of* is used to indicate a *whole*: e.g., “Percent of adults.” If there isn’t room for a relative clause, the part is shown as a gerund or a trailing modifier: e.g., “Percentage of adults employed” or “Percentage employed.” If there isn’t room for *percentage*, then “*Percent [part]*” is used: e.g., “Percent unemployed.”

17. GENERAL RECOMMENDATIONS

In 1949, Jenkinson outlined an “experiment” to establish rules for titling tables of rates and percentages. The Jenkinson “experiment” must be carried out. The inability of readers to properly decode such tables must be studied. Categories, constructs and rules such as those presented here must be verified and validated in a more top-down fashion. Educational materials must be created based on generally acceptable guidelines.

Data producers should pursue both operational goals (to collect and produce timely, relevant and quality data) and strategic goals (to facilitate the public’s ability to read such data when presented in tables).

REFERENCES

Chapman, Myra and Cathy Wykes, 1996. *Plain Figures*. 2nd Edition. The Stationary Office, London.

Jenkinson, Bruce L., 1949. *Bureau of the Census Manual of Tabular Presentation*. United States Government.

Nicol, Adelheid A. M. and Penny M. Pexman, 1999. *Presenting Your Findings: A Practical Guide for Creating Tables*. American Psychological Association.

American Psychological Association, 1994. *Publication Manual of the American Psychological Association*. American Psychological Association, Fourth Edition.

Kelley, David (1994). *The Art of Reasoning*. 2nd ed. W.W. Norton & Co.

Schild, Milo (1999). *Simpson’s Paradox and Cornfield’s Conditions*, ASA 1999 Proceedings of the Section on Statistical Education, p. 106.

Schild, Milo (2000a). *Statistical Literacy and Mathematical Thinking*, ICME-9, Tokyo, Japan.

Schild, Milo (2000b). *Statistical Literacy: Difficulties in Describing and Comparing Rates and Percentages*. ASA 2000 Proceedings of the Section on Statistical Education, p. 76.

Acknowledgments: To my students at Augsburg College who motivated me to learn how to read ratio tables. To Thomas V.V. Burnham and Linda Schield for their analysis, criticisms, suggestions and support.

To Glenn King, US Bureau of the Census Administrative and Customer Services Division, for introducing me to the *Manual of Tabular Presentation*. To the U.S. Bureau of the Census for providing copies of the U.S. Statistical Abstract to Augsburg students. To Donald Macnaughton for reviewing a draft of this paper.

To the W.M. Keck Foundation for their grant to promote Statistical Literacy. Dr. Schield can be reached at schild@augsborg.edu. This paper is available at www.augsburg.edu/ppages/~schild.

APPENDIX: Tabular Presentations

[The following is from the *Bureau of the Census Manual of Tabular Presentations* by Jenkinson (1949). Due to space limits, this Appendix is not in the ASA Proceedings. As a US Government publication, this material is in the public domain. It is reprinted here to indicate the historic awareness of the problems involved in titling tables containing rates and percentages.]

Foreward by J. C. Capt, Director, Bureau of the Census. "In the final analysis there are only two rules of tabular presentation that should be applied rigidly: *First*, the use of common sense when planning a table, and *second*, the viewing of the proposed table from the standpoint of the user. The details of the mechanical arrangement must be governed by a single objective; that is, to make the statistical table as easy to read and to understand as the nature of the material will permit."

INTRODUCTION

General Problem of Tabular Presentation

Tabular Presentation as an end-product. – ... the published tables and their accompanying text, tend to be the only results of the expenditure of effort, time and money which the using public sees; hence the user's evaluation of all operations tends to be based on the final step, that is, on the statistics as presented. In itself, this makes sound presentation vitally important. Good presentation transmits the quality of the previous operations and, in turn, assists, in making clear the meaning of the material. Poor presentation not only fails to clarify the data, it tends to impair or destroy the effect of the work that has gone on before.

Basic objective. – In the design and preparation of statistical tables the basic objective is to arrange and present data in such a way that their meaning and significance can readily be grasped by the user. ... The dividing line between good and bad presentation may best be determined by the answer to the question: "Which method is simple, focuses the reader's attention on the data (rather than on the form), and makes clear the meaning and significance of the information?" That which clarifies the meaning of the statistics and concentrates the attention upon them is good presentation; that which obscures the meaning of the data or distracts the attention from them is poor presentation.

Fundamental approach – Tabular design must be approached from the standpoint of the user rather than from that of the producer. It is not enough that the table be clear to the designer who necessarily has an intimate knowledge of the statistics and their qualifications. The imperative requirement is that the table be clear to the

user; that is, the table must be so designed that the user will be able to obtain from it that which he wishes and needs to know, so far as the desired information is present. Conformity with this requirement is not always simple since it is difficult for the table designer to discount his own knowledge and review his own work through the eyes of the using public. Yet that task must be performed if the purpose of the table is to be achieved.

Comparative roles of content and mechanics. – The purely mechanical aspects of tabular presentation, although important, can be overemphasized. Mechanics must be consistently thought of as a tool of presentation, not as an end in itself. Proper attention to mechanical style and form is essential, but forcing statistics into rigid physical pattern, or an overreliance upon mechanical tricks, can easily defeat the whole purpose of presentation. In the last analysis, the only purpose of mechanics in this field is to help in making clear to the reader the significance of the data presented; that is, the content of the table. The following statements may aid in clarifying the comparative roles of content and mechanics: (1) The point of departure for both content and mechanics is the meaning and significance of data. (2) Content is the reason for the table's existence; mechanics merely translates requirements of content within physical limits. (3) Content determines inclusion and relative placement of data; mechanics assists in presentation of content by means of typing, drafting and printing aids. (4) Content poses problems of space; the function of mechanics is to solve them in a fashion which aids in, rather than detracts from, the presentation of content. (5) Content is limited by space only in terms of the amount to be presented, never in terms of the meaning of the data. Mechanics is limited by space in terms first of preserving or enhancing meaning, and second by rigid physical and operational requirements. (6) Content is restricted by mechanics only so far as mechanical limitations add to, or detract from, the presentation of meaning. Mechanics is completely bound by content since a perfect physical pattern fitting the space is useless if meaning is impaired or destroyed.

Chapter 1:

CLASSES OF TABULAR PRESENTATION

(101-133)

101. Tabular Presentation Defined. – Tabular presentation is a means of bringing together and presenting related information or other information in columns and rows. Its object is to present in a concise and orderly fashion information that could not be presented so clearly in any other way. Since the tabular arrangement facilitates reference, comparison, and interpretation of the data, it is particularly useful in presenting large masses of related statistics.

Chapter 3: THE TABLE TITLE (301-374)

311: There are three common method of titling tables. In the conventional method, the title consists of “what, how classified, where and when.” In the key phrase method, the *what* is placed in the subject phrase and that is set off from the *how classified, where* and *when* by a dash. In the separated unit method, the universe is displayed in the title while the classification segment is displayed in the subtitle.

Sec. 3-G. Title Description of Percent Data (371-374)

371. **General.** – Correct identification of the base is the paramount consideration in describing percentages. A second requirement is correct description of the class-groups into which each universe, or sub universe, is distributed. Since these objectives must be achieved by fine distinctions in phrasing and word-order, flexibility in title treatment is reduced drastically in comparison with titles for frequency use. Percentage description usually must be independent of description of frequencies presented in the same table; that is, although certain words or phrases may be identical for both descriptions, they usually must be repeated separately for each. The need for independent description for each type of data naturally adds to title length and complexity.

372: **Percentages and frequencies appearing in same table.** Many tables contain both percentage and frequency data. In general, title treatment depends on the primary purpose of the table as a whole and the conditions or circumstances of inclusion of the two types of data.

a. **Frequency table including percentages.** – Here the primary purpose of the table is to show frequencies. The percentages are of secondary importance and are inserted for reader-convenience or to clarify, or add significance to, the frequency information. *Rule:* Confine title to description of frequencies. Omit title reference for the percentage data unless they – (1) Are of outstanding significance in themselves, *or* (2) Are not a natural byproduct of the frequencies shown; that is, they comprise percentages which would more logically have been placed elsewhere and which have been included in the given table only because of available space.

b. **Percentage table including frequencies.** – Here the primary reason for table existence lies in the percentage presentation. The frequencies are of secondary significance. Thus, it may be desirable to include in a percentage table the frequency rims used as the bases of the percentages. *Rule:* Confine title to description of percentages. Omit title-reference for the frequency data unless they – (1) Are of outstanding significance in themselves, *or* (2) Are not the bases of the percentages but comprise “supplementary” frequency information which appear in the given table because of difficulty of inclusion in some other table where they logically belong, *or* because their nature is such that they cannot be presented in a separate table.

c. **Doubtful cases.** – Where the amount of frequency and percentage data is about equal, or where the primary purpose of the table (in terms of presenting frequency or percentage data) is in doubt, the general rule is to assume that it is a “frequency table including percentages” and title it according to the rule in paragraph 372a, above.

373. **Problem of distinctive phrasing of percentage titles.** When a number of percentage tables which have been titled at various points in time are compared with each other, it may be found that the descriptive terminology has not been used consistently, even when all the tables were titled by the same person. Thus, identical phrasing (in respect to essentials) may have been used to describe basically different distribution patterns, or different phrasing may have been used to describe identical patterns, or both. Obviously then, some of the tables have not been titled properly, since the reader cannot ascribe the same meaning at all times to the same phrasing.

a. **Use of a predetermined method.** – Theoretically, the solution to the above problem would be to title all tables in accordance with some predetermined scheme broad enough in scope to provide distinctive phrasings for at least the most common situations.

In this connection, it must be remembered that the title-writer and the user approach the title from opposite directions. The title-writer works from table to title; that is, knowing the content of the table, he tries to describe it. The user, however, works from title to table; that is, he tries to deduce the content from the descriptive title. It follows that the title-writer should avoid terms which he and the user will not interpret in the same way.

The first prerequisite to a common understanding is the consistent use of descriptive phrases. Use of a standard scheme would promote consistency not only among titles written by the same person, but also among titles written by different persons. However, any scheme which attempts to differentiate between all major possibilities tends to become cumbersome where applied. Also, the distinctions in phrasing may become too fine-drawn to be useful to most readers. (see par 336.)

An experimental method is illustrated in fig. II, pp 57-58. This scheme states the problem; it does not solve it. The phrasings provided there are better suited for the description of simple than of complicated tables.

374: **Appending percentage description to frequency titles** – Where it becomes necessary to call specific attention to the presence of percentages in a frequency table, the reference usually is inserted preceding the area segment (assuming area specification is the same as for the frequencies) and is introduced by WITH.

Right: AGE OF THE POPULATION BY COLOR AND SEX, WITH PERCENT DISTRIBUTION BY AGE, FOR THE STATE: 1940

Figure 11. **Experiment in Formal Terminology for Percent Tables.** This figure illustrates the problems involved in the development of any standard scheme for phrasing of titles of percentage tables. In each case, the first title stresses the class being distributed (the base); the second title stresses the characteristic by which classified. These examples comprise an experiment in a search for – (1) A systematic presentation which focuses attention on basic problems of percentage description; (2) A testing ground for any system of terminology it may seem desirable to develop; and (3) A guide for optional use where circumstances seem to warrant.

The skeletal titles shown represent the basic terminology for the percentages described. The three dots at the beginning of each title (except tables 5 and 9) represent the words PERCENTAGE DISTRIBUTION; those at the end represent the area and time specification. Throughout, X represents the universe total (as total employed workers); A represents sex; B, color; and C, age.

(1) X distributed according to the characteristic B.

Table 1: Percentage Distribution of X by B
Percentage Distribution by B for X

(2) X classified by A and each resultant group distributed according to characteristic B.

Table 2: Percentage Distribution of A of X by B
Percentage Distribution by B for A groups of X

(3) X cross-classified by A and C and each resultant group distributed according to B.

Table 3: Percentage Distribution of C-A of X by B
Percentage Distributions by B for C-A groups of X

(4) X distributed according to groups formed by cross-classification by A and B (single column arrangement)

Table 4: Percentage Distribution of X by A and B.
Percentage Distributions by B and A for X

Table 5: Percent of X in each A-B group. [See Table 4]

Table 6: Percentage Distribution of X by A and by B.

Table 7: Percentage Distribution of A and B of X by B.

Table 8: Percentage Distribution of A by B and A by C.

Table 9: Percent A1 in each B-C group of X.