1   12/03/2001

Reading Census Data — Statistical Literacy

Student Difficulties Reading Tables of Ratios

U.S. Bureau of the Census
December 3, 2001

MILO SCHIELD
Project Manager
W.M. Keck Statistical Literacy Project
Professor, Business Administration
Augsburg College
www.augsburg.edu/ppages/~schield
schield@augsburg.edu

---

2   12/03/2001

Reading Census Data — Statistical Literacy

Topics

1. *Reading Ratio Tables: slides 3-11*
2. *Interpreting Averages: slide 12*
3. *Interpreting Associations: slide 13*
4. *Census Bureau Mission: slides 14-15*
5. *Recommendations: slide 16*
6. *References, Acknowledgments: slides 17-18*

---

3   12/03/2001

Reading Census Data — Statistical Literacy

Rate Tables
Ambiguity of '*by*'

No. 149. Death Rates for Injury by Firearms, Sex, Race, and Age: 1995
[Death rate per 100,000 population. Deaths classified according to the ninth revision of the *International Classification of Diseases*]

ITEM	5-14 yrs. old	15-24 yrs. old	25-34 yrs. old	35-44 yrs. old	45-54 yrs. old	55-64 yrs. old	65-74 yrs. old	75-84 yrs. old
MALE								
Firearms: White	2.5	31.4	26.1	21.2	19.6	19.9	26.1	39.8
Black	5.5	140.2	94.4	46.6	32.1	24.3	22.0	20.9
Accidents: White	0.7	1.8	0.8	0.6	0.5	0.4	0.6	0.7
Black	0.8	4.3	1.5	(B)	(B)	(B)	(B)	(B)
Suicide: White	0.8	15.4	15.1	14.2	14.9	16.6	23.9	38.2
Black	(B)	13.2	11.9	7.6	6.9	7.5	10.2	13.9
Homicide: White	0.9	13.6	9.8	6.3	4.0	2.8	1.5	0.8
Black	4.1	121.0	80.7	38.3	24.6	15.9	10.8	(B)

'by firearms' means '**by means of** firearms'
'by sex' means '**categorized by** sex'
Source: 1998 US Statistical Abstract. (See Table 152 for a better title)

---

4   12/03/2001

Reading Census Data — Statistical Literacy

Rate Tables
Ambiguity of '*by*'

No. 139. Age-Adjusted Death Rates, by Selected Causes: 1990 to 1996
[Rates per 100,000 population. For explanation of age-adjustment, see text, Section 2. The standard population is the total population of the United States enumerated in 1940. See also headnote, Table 138]

CAUSE OF DEATH	1990	1991	1992	1993	1994	1995
All causes	**520.2**	**513.7**	**504.5**	**513.3**	**507.4**	**503.9**
Major cardiovascular diseases	189.8	185.0	180.4	181.8	176.8	174.9
Diseases of heart	152.0	148.2	144.3	145.3	140.4	138.3
Rheumatic fever and rheumatic heart disease	1.5	1.4	1.3	1.3	1.2	1.1
Hypertensive heart disease	4.8	4.7	4.8	4.9	5.0	5.1
Hypertensive heart and renal disease	0.5	0.5	0.5	0.5	0.5	0.4
Ischemic heart disease	102.6	99.1	95.7	94.9	91.4	89.5
Other diseases of endocardium	2.5	2.5	2.6	2.6	2.6	2.6
Acute myocardial infraction	53.7	51.5	49.1	47.5	45.6	43.8
Old mycardial infraction and other	47.8	46.6	45.7	46.5	45.0	44.9
Hypertension	1.9	1.9	2.0	2.2	2.2	2.3

'by' means 'distributed by'
Source: 1998 US Statistical Abstract. Data for 1996 omitted to improve visibility of title.

---

5   12/03/2001

Reading Census Data — Statistical Literacy

Percentage Tables
Difficulties Reading

Percent of Women, 15 to 44,
Who Received Selected Medical Services

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
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: For 1995. Taken from Table 198 in 1998 U.S. Statistical Abstract.

CHOICES: Among these women ages 15-44,
a. 10% of those receiving an HIV test are ages 40-44.
b. 10% of those ages 40-44 received an HIV test.

---

6   12/03/2001

Reading Census Data — Statistical Literacy

Ambiguous Grammar
Percentages

A. *25% of males are smokers*
1. The percentage *of males* among smokers …
2. The percentage *of males* who smoke …

B. *Among smokers, the percentage of males is 45%.*
1. 45% *of males* are smokers.
2. 45% *of smokers* are males

C. *Among teens, the percentage of females who smoke*
1. Among teens, the percentage *of female* smokers …
2. Among female teens, the percentage *of smokers* ...

In each set, which statements have the same meaning?

---

7   12/03/2001

Reading Census Data

Statistical Literacy

Percentage Tables Difficulties Reading

Percentage of Smoking *Prevalence* Among U.S. Adults, 18 and Older: 1955 - 1990

Year	All	Males	Females	Whites	Blacks
1955	--	56.9	28.4	--	--
1965	42.4	51.9	33.9	42.1	45.8
1970	37.4	44.1	31.5	37.0	41.4
1980	33.2	37.6	29.3	32.9	36.9
1990	25.5	28.4	22.8	25.6	26.2

Among U.S. adults 18 and over in 1965, 45.8% is
a. the percentage of smokers who are black
b. the percentage who are black smokers
c. the percentage of blacks who are smokers

---

8   12/03/2001

Reading Census Data

Statistical Literacy

Grammar Differences: Percentage and Rates

1. Adjectives:
   "The accident rate" vs. "The accident percentage"

2. 'Of':
   "rate of inflation" vs. "percentage of inflation"

3. 'Of' and relative clause:
   "The rate of workers who are unemployed"
   "The percentage of workers who are unemployed"

4. 'Of' and 'among:
   "The rate of unemployment among workers"
   "The percentage of unemployment among workers"

---

9   12/03/2001

Reading Census Data

Statistical Literacy

Percentage tables Poor part-whole grammar

Low Birth-Weight Births by State: 1990, 1995 and 1996.

STATE	PERCENT OF BIRTHS WITH LOW BIRTH WEIGHT		
	1990	1995	1996
U.S.	7.0	7.3	7.4
AL.	8.4	9.0	9.3
AK.	4.8	5.3	5.5
AZ.	6.4	6.8	6.6
AR.	8.2	8.2	8.5
CA.	5.8	6.1	6.0

Source: Adopted from 1998 US Statistical Abstract (Some data omitted)

In the US in 1996,

a. 6% of low-weight births are in California.
b. 6% of California births are low-weight.

---

10   12/03/2001

Reading Census Data

Statistical Literacy

Presentation Rules for Ratio-Tables

1. Never use *percent* if *percentage* is needed.
   "The *percent* of women who received services…"
2. Never omit relative clause indicators.
   "The percentage of men *with* disabilities..."
3. Never use an adjective-noun for part-whole.
   "The percentage of *unemployed men*…"
4. Never use just "*by*" for a part.
   Use '*distributed by*' or appropriate grammar.
   "Death Rates [classified] by specific causes"

---

11   12/03/2001

Reading Census Data

Statistical Literacy

Jenkinson Experiment

Jenkinson (1949) notes the difficulty of
titling tables of percentages.

Jenkinson proposed
"a search for a systematic presentation
which focuses attention on basic problems
in **percentage description**."

This Jenkinson "experiment" must be
carried out.

---

12   12/03/2001

Reading Census Data

Statistical Literacy

Interpreting Averages Frequency vs. Severity

The average *US family* spends:
- more on pets than on alcohol.
- X% of their income on interest.

The *average American* is
- drinking less and working less.

**Action: If significant fraction are not
involved in the activity or condition of
interest, then show that fraction.**

---

13 12/03/2001

Reading Census Data

Statistical Literacy

Presenting Associations

Percentage of **low birth-weight births** is **higher**:
- among **black moms** than **white moms**
- among **teen moms** than **older moms**
- among **unmarried** moms than **married** moms

How strongly does this support claims of racism?

Actions:
1. **Show multivariate model of associations.**
2. **Never classify by a lesser factor without also classifying by all greater factors.**

---

14 12/03/2001

Reading Census Data

Statistical Literacy

Mission of US Bureau of the Census

*"Our mission is to be the preeminent collector and provider of **timely, relevant and quality data** about the people and economy of the United states.*

1999 Financial Report, Bureau of the Census

---

15 12/03/2001

Reading Census Data

Statistical Literacy

Purpose of Census Publications

*"The primary purpose of [Census] publications is to provide reliable and needed information to the users of statistics – a purpose that can be best served if the tabular materials are made **easy to read and to understand."***

J.C. Capt, Director, Bureau of the Census

---

16 12/03/2001

Reading Census Data

Statistical Literacy

Recommendations

Expand Mission
As the premier producer of data in the world, the US Bureau of the Census should produce data that is accurate, **useful,** *and understandable.*

- *Survey readers on readability of tabular data*
- *Identify rules for titling tables of ratios*
- *Identify fraction of groups who don't participate*
- *Provide information on priority of association*
- *Model data using multivariate regression*

---

17 12/03/2001

Reading Census Data

Statistical Literacy

References

"Describing Rates and Percentages in Tables"
    Schield, 2001 Business of Communications

"US Census Bureau Proposal" Schield, 2001 ASA

TECHNICAL
"Reading Tables of Rates and Percentages"
    Schield, 2001 ASA JSM

"Difficulties in Describing and Comparing Rates and Percentages" Schield, 2000 ASA JSM

Source: www.augsburg.edu/ppages/~schield

---

18 12/03/2001

Reading Census Data

Statistical Literacy

Acknowledgements

---

19 12/03/2001

Reading Census Data

Part-Whole Grammar of Percentage and Rates

Statistical Literacy

%1 X% of [whole] are \part/.

%2 Among [whole], X% are \part/.

P1 X% is the *percentage* of [whole] who are \part/.

P2 Among [whole], X% is the *percentage* who are \part/

P3 Among [whole], X% is the *percentage* of \part/.

R1 The *rate* of \part/ among [whole] is X per N.

R2 The \part/ *rate* among [whole] is X per N.

R3 The \part/ *rate* of [whole] is X per N.

---

20 12/03/2001

Reading Census Data

Ambiguous Grammar Rates

Statistical Literacy

1. The <u>accidental death</u> rate among *teenagers*

2. The *teenagers'* <u>accidental death</u> rate is …

3. The <u>accidental death</u> rate of *teenagers* is ...

---

4. The *teenager* <u>accidental death</u> rate is ...

5. The rate of *teenager* <u>deaths</u> is …

4 and 5 are ambiguous; possessive is unstated.

---

21 12/03/2001

Reading Census Data

Ambiguous Grammar Percentages

Statistical Literacy

A. 25% <u>of males</u> are smokers

1. The percentage *<u>of males</u>* among smokers …

2. The percentage *<u>of males</u>* who smoke …

B. Among smokers, the percentage <u>of males</u> is 45%.

1. 45% <u>*of males*</u> are smokers.

2. 45% <u>*of smokers*</u> are <u>males</u>

C. Among teens, the percentage <u>of females</u> who smoke

1. Among teens, the percentage *<u>of female</u>* smokers …

2. Among <u>female</u> teens, the percentage *of smokers* ...

Semantics: #1 never matches; #2 always matches

---

22 12/03/2001

Reading Census Data

Grammar Differences: Percentage and Rates

Statistical Literacy

1. Adjectives:
   a. Accident rate   b. Accident percentage

2. 'Of':
   a. Rate of inflation   b. Percentage of inflation

3. 'Of' and relative clause:
   a. Rate of workers who are unemployed
   b. Percentage of workers who are unemployed

4. 'Of' and 'among:
   a. Rate of unemployment among workers
   b. Percentage of unemployment among workers

---

23 12/03/2001

Reading Census Data

Ratio Table Margin Value Rules

Statistical Literacy

- **Margins** summarize data from table cells.
- **Margin values** are either **sums** or **averages**
- **Total** means 'all subjects'– not '**sum**'

If a margin value for rates or percentages:

1. is a **sum**, then the pieces are parts

2. is a **100% sum**, then the pieces are parts and their group is whole

3. is an **average**, then the pieces are wholes

A corner margin value can be a sum one way and an average the other

---

24 12/03/2001

Reading Census Data

Statistical Literacy

Statistical Literacy

Is based on three kinds of reasoning*:*

1. Proportional: describing rates and percents

2. Conditional: selecting groups (c.f. tables)

3. Contextual: making relevant comparisons

To be statistically literate, one must be able to describe and compare rates and percentages as found in tables and graphs.