Foundational Problems in Statistical Literacy

JSM 2000 ASA
August 13, 2000

MILO SCHIELD
Augsburg College

www.augsburg.edu/ppages/schield
schield@augsburg.edu

Ordered Relations 
Conditional Thinking

Conditional Probability

What is part? What is whole?

JSM 2000 ASA
August 13, 2000

MILO SCHIELD
Augsburg College

www.augsburg.edu/ppages/schield
schield@augsburg.edu

Ordered Relations 
Conditional Thinking

Difficulties Reading Rates 
and Percentages in Tables

Percentage of Smoking Prevalence Among U.S. Adults, 18 Years of Age and Older:

<table>
<thead>
<tr>
<th>Year</th>
<th>All</th>
<th>Males</th>
<th>Females</th>
<th>Whites</th>
<th>Blacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1955</td>
<td>--</td>
<td>56.9</td>
<td>28.4</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>1965</td>
<td>42.4</td>
<td>51.9</td>
<td>33.3</td>
<td>42.1</td>
<td>45.8</td>
</tr>
<tr>
<td>1970</td>
<td>37.4</td>
<td>44.1</td>
<td>31.5</td>
<td>37.0</td>
<td>41.4</td>
</tr>
<tr>
<td>1980</td>
<td>33.2</td>
<td>37.6</td>
<td>29.3</td>
<td>32.9</td>
<td>36.9</td>
</tr>
<tr>
<td>1990</td>
<td>25.5</td>
<td>28.4</td>
<td>22.8</td>
<td>25.6</td>
<td>26.2</td>
</tr>
</tbody>
</table>

CHOICES:
- a. percentage of these adults who are black smokers
- b. percentage of these black adults who are smokers
- c. percentage of these adult smokers who are black.

Ordered Relations 
Conditional Thinking

Reading Percentages:
Ambiguity of 'with' and 'to'

Source: 1998 US Statistical Abstract (Section on unmarried women omitted)

Ratio Chance-
SOURCE % of Rate Percentage Probability
1. Intro Statistics Text 55 0 90
2. Popular Essays 30 20 10 40
3. Data: 1998 U.S. Statistical Abstract 40 40 20 0

Percents are estimates at this time
Intro Statistics text: Anderson & Sweeney.

Ordered Relations 
Conditional Thinking

Conditional Thinking:
Students lack the basics

Basic
- Tables
- Graph Series
- %, Rates, Percentages, Chance, Odds, Risk
- Bayes Rule (counts)
- Arithmetic Comparisons, Likely, Attributable

Advanced
- Mean, Std.Deviation, Percentile, Z,
- Bayes Rule (Algebra)
- Correlation, Linear Regression
- ANOVA
- Logistic Regression
Describing & Comparing Rates & Percentages

Ordered Relations, Conditional Thinking

Arithmetic Comparisons of Counts and Named Ratios

- **Simple Difference**: \[(\text{Test} - \text{Base})\]
- **Simple Ratio**: \[(\text{Test} / \text{Base})\]
- **Relative Difference**: \[(\text{Test} - \text{Base}) / \text{Base}\]

**Named Ratios:** Percent, Rate, Percentage, Chance, Risk, Probability

**Attributable To:** Count or Percent
**Arithmetic Comparisons**
**Likely Family:** Risky, Probable

Grammar Difference: Rates versus Percentages

1. **Adjectives**: "accident rate" or "accident percentage"
2. **'Of'**: "Rate of inflation" or "Percentage of inflation"
3. **'Of followed by a relative clause**: "Rate of workers who are unemployed" or "Percentage of workers who are unemployed"
4. **'Of' and 'among'**: "Rate of unemployment among workers" or "Percentage of unemployment among workers"

Grammar of Rates Problems

1. The *accidental death* rate among teenagers
2. The *teenagers' accidental death* rate is …
3. The *accidental death* rate of teenagers is …
4. The *teenager accidental death* rate is …
5. The *rate of teenager deaths* is …

*These are ambiguous; possessive is unstated.*

Grammar of Rates Part is underlined

1. **The part** can be indicated several ways:
   - **Rate of**: *The rate of births is ...*
   - **Rate Modifier**: *The birth rate is ...*
   - **Verb**: *People died at a rate of*
   - **Predicate**: *People are dying at a rate of*

2. **The whole** can be indicated several ways:
   - **Rate of**: *The death rate of teenagers is ...*
   - **Possessive**: *The teenagers’ death rate ...*
   - **Among**: *Among teens, the death rate ...*

Statistical Literacy Focus on Ordered Relations

Students have difficulty with conditional probability.

- Hypothesis tests and p-values: \[P(z > k \mid H_0 \text{ is true}) \text{ with } P(H_0 \text{ is true} \mid z > k)\]
- Confidence Intervals:
  - \[P(\text{sample mean will be in interval} \mid \mu) \text{ with } P(\mu \text{ will be in the interval} \mid \text{sample mean})\]

- David Moore "What is Statistics?" MAA Notes #21

Conclusion for Statistical Literacy

Greater focus on Named Ratios:
- Percents, Rates, Percentages,
- Chance, Risk, Odds and Probability.

- Describing and comparing
- Separating association & causation,
- Separating spurious from biased,
- "Check your assumptions..."
"Percent(age) of" normally indicates a whole:
52% OF males are smokers
The percentage OF males who are smokers is 20%

"Percentage of" can indicate the part:
Among males, the percentage of smokers is 20%

Sometimes it is hard to tell.
Among men, the percentage OF smokers who run

"by" means of 'versus' categorized by
Source: 1998 US Statistical Abstract (See Table 152 for a better title)

"by" means "among" -- not 'distributed by'

Arithmetic Comparisons

<table>
<thead>
<tr>
<th>COMMON ERRORS</th>
<th>DIFFERENCE (MORE THAN)</th>
<th>RATIO (AS MUCH AS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 is 1 more than 2</td>
<td>3 is 50% more than 2</td>
<td>3 is 1.5 times [as much as] 2</td>
</tr>
<tr>
<td>3 is 0.5 times more than 2</td>
<td></td>
<td>3 is 150% times [as much as] 2</td>
</tr>
<tr>
<td>3 is 150% more than 2</td>
<td>3 is 1.5 times more than 2</td>
<td>2 is 1.5 times less than 3</td>
</tr>
</tbody>
</table>