Chapter 3: Overview

Statistical Literacy 2009
Chapter Summaries
by Milo Schield
www.StatLit.org/pdf/...
2009StatLitTextHandoutCh3.ppt
2009StatLitTextHandoutCh3.pdf

Ch 1. Review

Statistics are generally used as evidence to support an argument.

The influences on a statistic are of four kinds:
Context, Assembly, Randomness or Error.

"All Statistics are Socially Constructed"
So, "Take CARE!!"

Statistics may be influenced by:
Confounding Assembly Randomness Error

Describe Distributions: Percentiles

Table 7. Distribution of Heights for U.S. Twenty-year olds

<table>
<thead>
<tr>
<th>Percentile</th>
<th>3rd</th>
<th>5th</th>
<th>10th</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
<th>90th</th>
<th>95th</th>
<th>97th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>64.5</td>
<td>65.0</td>
<td>66.0</td>
<td>67.7</td>
<td>69.6</td>
<td>71.5</td>
<td>73.2</td>
<td>74.2</td>
<td>74.9</td>
</tr>
<tr>
<td>Female</td>
<td>69.5</td>
<td>60.1</td>
<td>61.0</td>
<td>62.6</td>
<td>64.3</td>
<td>66.0</td>
<td>67.6</td>
<td>68.6</td>
<td>69.1</td>
</tr>
</tbody>
</table>

Describe Distributions: Comparisons

2000 U.S. Family Incomes by Number of Wage Earners

<table>
<thead>
<tr>
<th># Earners</th>
<th>Median Income</th>
<th>Mean Income</th>
<th>Income per family member</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>21,916</td>
<td>27,720</td>
<td>12,054</td>
</tr>
<tr>
<td>1</td>
<td>34,423</td>
<td>50,188</td>
<td>16,779</td>
</tr>
<tr>
<td>2 or more</td>
<td>67,600</td>
<td>82,267</td>
<td>23,762</td>
</tr>
<tr>
<td>2</td>
<td>63,816</td>
<td>79,113</td>
<td>24,965</td>
</tr>
<tr>
<td>3</td>
<td>76,566</td>
<td>90,330</td>
<td>21,270</td>
</tr>
<tr>
<td>4 or more</td>
<td>91,709</td>
<td>103,678</td>
<td>19,375</td>
</tr>
<tr>
<td>ALL</td>
<td>50,890</td>
<td>65,574</td>
<td>20,865</td>
</tr>
</tbody>
</table>

Describe Distributions: Mean, Median & Mode

Context: Related factors taken into account; the confounders not taken into account.

Assembly: Choice in definition, measurement or presentation.

Randomness: Influence of chance.

Error: Systematic deviation of statistics from the underlying reality.

Review of C.A.R.E.

Hypothetical Distribution of Houses by Price
Compare Distributions: Trends

**Weight vs. Height**

College students

Ave Weight: 140.4 lbs  
Ave. Height: 68.3"

Pulse.mtw dataset

**Standardizing Totals:**

*taking into account*

State Prison Operating Expenses: CA vs. NY

<table>
<thead>
<tr>
<th>State</th>
<th>Total</th>
<th># Inmates</th>
<th>Per Inmate</th>
<th>Total</th>
<th>Per Inmate</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>$2.9B</td>
<td>136K</td>
<td>$21,385</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NY</td>
<td>$1.9B</td>
<td>69K</td>
<td>$28,426</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

50% more than 25% less than

Controlling for prison population reverses the association.

State Prison Operating Expenses: MD vs. KS

<table>
<thead>
<tr>
<th>State</th>
<th>Total</th>
<th># Inmates</th>
<th>Per Inmate</th>
<th>Total</th>
<th>Per Inmate</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD</td>
<td>$481M</td>
<td>21,623</td>
<td>$22,245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KS</td>
<td>$159M</td>
<td>7,148</td>
<td>$22,245</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3 times as much as as

Controlling for prison population nullifies the association.

**Standardizing Totals:**

*taking into account*

State Prison Operating Expenses: MN vs. ME

<table>
<thead>
<tr>
<th>State</th>
<th>Total</th>
<th># Inmates</th>
<th>Per Inmate</th>
<th>Total</th>
<th>Per Inmate</th>
</tr>
</thead>
<tbody>
<tr>
<td>MN</td>
<td>$184M</td>
<td>4,865</td>
<td>$37,825</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ME</td>
<td>$484M</td>
<td>1,424</td>
<td>$33,711</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Controlling for prison population decreases the association.

State Prison Operating Expenses: MN vs. IA

<table>
<thead>
<tr>
<th>State</th>
<th>Total</th>
<th># Inmates</th>
<th>Per Inmate</th>
<th>Total</th>
<th>Per Inmate</th>
</tr>
</thead>
<tbody>
<tr>
<td>MN</td>
<td>$184M</td>
<td>4,865</td>
<td>$37,825</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IA</td>
<td>$144M</td>
<td>5,929</td>
<td>$24,286</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Controlling for prison population increases the association.

**Standardizing Averages:**

*taking into account*

**NAEP 2000 8th Grade Math Scores: VA vs. TX**

<table>
<thead>
<tr>
<th>State</th>
<th>Encyclopedia at home</th>
<th>All (%)</th>
<th>Yes (%)</th>
<th>No (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virginia</td>
<td>275 (100%)</td>
<td>276 (81%)</td>
<td>241 (19%)</td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td>273 (100%)</td>
<td>279 (73%)</td>
<td>242 (27%)</td>
<td></td>
</tr>
</tbody>
</table>

Virginia students did better than Texas students.

After taking into account encyclopedias at home, Texas students did better than Virginia students.

**Standardizing Averages:**

*taking into account*

**SAT Verbal Scores by Race: 2002 vs. 1981**

<table>
<thead>
<tr>
<th>GROUP</th>
<th>1981</th>
<th>2002</th>
<th>CHANGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>519</td>
<td>527</td>
<td>+8</td>
</tr>
<tr>
<td>Black</td>
<td>412</td>
<td>431</td>
<td>+19</td>
</tr>
<tr>
<td>Asian</td>
<td>474</td>
<td>501</td>
<td>+27</td>
</tr>
<tr>
<td>Mexican</td>
<td>438</td>
<td>446</td>
<td>+8</td>
</tr>
<tr>
<td>Puerto Rican</td>
<td>437</td>
<td>455</td>
<td>+18</td>
</tr>
<tr>
<td>American Indian</td>
<td>471</td>
<td>479</td>
<td>+8</td>
</tr>
<tr>
<td>ALL Test takers</td>
<td>504</td>
<td>584</td>
<td>ZERO</td>
</tr>
</tbody>
</table>

SAT scores were the same in 2002 as in 1981.

After taking into account race, SAT scores were higher in 2002 than in 1981.

**Single Weighted Average**

**Weighted-Average Graph: Silverware (Jill)**

- No Percentage of Items which are Knives
- Yes Percentage of Items which are Knives
Comparing Weighted Averages

Weighted-Average Graph: Silverware (Both)

Standardizing: “Same mix”

Weighted-Average: Silverware Standardized

Family Income: Plotting the data

Mean Family Income by Race & Structure

Race  | Married | Single
--- | --- | ---
White (100%) | $60,600 (82%) | $26,700 (18%)
Black (100%) | $53,900 (48%) | $14,000 (52%)

Calculate Averages

Mean Family Income by Race & Structure

Race  | Married | Single
--- | --- | ---
White (100%) | $60,600 (82%) | $26,700 (18%)
Black (100%) | $53,900 (48%) | $14,000 (52%)

Standardizing

78% of all US families are headed by a married couple

Comparisons: Black-White Income Gap

Average Income | Before | After
--- | --- | ---
Whites | 55K | 55K
Blacks | 33K | 45K
Difference | 22K | 8K

Of the $22K black-white income gap, 14K (22-8) is explained by family structure.
67% (14/22) of the black-white income gap is explained by marital status.
Three methods

If you’re having difficulty using the graphical approach, you can use either proportional reasoning or the algebra of weighted averages. As you’ve seen, they give the same result as the graphical approach. A common error in using either is to multiply by the percentages. The proper approach is to convert the percentages to decimals before multiplying.

Here are problems associated with each of these three methods.

• Graphically: a common problem is identifying what numbers one places on the right and the left sides of the graph.
• Proportional reasoning: a common problem is identifying whether to add onto the smaller or subtract from the larger.
• Algebra: a common problem is deciding which percentage to apply to which value.

Summary

Context involves what is (not) taken into account.

What is taken into account can influence

• Counts or totals (by forming ratios)
• Averages (by selection or standardizing)

Hypothetical thinking is required to think of what could have been taken into account (confounders).

“Presenting Confounding and Standardization Graphically”