Teaching the Social Construction of Statistics

Milo Schield, Augsburg College
Dept. of Business Administration. 2211 Riverside Drive. Minneapolis, MN 55454

Abstract

Abstract: Citizens in our modern society are deluged by a torrent of statistics and they often treat these statistics as numerical facts. Yet in "*Damned Lies and Statistics*" sociologist Joel Best noted that statistics are socially constructed: they are selected, shaped and presented by human beings. In some cases the effects of social construction can be found in the data; in many cases it cannot even with all the data. And when given only selected summary statistics, most students fail to see that these statistics are most susceptible to the influence of social construction. Statistical educators have done little to help students appreciate the consequences of this fact and to realize that analyzing the influence of social construction requires hypothetical thinking: thinking about alterative ways of collecting, counting, measuring and presenting data. Hypothetical thinking is a skill that requires training. One way to develop realistic hypothetical thinking skills is through factual exercises – exercises with right-wrong answers – that help students see how small changes in selection, definition, grouping, measurement or presentation can create large changes in numeric outcomes. Examples of such factual exercises are presented. Educators in quantitative literacy are encouraged to include the social construction of data as a fundamental part of QL.

Introduction

The most-recent guidelines proposed by the American Statistical Association for introductory statistics courses noted that, "*The desired result of all introductory statistics courses is to produce statistically educated students which means that students should develop statistical literacy and the ability to think statistically*."

The College Guidelines Introductory Statistical Education (GAISE, 2006) report defined statistical literacy as "*understanding the basic language of statistics (e.g., knowing what statistical terms and symbols mean and being able to read statistical graphs), and understanding some fundamental ideas of statistics.*

But what are the fundamental ideas of statistics?

Joel Best

In "Damned Lies and Statistics," a popular book well-grounded in academic traditions, Professor Joel Best (2001) claimed that "*statistics – even official statistics such as crime rates, unemployment rates and census counts – are products of social activity,*" that "*All statistics are created through people's actions*" and that "*All statistics are social products, the results of people's efforts.*" (pp 26 & 27) At this point one might conclude that this claim is more of a fact than a big idea.

In his address to the American Statistical Association at their Joint Statistical Meeting, Professor Best (2002a) mentioned 'social construction.' He noted that, "*Numbers do not exist independent of people; understanding numbers requires knowing who counted what, and why. This is what is meant by saying that statistics are* ***socially constructed****. Sociologists use* ***social construction*** *to refer to the process by which people assign meaning to the world. This term means different things in different disciplines, but when sociologists use it, they do not imply that something is false or fanciful. All statistics, from the best to the worst, are socially constructed. All statistics are products of choices and compromises that inevitably shape, limit, and distort the outcome*."

In his talk at Augsburg College, Professor Best (2002b) was even more explicit about the nature and role of social construction. He said, "*Statistics are socially constructed: the products of social activities. There's a tendency in our culture to believe that statistics – that numbers – are little nuggets of truth. That we can come upon them and pick them up very much the way a rock collector picks up stones. A better metaphor would be to suggest that statistics are like jewels; that is, they have to be selected, they have to be cut, they have to be polished, and they have to be placed in settings so that they can be viewed from particular angles*."

He noted that, "***Social construction*** *is a term that became fashionable starting in the mid-sixties. It began in sociology but then spread to a bewildering variety of disciplines: science studies and literary theory, and so on. The term means different things in different disciplines. For some people, social construction has become a synonym for bogus, fraudulent, phony, made-up, fictional, or whatever. That's not my point. My point is not that bogus statistics are socially constructed. My point is that all statistics are socially constructed, in the sense that people make them*."

Best (2002a) claimed that, "*Statistics instruction needs to address this social process. It needs to concern itself with matter of construction – as well as calculation.*" It seems he is claiming that knowing that all statistics are socially constructed is one of the most important things to know about any statistic.

Social construction of bad statistics

Professor Best (2001, p. 32) indicated some ways that bad statistics may be socially constructed. He noted that "O*ften the ways that people create statistics is flawed: their numbers may be little more than guesses; or the numbers may be a product of poor definitions, flawed measurements, or weak sampling. These are the four ways to create bad social statistics*."

#1: Weak sampling: Best (2002, p. 52), reviews the need for good samples (both random and representative) and the problem of obtaining such samples. Convenience samples are introduced as an example of weak sampling. Statistical educators discuss these topics under the heading of sample bias.

#2: Flawed measurements: Best (2002, p. 45) focuses on the wording of questions and on the selection of possible answers as the basis for flawed measurements. Statistical educators may also discuss these topics under the heading of measurement bias.

#3: Guessing: It would seem that 'guessing' has no status as a fundamental idea in statistics. But adjusting a sample for non-representativeness is a relevant topic under sampling bias in statistics. A sample of families in Seattle that happens to include Bill and Melinda Gates might be considered non-representative. Statistical educators would not oppose removing them in order to make the sample more representative of a known or presumed population. But statistical educators are generally quiet on whether one should over-sample the poor and homeless to insure the sample is representative of the entire population when the population is really unknown. The former involves omitting an actual member of the sample; the latter involves adding members of the population that might be under-represented in the sample. The latter is closer to guessing than the former. A different variety of guessing might be estimating the deaths attributable (due) to second hand smoke when confounders are not taken into account.

#4: Poor definitions: Best (2002, p. 39) noted that "*whenever examples substitute for definitions, there is a risk that our understanding of the problem will be distorted*." And even when a complete definition is offered, Best (2002, p. 40) noted that "*No definition of a social problem is perfect.*"

Moore and the Social Construction of Statistics

David Moore (2002) indicated that the social construction of statistics may be a major element of statistical literacy. He defined statistical literacy as "*what every educated person should know [about statistics*]." In describing statistical literacy, he admonished: "*Think Broadly*." As examples of broad thinking he included,

(1) "*Is this the right question? (Who is unemployed?*") or

(2) "*Does the answer make sense? (Only 15% of new entrants into the work force will be native white males*").

Note that both examples use terms that need clear definitions: 'unemployed' in the first example, 'new entrants into the work force' in the second example.

(1) Defining 'unemployed' as "workers who are out of work and are actively looking for work" gives a smaller number than defining 'unemployed' as "workers who are out of work and who want work regardless of whether or not that are actively looking."

(2) If 'new entrants' is defined as all new employees, then native white males should be over 30% of new entrants. If 'new entrants' is defined as new employees who are not replacing a retiring employee of the same sex, race and country of origin, then native-born white males could be less than 15% of these new entrants.

In talking about the big ideas involving statistical literacy, Moore (2001) mentions, "*Where does the data come from?*" Again, Moore is clearly focusing on the process by which statistics are generated.

Based on this evidence, it appears that Moore would agree with Best that the social construction of statistics is a fundamental idea in understanding statistics.

Examples of the Social Construction of Statistics

Notice the difference in the following pairs of statistics. Notice how a small change in syntax can create a big change in semantics.

- OPEC countries supply 50% of US oil *imports*,[1] but only 30% of US oil *usage*[2].
- The average US farm was 440 acres; the average US *family* farm was 326 acres[3].
- The US median *household* income is $43,318[4]; median *family* income is $52,680[5] and median *married-couple* income is $62,261.[6]
- In 2005, the world gained 2.3 people per second (over 74 million people per year).[7]

Other Kinds of Social Construction

Moore (2001) noted another category of social construction: the failure to include a lurking variable. Moore noted two clusters of big ideas:

[1] Figure 19.2 and Table 900. 2006 U.S. Statistical Abstract

[2] Table 889, 2006 US Statistical Abstract. 12.03 QBTU Production versus 21.08 QBTU Import.

[3] Table 798 2006 U.S. Statistical Abstract. In 2002, there were 2.13 billion farms with 938 billion acres. Family farms numbered 1.91 billion and had 622 billion acres.

[4] Table 675, 2006 US Statistical Abstract for 2003.

[5] Table 678, 2006 US Statistical Abstract for 2003.

[6] Table 683, 2006 US Statistical Abstract for 2003.

[7] Figure 30.1, P. 861, 2006 US Statistical Abstract

- One cluster included "The omnipresence of variation", "Conclusions are uncertain", "Avoid inference from short-run irregularity" and "Avoid inference from coincidence."
- The other cluster included: "Beware the lurking variable," "Association is not causation", "Where did the data come from?" and "Observation versus experiment."

Is there anything in this list that influences the size of a statistic that is not included or includable under Best's four categories?

Suppose we allow Best's 'weak sampling' to include all of Moore's ideas involving variation and statistical inference. And suppose we allow that "association is not causation" and "observation versus experiment" do not generate bad social statistics but are involved in thinking critically about the statistics that are generated.

This leaves one item: "Beware the lurking variable." For more on lurking variables, see Schield (2006). Associations that fail to take into account relevant factors are not false or bad statistics, but their value is strongly influenced by that choice. As such, the influence of a relevant confounder may be a key element in the social construction of statistics.

Seeing the Social Construction of Statistics

Given that the social construction of a statistic is extremely influential in understanding its value, is there a sign of this construction in the data?

When given adequate data, one can calculate some influences of social choices in constructing a statistic:

- choice of mean versus median
- choice of focusing on selected sub-group
- choice of how to group subgroups
- choice of including/excluding outliers
- choice of what cutoff to use in forming groups
- choice of P(A|B) versus P(B|A)

But even with all the detailed data, there are some effects or influences of social construction that remain outside the data and thus are often unknown.

- sampling bias in random sampling
- non-response, respondent or researcher bias
- measurement bias (changing a question)
- changing the target or sampled population
- changing the sample size
- changing the definition of a group or measure embedded in the original data
- including a plausible confounder

The less data available (e.g., the more the data is summarized), the less that can be known about the effects of social construction. Media stories typically present only a few carefully-selected summary statistics so most of the influence of social construction on these statistics cannot be seen in the data presented. In such cases, readers must be most careful in drawing conclusions from such summaries.

Hypothetical Thinking

This lack of access to the underlying data requires hypothetical thinking in order to analyze or evaluate essays that use statistics as evidence. This hypothetical thinking is absolutely critical once one accepts that all statistics are socially constructed – they are not numerical absolutes: they are selected, defined and presented by people who have motives in seeing the statistics be large or small.

In most quantitative courses, students have all the information needed in the problem or case; it is a matter of extracting and manipulating the numbers appropriately. But in dealing with the social construction of the data, the evaluation necessarily involves hypothetical thinking about how the data was constructed – or how it could have been constructed. Hypothetical thinking involves all aspects of the social construction of data.

Hypothetical thinking is typically inductive. To be useful, alternatives must be plausible, relevant and proximate rather than arbitrary, irrelevant or distant.

Hypothetical thinking includes evaluating the influence of assembly: the choice of what to count and how to measure in generating a statistic.

- Consider evaluating the prevalence of 'bullying at school.' Was 'bullying' defined narrowly (e.g., just physical harm to school children while at school), or was it defined broadly (e.g., physical or psychic harm to the person being bullied or by the person doing the bullying in any school-related situation)? Would this difference result in a material difference in the statistic? See Best (2001 and 2004) for a complete presentation and many examples on the social construction of statistics.
- Consider evaluating the average income of a group. Was the group defined narrowly to include just the highest paid workers (e.g., permanent, full-time working adults) or broadly to include the entire population and would that change the average income by a factor of two? Was 'income' defined narrowly to include just the minimum income (e.g., W-2 earnings from wages) or broadly to include all sources of cash (e.g., including the use of credit, the sale of assets and illegal activities) and would this difference change the average income by a factor of two?

Hypothetical thinking also includes evaluating the influence of confounding: the choice of what could have been taken into account.

- In evaluating the percentage of babies which have low birth weight, did the researchers take into account whether the mother smoked, used drugs, was anorexic, wanted the baby or was under 18? Which of these is most influential? Would taking any of these into account nullify or reverse an observed association?

- In evaluating average family incomes did the researchers take into account the influence of family structure (was the family headed by a single parent or by a married couple) or work status (was the head of the family employed full time or not). Which of these is most influential? Would taking any of these into account nullify or reverse an observed association.

Hypothetical thinking also includes the assembly of presentation: the choice of how to present the data. The choice may depend on several factors: whether the item involved is a count or rate, whether something is to be diminished or enhanced, and whether the numbers being compared are small or large.

- A count can be diminished by showing it as a percentage of a much larger group (e.g., less than 1% of new-born babies die from birth defects) or enhanced by showing it as a count in a large population (e.g., over 10,000 babies die each year from birth defects world wide).
- In comparing small numbers, percentages or rates for two groups, the difference can be minimized by presenting their difference (The percentage of children who are autistic by age five is less than 3 percentage points greater for males than for females) and maximized by comparing their ratio or percentage difference (boys are three or four times as likely to become autistic by age five as are girls).
- In comparing large numbers, percentage or rates for two groups, the difference can be minimized by using a percentage difference (the distance between the earth and the sun varies by less than 4% over the course of a year) or maximized by using a simple difference (the sun is 5 million miles closer to the earth when it is closest than when it is furthest away).

Isaacson (2005) noted that hypothetical thinking involves critical thinking in ways that students find unfamiliar. Answers are seldom right or wrong, but are more or less plausible or have more or less influence.

Hypothetical thinking is required to overcome these limitations and think about other possibilities that would influence the size of a statistic – especially in dealing with the selected summary social statistics commonly found in the media. Schield (2007)

When the evidence of social construction is not available (either the data is not provided, or the evidence is outside the data), then a statistically-literate reader must think hypothetically about social construction. Evaluating statistics requires hypothetical thinking on how the statistics could have been constructed.

Can Hypothetical Thinking be Taught?

Consider these questions:

1) How can students be taught:
   - to see that all statistics are socially constructed?
   - to think hypothetically about alternatives?

2) How can student be taught to distinguish:
   - between plausible and arbitrary?
   - between material and trivial?

3) How can student be taught to focus on social constructions that involve the smallest, most plausible alternatives that would substantially change the size of a statistic?

This is all very difficult for students because all too many of their courses focus on questions with a single correct answer. Students lack training in hypothetical thinking:

- estimating magnitudes or ranges
- estimating associations or correlations
- comparing the influence of different factors
- distinguishing between plausible and arbitrary.

Teaching Hypothetical Thinking

Hypothetical thinking that is plausible and productive typically requires maturity and experience in the students. How can hypothetical thinking be taught to students who lack maturity and experience?

One way may be to present students with problems with correct-incorrect answers that illustrate the effect of social construction. Hopefully a student will see how small changes in the collection, definition, classification, measurement and presentation of data can make big changes in the size of the statistic. And having seen how small details in definition can make big changes in a statistic, they will become better readers in looking for the signs of social construction.

Two kinds of hypothetical thinking exercises are envisioned. One kind involves hypothetical situations in which the factors involved have direct quantitative effects so that the influence of a hypothetical factor can be measured. The other kind involves situations in which the factors involved do not have direct quantitative effects so the influence of a hypothetical factor cannot be measured. The first is more deductive with factual (right-wrong) answers; the second is more inductive or open-ended.

Teach Hypothetical Thinking with Factual Problems

To develop skills in hypothetical thinking, students must practice calculating the influence of small changes in factors that are known. Students need drill with factual questions to practice deductive thinking. They need to identify the influence of choices

1. on numeric answers
2. on numerically-based answers
3. on math-related ideas.
4. on factors that could influence statistics

1) Students need arithmetic problems that measure the influence of social construction on a number.

- How does the choice of a basis for comparison influence the size of the comparison?
- How does the definition of a group or measure influence the average, standard deviation, Z-score and effect size?
- How does taking into account the influence of a related factor change the size of an association?

2) Students need drill on factual questions that have a single non-numeric answer.

- Which definition gives the larger count or rate?
- Which choice of comparison gives bigger #?
- Which choice of part & whole gives bigger %? *percentage of male smokers who are runners vs. percentage of smokers who are male runners*

3) Students need to see how small changes in syntax can create big changes in semantics

- Compare numbers: 8 is 300% (3 times) more than 2.
- Compare percentages: 6% is 50% (2 percentage points) more than 4%.
- Compare counts using ordinary English:
  Gun deaths *each year* doubled in the last 50 yrs.
  Gun deaths doubled *each year* in the last 50 yrs.

4) Students need to see how changes in study design can influence what factors can or cannot influence the size of a statistic. What kinds of alternates are eliminated by

- studies being experiments vs. observational?
- studies being controlled vs. uncontrolled?
- studies being longitudinal vs. cross-sectional?
- outcomes being counts vs. ratios?

What are some possible factual exercises that involve the social construction of statistics?

Exercises involving Social Construction

Developing factual exercises involving the social construction of statistics has been a major activity of the W. M. Keck Statistical Literacy project. The Appendix presents examples of these factual problems involving the choice of the definition. These exercises are taken from a much larger set of exercises – all of which are available online to students studying statistical literacy. See Schield (2007).

Factual problems are not sufficient to prepare students for hypothetical thinking. In some cases there is no single right answer. Consider asking students two questions: (1) Are you happy **at** your College? (2) Are you happy **with** your College?

We might get a higher percentage of students who are happy *at* their college than who are happy *with* their college (or vice versa) depending on the location and status of their college. Students at a high-status, poor-location college might answer differently from students at a low-status, good-location college.

Objections

One may question the primacy of the claim that all statistics are socially constructed. CPAs have very clear standards on what constitutes an asset, liability, income or expense so that social construction at the individual level is minimal. In many areas of the physical sciences, a profession has very clear standards on how data is classified and labelled even though such standards may be absent in the social sciences and the application of epidemiological methods to social issues.

But, the desire for funding and for increased (or decreased) public interest may be opening doors for enhanced social construction of statistics even in the physical sciences. In counting the number of endangered species, the term 'species' – a very solid biological concept – was extended to included geographic subspecies (e.g., the Northern spotted owl) which in turn increased the number of endangered 'species.' The number of deaths *attributable to* second hand smoke becomes the number of deaths *due to* second-hand smoke, which becomes the number of deaths *caused by* second-hand smoke. There are no standards for journalists in reporting the results of observational studies (e.g., in going from "the prevalence of breast cancer among those eating nuts was half that among those who didn't eat nuts" to saying "eating nuts cuts the risk of breast cancer in half").

A second objection to teaching students hypothetical thinking it that it just increases their native scepticism to the point they become nihilists – seeing no value in any number or statistic used to support a claim. This outcome is certainly not a desirable goal for those who uphold science as a search for truth. But, if the focus is on which study is stronger or could be made stronger rather than on showing that all studies fall short of perfection then perhaps students can become properly cautious without becoming complete sceptics.

Conclusion

Educators in numeracy, quantitative literacy and quantitative reasoning have a choice in whether to include the following ideas as big ideas in their areas.

1. All statistics are socially constructed.
2. The social construction of all statistics influences the size of a statistic or a comparison of statistics.
3. Analyzing and evaluating the social construction of a statistic often requires hypothetical thinking.
4. Students need exercises showing how small changes in construction can yield big changes in numbers, rates or percentages.

Well-established mathematics and statistics organizations may decide to continue to omit or mention these ideas only incidentally. Newer interdisciplinary organizations such as the National Numeracy Network (NNN) might consider making these ideas a key element in their approach to numeracy across the curriculum.

Recommendations

Educators interested in quantitative literacy should consider including the social construction of statistics as a fundamental idea. Including this idea could distinguish quantitative literacy from quantitative reasoning, mathematics for liberal arts or traditional statistics. Educators should work together in developing and assessing exercises to help students think hypothetically in ways that are realistic and productive. Problems that allow students alternatives in defining groups or building models to achieve an outcome must be developed.

Acknowledgments

This paper was presented in the Quantitative Literacy session at the Midwest Sociology Association April 4-7 in Chicago. This work was performed as part of the W. M. Keck Statistical Literacy project: a project "*to support the development of statistical literacy as an interdisciplinary curriculum in the liberal arts*."

References

Best, Joel (2001). *Damned Lies and Statistics.* The Free Press.

Best, Joel (2002a). *People Count: The Social Construction of Statistics.* The 2002 Joint Statistical Meeting of the American Statistical Association.[8]

Best, Joel (2002b). *People Count: The Social Construction of Statistics.* Delivered at Augsburg College.[9]

Best, Joel (2004). *More Damned Lies and Statistics.* The Free Press.

GAISE (Guidelines for Assessing Introductory Statistics Education) 2006. Published by the ASA.

Isaacson, Marc (2005). *Statistical Literacy – An Online Course at Capella University.* ASA Proceedings of Section on Statistical Education, p. 2244-2252.

Moore, D. (2001). *Statistical Literacy and Statistical Competence in the New Century*. IASE Satellite Conference on Statistical Literacy. Seoul, Korea.[10]

Schield, Milo (2006). *Beware the Lurking Variable: Understanding Confounding from Lurking Variables Using Graphs.* STATS magazine published by the ASA, Fall 2006, Issue 46.[11]

Schield, Milo (2007). *Statistical Literacy: Factual Assessment to Support Hypothetical Thinking.* IASE Guimarães, Portugal.[12]

Appendix: Hypothetical Thinking Exercises

The following exercises all have factual (right-wrong) answers. They illustrate how differences in choices can influence the size of a statistic.

These exercises are classified into three groups: counts, shares within a group and averages of different groups. Counts and shares can be added whereas averages cannot.

Social Construction involving Counts

Ex #1: Which definition gives the larger count?
1.1: teens: 'those 13-16' versus 'those 13-19.'
1.2: heat-wave deaths: 'deaths caused by the heat wave' versus 'deaths occurring during a heat-wave.'
Answer: In both cases, the latter is the less restrictive definition and thus has the larger count.

Table 1: US Women (in millions) who had a child in 2004 by family income

< $10K	10K-19.9K	20K-24.9K	25K-29.9K	30K-34.9K	35K-49.9K	50-74.9K	75K and up
4.2	6.2	3.4	3.8	3.6	8.9	10.6	12.5

Source 2006 US Statistical Abstract. Table 88.

Ex. #2: Did rich or poor mothers have the most babies?
2.1. Define 'Rich' as 35K and up; 'Poor' as under 35K.
2.2. Define 'Rich' as 75K and up; 'Poor' as under 25K.
Answers: In 2.1, Rich mothers have more babies than Poor. In 2.2, Poor mothers have more babies than Rich.

Table 2. Estimated US persons living with AIDS by race/ethnicity for 2003 Numbers in thousands.

ALL	Non-Hispanic White	Non-Hispanic Black	Hispanic	Other
406	147	172	81	5

Source 2001 US Statistical Abstract. Table 180.

Ex. #3: Which group had the most cases of AIDS? Assume that 75% of Hispanics are white.
3.1. a. Non-Hispanic White b. Non-Hispanic Black
3.2. a. White (including white Hispanics) b. Black
Answers: In 3.1, the answer is Non-Hispanic Black. In 3.2, the answer is White (including Hispanic White).

Ex #4. Which gives the largest number in the following comparisons using the smaller as the base: simple difference, times ratio or times difference?
4.1 Compare 0.6 with 0.2.
4.2 Compare 6 million with 2 million.
Answer: In 4.1, the times ratio (3 = 0.6/0.2) is largest. In 4.2, the simple difference (4 million) is the largest.

For counts, the principle is this: the broader (the less restrictive) the definition, the greater the count provided the groups are overlapping. See Best (2002, p. 44).

Social Construction involving Shares

The term 'shares' is used to indicate shares within a single common whole.

[8] www.StatLit.org/pdf/2002BestASA.pdf
[9] www.StatLit.org/pdf/2002BestAugsburg.pdf
[10] www.stat.auckland.ac.nz/~iase/publications/7/Moore.pdf
[11] www.StatLit.org/pdf/2006SchieldSTATS.pdf
[12] www.StatLit.org/pdf/2007SchieldIASE.pdf

Table 3: Distribution of US Family Income by quintiles

1st	2nd	3rd	4th	5th
4.1%	9.6%	15.5%	23.2%	47.6%

Source Table 680 in 2006 US Statistical Abstract. Data for 2003.

Ex #5: Define and compare the incomes of the 'rich' and 'poor'. Quintiles are fifths. The 1st quintile is the bottom 20%; the 5th quintile is the top 20%.

What is the income share of the *rich*?
5.1 Define *rich* as the top 20%.
5.2 Define *rich* as the top 40%.

Compare the income share for *rich* and *poor* families as a simple ratio using the share of *poor* as the base.

5.3 Define *Rich* as the top 40%; *Poor* as bottom 60%.
a. 0.9 b. 1.25 c. 2.42 d. 5.17 e. 11.6

5.4 Define *Rich* as the top 20%; *Poor* as bottom 80%.
a. 0.9 b. 1.25 c. 2.43 d. 2.82 e. 11.6

5.5 Define *Rich* as the top 20%; *Poor* as bottom 20%.
a. 0.9 b. 1.25 c. 2.43 d. 2.82 e. 11.6

Shares are like counts: the less restrictive the definition, the greater the share (the more restrictive the definition, the smaller the share). But as a percentage, share is not proportional to count when comparing different groups.

Social Construction involving Averages
This category includes both the averages of measures and the percentages or rates (but not shares) of groups.

Ex #6. Suppose that family incomes have a mean of $60,000 and a standard deviation of $20,000. If the distribution of family incomes is normally shaped, then here are the averages for various groups:[13]

Top 1%	Top 2.5%	Top 5%	Top 10%	Top 20%
$108,600	$102,800	$97,200	$91,000	$83,600

Bottom 1%	Bottom 2.5%	Bottom 5%	Bottom 10%	Bottom 20%
$11,400	$17,200	$22,800	$29.000	$36,400

Compute the simple ratio of average incomes for rich families to poor families. Use poor families as the base.

6.1 Define *Rich* as the top 1%; *Poor* as the bottom 1%.
a. 2.30 b. 3.14 c. 4.26 d. 5.98 **e. 9.53**

6.2 Define *Rich* as the top 10%; *Poor* as bottom 10%.
a. 2.30 **b. 3.14** c. 4.26 d. 5.98 e. 9.53

6.3 Define *Rich* as the top 20%; *Poor* as bottom 20%.
a. 2.30 b. 3.14 c. 4.26 d. 5.98 e. 9.53

[13] Approximate average value of Z for various top percentages under a Normal distribution. Top 1% (AveZ=2.43), Top 2.5% (2.14), Top 5% (1.86), Top 10% (1.55), Top 20% (1.18) and Top 50% (0.56). These Z scores are approximate since they were obtained empirically and not analytically. The bottom groups have Z scores that are the negations of these (e.g., Ave Z of bottom 1% = -2.43).

Principle: Let x and y be continuous where y is the outcome of interest. Suppose y = f(x) increases monotonically over the range of x. Consider two symmetric groups of x (in percentile) starting from the two extremes where the top group and bottom groups each contain XPercentile. The further two groups are from the center of a distribution (the smaller the percentile included in each group), the greater the ratio of their average y.

Ex #7. Alzheimer's is the most common form of dementia. "'Dementia' is the general term for a group of disorders that cause irreversible cognitive decline as a result of various biological mechanisms that damage brain cells."[14] Increasing age is the greatest risk factor for Alzheimer's.

7.1 Which group has the larger count?
a. Adults over 65 b. Adults over 85

7.2 Which group has the larger count?
a. Adults knowing a person who has Alzheimer's.
b. Adults with a family member with Alzheimer's
c. Families with a family member with Alzheimer's.

7.3 Which has the higher prevalence of Alzheimer's?
a. Adults over 65 b. Adults over 85

7.4 Which has the larger number with Alzheimer's?
a. Adults over 65 b. Adults over 85

Multi-Factor Percentages
Percentages are just averages, but the grammar that indicates numerator and denominator is quite different. These exercises show how small changes in syntax can generate big changes in semantics when there are three or more factors. Moving just one factor from part to whole necessarily increases the size of the percentage.[15]

8.1 Which is larger?
A. The percentage of school-age teens who are unemployed dropouts.
B. The percentage of school-age teen dropouts who are unemployed.

8.2 Which is smaller?
A. The percentage of teen-age mothers who have low-weight babies.
B. The percentage of mothers who are teenagers with a low-weight baby.

8.3 Which is larger?
A. The percentage of adult workers who are unemployed without a high school diploma.
B. The percentage of unemployed adult workers who lack a high school diploma.

[14] www.alz.org/alzheimers_disease_alzheimer_statistics.asp
[15] P(A|BC) is necessarily greater than or equal to P(AB|C).