Call for a Dissertation or Research on Using Ordinary English to Distinguish Association from Causation and to Describe and Compare Numbers and Ratios

Students need guidance on using ordinary English to distinguish association from causation and to describe and compare statistics.

I would be most pleased to have a dissertation or research by someone, well-trained on the use of English grammar, to critically review our results as presented in our published papers:

My colleague, Tom, and I downloaded sentences and related information involving various quantitative words from the Harper-Collins World-Banks corpus: the world's largest database (at that time) of written and spoken English in which every word was tagged with its part of speech. We used the observed patterns to formulate guidance for students on ways to use ordinary English to describe and compare statistics using what we called "named ratios".

1. Title: Common Errors in Forming Arithmetic Comparisons (1999)
   Updated: www.statlit.org/pdf/1999SchieldAPDU2.pdf (Fixed error in original)
   Original: www.statlit.org/pdf/1999SchieldAPDU2-Original.pdf

2. Title: Statistical Literacy: Describing and Comparing Rates and Percentages (2000)
   Paper: www.statlit.org/pdf/2000SchieldASA.pdf

3. Title: Statistical Literacy: Reading Tables of Rates and Percentages (2001)
   Paper: www.statlit.org/pdf/2001SchieldASA.pdf
   With Appendix: www.statlit.org/pdf/2001SchieldASA2.pdf

4. Title: Grammar of Statements Involving "Chance" (with Tom Burnham, 2007)
   Paper: www.statlit.org/pdf/2007SchieldBurnhamASA.pdf

5. Title: Describing Quantitative Relationships Using Informal Grammar (2011)
   Paper: www.statlit.org/pdf/2011SchieldASA.pdf

6. Title: Quantity Words Without Numbers: Why Students use "Many" (2005)
   Paper: www.statlit.org/pdf/2005SchieldCarleton.pdf

7. Title: Numbers in the News: A Survey (with Cynthia Schield, 2007)
   Paper: www.statlit.org/pdf/2007SchieldASA.pdf

8. Title: Numbers in the News: A Survey (with Bob Raymond, 2008)
   Paper: www.statlit.org/pdf/2008RaymondSchieldASA.pdf

9. Title: Distinguishing Association from Causation in Media Headlines (w. Raymond, 2009)
   Paper: www.statlit.org/pdf/2009SchieldRaymondASA.pdf

Researchers would also have access to the downloaded materials (*.txt) used to generate the rules contained in these papers. [1999-Cobuild 1GB; 2011-World-Banks-Final 2GB] For a given keyword, there were typically three to six downloaded files based on searches at L1, R1, L2 and R2 from the node word or phrase for text and for part of speech. Finally all sentences of up to 10k involving the selected keyword were downloaded. These sentences were analyzed to identify the most common patterns.

Call for a Dissertation or Research on Using Ordinary English to Distinguish Association from Causation and to Describe and Compare Numbers and Ratios

Sample grammar questions involving descriptions or comparisons of named ratios:

1. Singular or plural verb? "One in five patients develop(s)…. 20% of patients develop(s)…..[S, P]
2. Are these the same? "The higher teachers' salaries, the higher student scores" versus "the higher student scores, the higher teachers' salaries"? [Yes]
3. Are these the same? "As teacher salaries increase, the higher student scores" versus "As we increase the salaries of teachers, the higher the student scores"? [No]
4. If "8 is 4 times as much as 2" then is it true that "8 is 4 times more than 2"? [Maybe]
5. If "8 is 4 times more than 2" then is it true that "2 is 4 times less than 8"? [Maybe]
6. Suppose that 25% of women smoke (20% of men). Can we write a one-line comparison of these two percentages without using the keyword percentage and without including these two amounts? [No]
7. Is it true that percent is an amount like volts, while percentage is a property like voltage? [Yes]
8. Is it true that a specific percent can never be an adjective of 'percentage'? [True]
9. Suppose that 5% of men run. If so, are these both valid? "A small percentage of men run"? "A small percent of men run"? [No, just the first]
10. Does "the percentage of men who run" equal "the percentage of men among runners"? [No]
11. Does "the percentage of runners who smoke among men" equal "the percentage of men who run and smoke" [No]
12. Does "the percentage of runners who smoke among men" equal "the percentage of smokers who run among men" [No]
13. Does "the death rate" equal "the rate of death"? [Yes]
14. Does "the death rate among men" equal "the male rate of death"? [No]
15. Suppose we flip a fair coin twice. What is the chance that a head on the first flip is followed by a head on the second? [Statistical educators were divided. Some said 50%, others said 25%.]
16. Is "chance of a male dying accidentally" the same as "chance that a male dies accidentally"? [Yes]
17. Is "chance of a male dying accidentally" the same as "chance of a man dying accidentally"? [Richard von Mises would say "No".]
18. If "accidental death is twice as likely to occur among men as [among] women" then is it true that "men are twice as likely to die accidentally as [are] women"? [Yes]
19. If "accidental death is twice as prevalent among men as [among] women" then is it true that "men are twice as likely to die accidentally as [are] women"? [Yes]
20. If "accidental death is twice as prevalent among men as [among] women" then is it true that "men are twice as likely to die **in** accidents as [are] women"? [Ambiguous]
21. If "car accidents are twice as likely to involve deaths among men as [among] women" then is it true that "men are twice as likely to die in car accidents as [are] women"? [Ambiguous]
22. Suppose that the auto death rate is higher in Hawaii than in Arkansas. Is this ambiguous? [Yes]

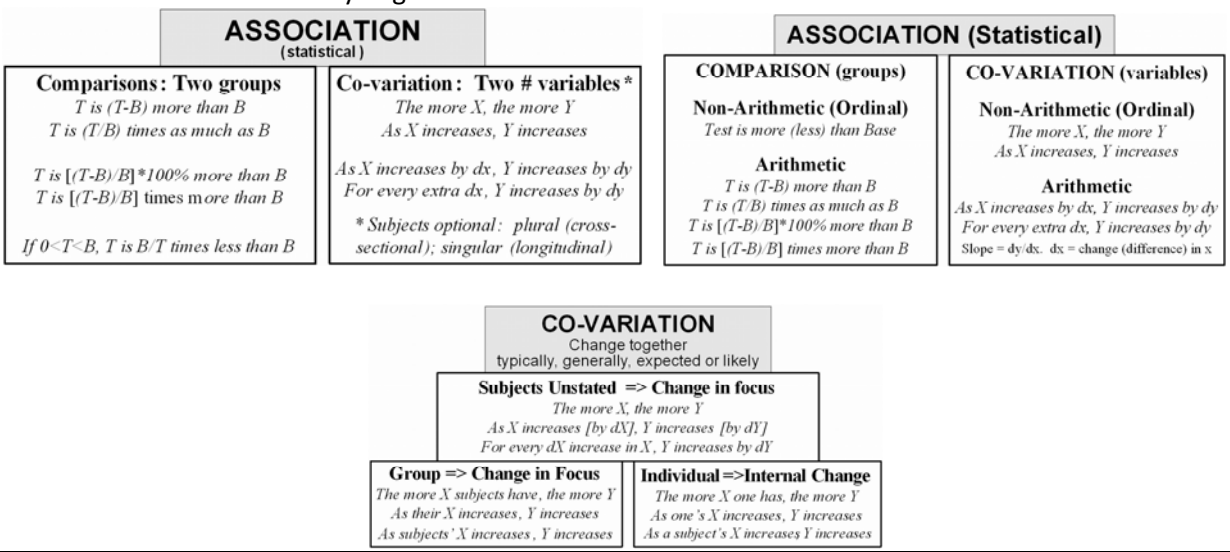
Call for a Dissertation or Research on Using Ordinary English to Distinguish Association from Causation and to Describe and Compare Numbers and Ratios

23. If "accidental death is twice as prevalent among men as [among] women" then is it true that "men are twice as prevalent to die accidentally as [are] women"? [Invalid grammar]

24. Q. Is a statistical difference or disparity sufficient to conclude there is improper discrimination?

Note: Chapters 4 and 5 of Schield (2011) contain many more examples of ambiguous ratio grammars.

Here is how we used ordinary English to describe statistical associations:



Here is how we distinguished association from causation (disparity from discrimination)

Semantics: Association is not [necessarily] Causation

A: Association	B: Between	C: Causation
Asserts an association;	Asserts an association	Asserts causation;
Says "what"	but suggest causation	Asserts "how" *
associated/association	increases, raises, ups; cut	cause, create, produce
correlation	"As x ↑, y ↓"; "more x, less y"	effect, result, consequence
Two-group comparisons:	before/after; linked, factor	Sufficient: prevent, stop
"Women live longer than men"	leads to; causal factor	"If X, then Y will happen"
"Men more likely to drink beer"	due to, because of	Contra-factual

Based on common usage by many today, but not "etched in stone" for all. * Other words OK in context. Schield V0K

Semantics: Differences or Disparities are not [necessarily] Discrimination

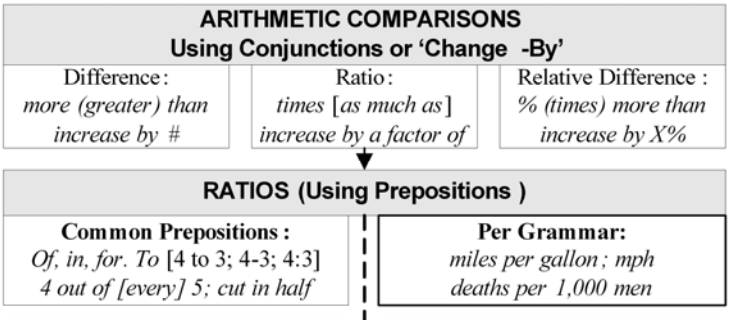
A: Association	B: Between (moral)	C: Causation (moral)
Math Differences	Descriptive Differences	Negative Differences:
Count/Rate/Amount	with a Moral Connotation	Evaluative or Judgemental
different, unequal	unequal/inequality	unfair/unjust/improper
Rank: first, second, last	disproportionate	inequity/inequitable
Superlatives: highest/lowest	discriminate: discern difference	discriminate: with prejudice
Comparatives: more, higher,	disparate/disparity	discrimination*
times as much, percent more	over/under represented	racism/sexism

* Discrimination: direct/intended (racist/sexist) vs indirect/unintended; individual vs social (systemic)

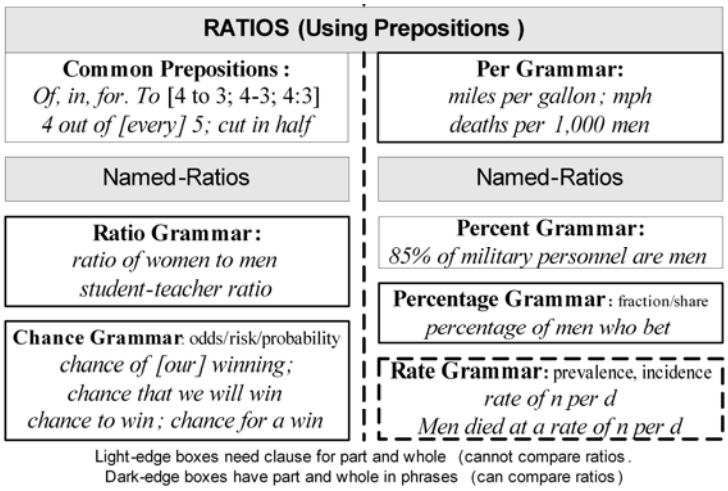
Based on common usage by many today, but not "etched in stone" for all. Schield V0K

Call for a Dissertation or Research on Using Ordinary English to Distinguish Association from Causation and to Describe and Compare Numbers and Ratios

Here is how we viewed the movement from comparisons to ratios using prepositions.



Here is how we viewed the movement from preposition ratios to named ratios.



Distribution grammar describes the allocation of a whole into or among parts in tables and graphs. It does not describe individual part-whole ratios. So it is not shown as a part-whole ratio grammar.

Note; the names shown for named ratios designate families. Here are their members:

- Percent: includes per cent.
- Percentage: includes proportion and fraction
- Rate: includes incidence and prevalence
- Chance: includes odds, risk, likelihood and probability.

"Share, proportion and fraction are used with percentage grammar at least 10% of the time they appear. But only proportion and fraction indicate a ratio. Share may use the grammar of percentage, but refers to only the part as do portion, piece, or slice. "A 10% share of the million dollar estate was his" means "A $100,000 share/portion was his." "His 10% share of the million dollar estate was ample" means "His share/portion, which was $100,000, was ample." Schield (2000)