Research in Statistics, English and Critical Thinking

March 13, 2000

MILO SCHIELD
Augsburg College
www.augsburg.edu/ppages/schield
schield@augsburg.edu

Grammar of Statistics: Rates, Percentages & Risks

Statistical Literacy
"Take Into Account"

CONTROL FOR
(Mental / Conditional)

Basic
- Tables
- Graphs
- Arithmetic Comparisons, %, Rates, Percentages, Chance, Odds, Risk

Advanced
- Mean, Std.Deviation, Percentile, Z,
- Correlation, Linear Regression
- ANOVA
- Logistic Modeling

Research Tool

Used the Cobuild corpus: the largest machine-readable corpus of English in the world.
Source: www.titania.cobuild.collins.co.uk/index.html

Grammar of Statistics

Statistical Literacy
Supporting Arguments

The Point
- Roof: point of dispute
- Walls: Support of the point assuming the reasons are true
- Floor: truth of the reasons

Control For: Experimental

Research Goal

To generate grammatical rules based on how English is actually used
- to compare any two statistics using differences, ratios and percents.
- to describe a ratio statistic percent, rate, percentage or probability.
- to compare two ratio statistics percents, rates, percentages and probabilities.

Corpus Output

Actual Text

Up to 512 characters per line centered on the node word

- the probability of point nine we'll get a hundr
- he believes that a greater probability of being caught would be a grea
- round was asked about the probability of suicide fishing on the reef.
- against over-stating the probability of falling victim. <p> Those n
- any want to increase the probability of keeping ahead of its compe
- end that there's a strong probability of waking up next April and fis
- y. And if you look at the probability of going in each direction, we
- C, D. Right? You have a probability of 80 percent of finding him is
- weather the probability of walking round town and wande
- t she had offered was the probability of a further visit to England`
- s to achieve the probability of achieving satisfactory perso
- uccessful launch and a 75 probability of a failed launch. These proba
- usiness-and even faced the probability of criminal action. Margaret to
- th problems, or where the probability of serious health consequences
- easy earnings now and the probability of steady income in the future.
Corpus Output  
Significance by Location

The node is the keyword selected (in this case 'RATE').
Locations are relative to the node. Each column is a word.
Words in columns are sorted by relative importance (t-scores)
1. interest NODE of interest
2. fixed NODE is return
3. exchange NODE for inflation
4. the NODE cut growth
5. any NODE mechanism tax
6. inflation NODE and which
7. growth NODE cuts 6
8. variable NODE rises 7
9. higher NODE cpr economic
10. cheap NODE at about

The left-1 position includes
• kind of rate: interest, exchange, inflation, etc.
• comparative measures: higher, cheap, etc., or
• appositives: e.g., the, any

Grammar Difference:
Rates versus Percentages

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

Expected the ratio statistics to form 2 families:
1. rate/chance family rate/chance of {part}
2. percentage family: percentage of {whole/part}

Found the ratio statistics to form 4 families:
1. standard percentage family: percentage of {whole}
2. rate family: rate of {part},
3. rate-percentage family: percentage of {part}
4. the chance family: chance of/that {part}.

Research Result #1:
Four Families of Ratios

Research Result #2:
Family Use Varies by Source

PERCENT (%), RATE, OR PERCENTAGE FAMILY.
Factual: “X% of this group have Y”
Generalization: “X% of the population have Y based on sample.”

CHANCE FAMILY [chance, risk, likelihood or probability]
Random Sampling Prediction:
“If you randomly sample from a population where X% of group had Y,
then the chance of the next one having Y is X%.”

Controlled Prediction:
If you start taking vitamins and if people who take vitamins are less
likely to get colds, then you will cut (reduce) your risk of colds.”

Classification of Arithmetic Comparisons
Grammar of Rates

Exception #1

Rate of' normally indicates the part (rate of deaths).
But if a modifier of 'rate' indicates part (death rate) then 'rate of' indicates the whole (death rate of men).

• the high divorce rate of their parents' generation
• the accidental drowning rate of children
• the dud rate of Air Force bombs
• the failure rate of hard disks,
• the population growth rate of the U.S.
• the occupancy rate of Kings Row

Exception #2

Usually a rate modifier is a part (death rate).
Sometimes the part is modified by a whole.

• The accidental death rate per 10,000 teenagers
• Among teenagers the accidental death rate ...
• The teenagers' accidental death rate is …
• The accidental death rate of teenagers ...
• The teenager accidental death rate is ...

Conclusion for Statistics

Statistical literacy requires more attention to the basics:
• Descriptive statistics
• Conditionality and proportionality
• Measuring association (comparison)
• Modeling of data
• Going from association to causation

Conclusion for English & Philosophy

Need more attention to grammar:
Clauses and phrases:
• Restrictive versus non-restrictive
• Relative versus subordinate
• Difference between prepositions
Prepositions:
• Meaning and role
• proper and improper use

Conclusion for Critical Thinking

Linguistics:
• Current meaning of words; rules for usage.
Statistical Literacy:
• Using statistics as evidence in arguments
Critical Thinking:
• Identify, evaluate & strengthen inductive arguments
• Relation of association and causation.
• Relation of nature, needs and values.

Conclusion for Augsburg: Interdisciplinarity

Capstone Model (Graduate school):
• Connecting different disciplines and perspectives.
• Focus on small problems using different methods
Foundation Model (Undergraduate):
• Common foundation for different disciplines
• Focus on common concepts of method:
  grammar, logic, math, computers, writing, speaking,
  critical thinking, speaking, arguing/persuading & statistics
• Focus on broad problems involving arguments:
  Philosophy, history, literature, ethics, business & politics.
Interdisciplinarity
Two Models

Capstone Model
Multiple Perspectives

1 2 3

Foundation Model
Common Skills

1 2 3

-----Majors-----

-----Majors-----

Grammar of Statistics

03/2000