Slide 1

03/2000

Grammar of Statistics

Grammar of Statistics: Rates, Percentages & Risks   1

Research in Statistics, English and Critical Thinking

March 13, 2000

MILO SCHIELD

Augsburg College

www.augsburg.edu/ppages/schield

schield@augsburg.edu

Slide 2

03/2000

Grammar of Statistics

Statistical Literacy Supporting Arguments   2

The Point

Roof: point of dispute

Walls: Support of the point assuming the reasons are true

Control Of: Experiment

Floor: truth of the reasons

Control For: Observational Study

Slide 3

03/2000

Grammar of Statistics

Statistical Literacy "Take Into Account"   3

CONTROL FOR (Mental / Conditional)

Basic
- Tables
- Graphs

Arithmetic Comparisons, %, Rates, Percentages, Chance, Odds, Risk Likely

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

Slide 4

03/2000

Grammar of Statistics

Research Goal   4

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.

Slide 5

03/2000

Grammar of Statistics

Research Tool   5

Used the Cobuild corpus: the largest machine-readable corpus of English in the world.

Source: www.titania.cobuild.collins.co.uk/index.html

During 30 days of on-line access, downloaded
- All lines involving rates, percentages, chance, odds, risk and probability (377,660 lines, 512 MB)
- All lines involving association and causation for words such as 'result of', 'due to', 'attributable to', 'contributed by', etc. (311,944 lines, 587 MB)

Slide 6

03/2000

Grammar of Statistics

Corpus Output Actual Text   6

Up to 512 characters per line centered on the node word

- ity of ninety per cent or probability of point nine we'll get a hundr
- e believes that a greater probability of being caught would be a grea
- rwood was asked about the probability of cyanide fishing on the reef.
- against over-stating the probability of falling victim. <p> Those m
- any wants to increase the probability of keeping ahead of its competi
- end that there's a strong probability of waking up next April and fin
- 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 in
- nd it's sunny so it's the probability of being sunny plus me in my ro
- round at leisure. Now the probability of walking round town and wande

- l she had offered was the probability of a further visit to England `
- n procedures increase the probability of achieving satisfactory perso
- drinks which increase the probability of sinus symptoms and these sho
- timate that there is a 25 probability of a successful launch and a 75
- uccessful launch and a 75 probability of a failed launch. These proba
- siness-and even faced the probability of criminal action. Margaret to
- th problems, or where the probability of serious health consequences
- eady earnings now and the probability of steady income in the future.

Slide 7

Grammar of Statistics

03/2000

Corpus Output
Significance by Location

7

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    <p>       economic
10.   cheap       NODE    at        about
```

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

Slide 8

Grammar of Statistics

03/2000

Grammar Difference:
Rates versus Percentages

8

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

Slide 9

Grammar of Statistics

03/2000

Research Result #1:
Four Families of Ratios

9

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}.

Slide 10

Grammar of Statistics

03/2000

Research Result #2:
Family Use Varies by Source

10

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

Intro Statistics text: Anderson & Sweeney.

Slide 11

Grammar of Statistics

03/2000

Research Result #3:
Inference Varies by Family

11

**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."

Slide 12

Grammar of Statistics

03/2000

Classification of
Arithmetic Comparisons

12
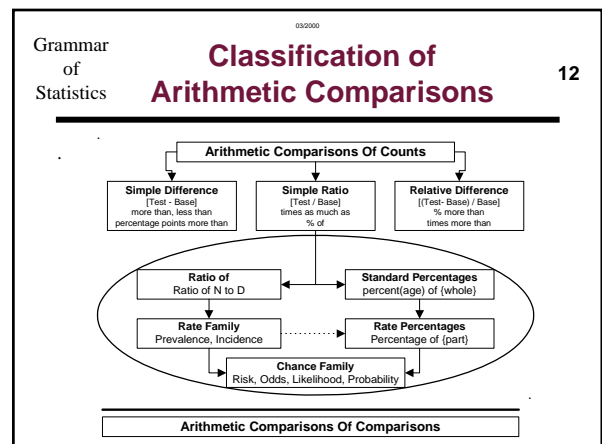
Grammar of Rates Exception #1

Grammar of Statistics — 13

Rate of' <u>normally</u> 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 <u>divorce rate</u> *of* their parents' generation
- the accidental <u>drowning rate</u> *of* children
- the <u>dud rate</u> *of* Air Force bombs
- the <u>failure rate</u> *of* hard disks,
- the population <u>growth rate</u> *of* the U.S.
- the <u>occupancy rate</u> *of* Kings Row

Grammar of Rates Exception #2

Grammar of Statistics — 14

Usually a rate modifier is a part (death rate). Sometimes the <u>part</u> is modified by a whole.

- The <u>accidental death</u> rate per 10,000 teenagers
- Among *teenagers* the <u>accidental death</u> rate ...
- The *teenagers'* <u>accidental death</u> rate is …
- The <u>accidental death</u> rate of *teenagers ...*
- The *teenager* <u>accidental death</u> rate is ...

Conclusion for Statistics

Grammar of Statistics — 15

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

Grammar of Statistics — 16

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

Grammar of Statistics — 17

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

Grammar of Statistics — 18

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.

03/2000

Grammar
of
Statistics

**Interdisciplinarity
Two Models**

19

Capstone Model
Multiple
Perspectives

------Majors-----

| 1 | 2 | 3 | | 1 | 2 | 3 |

------Majors-----

Foundation
Model
Common Skills