03/17/2000

Conditional Thinking

Teaching Conditional Thinking Using Ratios

1

Western Statistics Teachers Conference 2000

March 13, 2000

MILO SCHIELD

Augsburg College

www.augsburg.edu/ppages/schield

schield@augsburg.edu

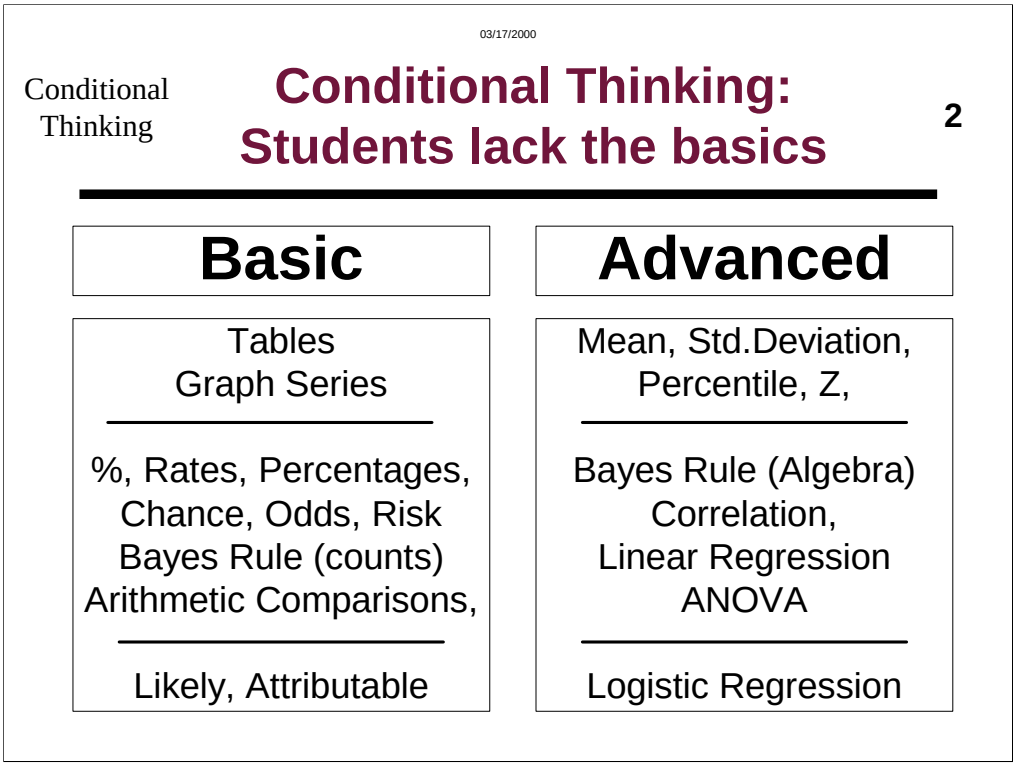
David Moore argued against teaching Bayesian reasoning in the introductory course. His primary reason was that conditional probability is "fatally subtle."

I agree with the truth of his reason.

But why is conditional probability so subtle?

Is it the probability or is it the conditional thinking.

While both are certainly involved,
I assert it is the latter more than the former.



There are different levels of conditional thinking.

We tend to teach conditional probability directly --
in confidence intervals and hypothesis tests.

We teach it using certain keywords:

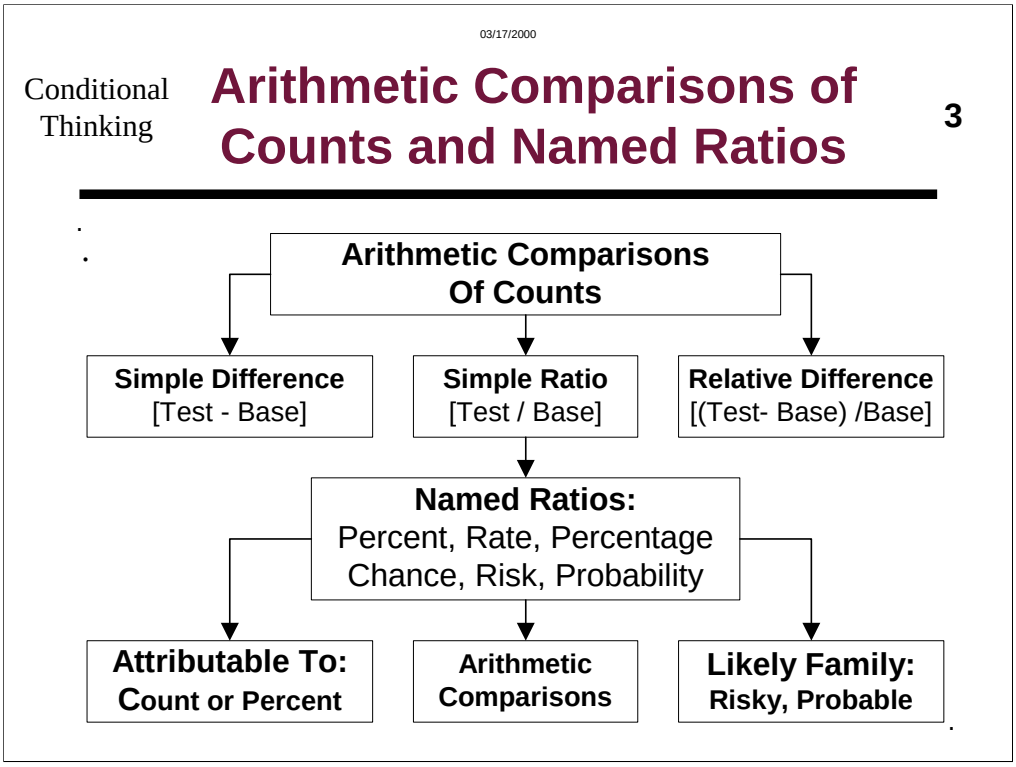
IF, WHEN and GIVEN.

These are words students use

but students use other words as well.

I assert

1. there is a much more basic level for teaching conditional thinking.
2. Student's have as much trouble at the basic level as the advanced.
3. Deficiencies in handling the basics explains most of the problems at the advanced.
4. The basics can be taught.
5. The basics should be taught before taking on conditional probability.



Consider Arithmetic Comparisons:
Three comparisons: simple difference, simple ratio and relative diff.
Each has a different grammar: more than, times as much & % more than

---

Some ratios are so common, they take into account things so elementary, that we give them names. These are what I call "Named Ratios".

These include percent, rate and percentage.
They also include the Chance family:
chance, risk, likelihood, odds and probability.
These named ratios have different grammars.

We can also do arithmetic comparisons on these named ratios.
But the grammar to handle the named ratio and the comparison is hard.

Finally there are two specialty forms of comparisons of named ratios:
1. Likely family (which is most common),
2. Attributable to family (which is increasingly common in epidem).

03/17/2000

Conditional Thinking

Grammar Difference: Rates versus Percentages

4

1. **Adjectives**: "accident rate" or "accident percentage"
2. '**Of**': "Rate of inflation" or "Percentage of inflation"
3. '**Of' followed by a relative clause:**
   "Rate of workers who are unemployed" or
   "Percentage of workers who are unemployed"
4. **'Of' and 'among:**
   "Rate of unemployment among workers" or
   "Percentage of unemployment among workers"

Exercises:

.

03/17/2000

Conditional Thinking

Named Ratio Usage Varies by Source

5

SOURCE	% of {whole}	Rate	Percentage of {part}	Chance-Probability
1. **Intro Statistics Text**	5	5	0	**90**
2. **Popular Essays**	30	20	10	**40**
3. **Data: 1998 U. S. Statistical Abstract**	40	40	**20**	0

Percents are estimates at this time

Intro Statistics text: Anderson & Sweeney.

Look first at the right columns in each row.

In the first row, statistics texts: 90% of named ratios are Chance family
And most of these use "probability".

Now look at the bottom row: the US Statistical Abstract.
The Chance family is never used.

Now look at the middle row:

03/17/2000

Conditional Thinking

Kind of Inference Varies by Named Ratio

6

PERCENTS (%), RATES, OR PERCENTAGES

Factual: "X% of this sample/group have Y"

Generalization: "X% of the population have Y."

CHANCE FAMILY: risk, likelihood or probability

Suppose smokers have a higher rate of colds [than non-smokers].

Random Sampling Prediction:

"A smoker has a higher risk of a cold [than does a non-smoker].

"If you smoke, you have a higher risk of a cold [than a non-smoker].

Controlled Prediction:

If non-smokers start smoking, they can expect to cut their risk of colds."

If you start smoking, you can expect to increase your risk of a cold."

If you smoke, you have a higher risk of a cold than if you don't smoke."

03/17/2000

Conditional Thinking

Grammar of Percentages
Part is underlined

7

"Percent(age) of" normally indicates a whole:

52% OF males are smokers

The percentage OF males who are smokers is 20%

With"among","percentage of" indicates the part:

Among males, the percentage of smokers is 20%

With "among" and a trailing relative clause, "percentage of" indicates a whole:

Among men, the percentage OF smokers who run

03/17/2000

Conditional Thinking

Percentages
Ambiguity of 'with' and 'to'

8

No. 103. Low Birth Weight and Births to Teenage Mothers and to Unmarried Women—States: 1990 to 1996

STATE	PERCENT OF BIRTHS WITH LOW BIRTH WEIGHT[1]			BIRTHS TO TEENAGE MOTHERS, PERCENT OF TOTAL		
	1990	1995	1996	1990	1995	1996
U.S.	7.0	7.3	7.4	12.8	13.1	12.9
AL	8.4	9.0	9.3	18.2	18.5	18.3
AK	4.8	5.3	5.5	9.7	11.2	11.2
AZ	6.4	6.8	6.6	14.2	15.1	15.0
AR	8.2	8.2	8.5	19.7	19.6	19.8
CA	5.8	6.1	6.0	11.6	12.4	12.0

Source: 1998 US Statistical Abstract (Section on unmarried women omitted)

Given these probabilities by race of murder,

the relative risk (1.13) is quite small.

Given these probabilities by race of victim,

the relative risk (2.6) is much larger.

But the telling condition is the fact that the high and low percentages of the death penalty by race of victim are outside the high and low percentages of the death penalty by race of murderer. This is what makes a Simpson's Paradox reversal likely in comparing the probability of the death sentence by the race of the murderer.

Recall, this was exactly what happened in our previous slide.

03/17/2000

Conditional Thinking

Grammar of Rates
Part is underlined

9

"Rate of" normally indicates the part:

The rate of births is …

A modifier of "rate" normally indicates the part:

The birth rate is …

When "rate of" is followed by a number in the predicate, then the subject and verb indicate the part:

Births occurred at a rate of 30 per 1,000...
People died at a rate of 20 per 1,000...

03/17/2000

Conditional Thinking

Grammar of Rates Exceptions

10

Sometimes the part is modified by a whole.
Sometimes "rate of" introduces 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** … **of whole*
- The *teenager* accidental death rate is ...

03/17/2000

Conditional Thinking

Rates
Ambiguity of 'by'

11

No. 149. Death Rates for Injury by Firearms, Sex, Race, and Age: 1995

[**Death rate per 100,000 population.** Deaths classified according to the ninth revision of the *International Classification of Diseases*]

ITEM	5-14 yrs. old	15-24 yrs. old	25-34 yrs. old	35-44 yrs. old	45-54 yrs. old	55-64 yrs. old	65-74 yrs. old	75-84 yrs. old
MALE								
Firearms: White	2.5	31.4	26.1	21.2	19.6	19.9	26.1	39.8
Black	5.5	140.2	94.4	46.6	32.1	24.3	22.0	20.9
Accidents: White	0.7	1.8	0.8	0.6	0.5	0.4	0.6	0.7
Black	0.8	4.3	1.5	(B)	(B)	(B)	(B)	(B)
Suicide: White	0.8	15.4	15.1	14.2	14.9	16.6	23.9	38.2
Black	(B)	13.2	11.9	7.6	6.9	7.5	10.2	13.9
Homicide: White	0.9	13.6	9.8	6.3	4.0	2.8	1.5	0.8
Black	4.1	121.0	80.7	38.3	24.6	15.9	10.8	(B)

'by means of' versus 'categorized by'

Source: 1998 US Statistical Abstract (See Table 152 for a better title)

Death and Death Rates for Injury
by Firearms, Race and Sex

Death by [means of] firearms is the part.

Race and sex are wholes [broken down by].

Solutions:

(1) Death Rates due to/for/from Firearm Injuries
by Race and Sex.

(2) Firearm-related Death Rates by Race & Sex.

.

03/17/2000

Conditional Thinking

Conclusion for Statistical Literacy

12

Greater focus on Named Ratios:
Percents, Rates, Percentages,
Chance, Risk, Odds and Probability.

- **Describing and comparing**
- **Separating association & causation,**
- **Separating spurious from biased,**
- **"Check your assumptions…"**

Descriptive Statistics: Must include strong emphasis on count-based statistics: counts, percentages and rates

Conditionality: Can be introduced naturally by using tables.

Proportionality: Very basic concept in mathematics. Use percents and rates.

Measuring association. The simplest form of association is the arithmetic comparison. Students must learn this before they take on correlation**.**

Data modeling: Modeling is another way to describe an association. Students must learn modeling before they study chance.

From association to causation: Students must learn to distinguish these two both grammatically and in reality.

03/17/2000

Conditional Thinking

Percent: Ambiguous Phrase

13

No. 113. Percent Low Birthweight, by Smoking Status, Age, and Race of Mother: 1993

[Low birthweight is defined as weight of less than 2,500 grams (5 lb. 8 oz.). Excludes California, Indiana, New York, and South Dakota, which did not require reporting of tobacco use during pregnancy]

SMOKING STATUS AND RACE OF MOTHER	All ages	AGE OF MOTHER								
		Under 15 years	15-19 years			20-24 years	25-29 years	30-34 years	35-39 years	40-49 years
			Total	15-17 years	18-19 years					
All races [1]	7.4	13.8	9.6	10.5	9.0	7.5	6.5	6.8	8.0	9.0
Smoker	11.8	14.7	10.8	11.4	10.5	10.4	11.5	13.6	16.1	17.8
Nonsmoker	6.6	13.8	9.3	10.3	8.6	6.8	5.6	5.7	6.8	7.9
Not stated	9.2	14.2	11.8	12.9	11.1	9.0	8.2	8.7	9.9	10.3
White	6.1	10.8	7.9	8.6	7.5	6.1	5.4	5.7	6.8	7.7
Smoker	10.1	14.0	10.3	11.0	9.9	9.2	9.4	10.9	13.3	14.7
Nonsmoker	5.2	10.3	7.1	7.9	6.6	5.2	4.6	4.9	5.9	6.9
Not stated	7.6	(B)	9.7	10.9	9.1	7.8	6.6	7.4	8.2	9.5
Black	13.4	16.1	13.4	13.9	13.0	12.3	13.2	14.8	16.6	17.4
Smoker	22.6	19.6	17.2	17.1	17.3	18.8	23.2	26.3	27.8	30.4
Nonsmoker	12.0	15.9	13.1	13.7	12.6	11.4	11.2	11.8	13.6	14.6
Not stated	16.9	(B)	17.5	17.5	17.4	13.9	16.9	18.4	23.7	22.7

"by" means "among" -- not 'distributed by'

Source: 1998 US Statistical Abstract

Death and Death Rates for Injury by Firearms, Race and Sex

Death by [means of] firearms is the part.

Race and sex are wholes [broken down by].

Solutions:

(1) Death Rates due to/for/from Firearm Injuries by Race and Sex.

(2) Firearm-related Death Rates by Race & Sex.

.

03/17/2000

Conditional Thinking

Rates
Non-standard usage of '*by*'

14

No. 139. Age-Adjusted Death Rates, by Selected Causes: 1990 to 1996

[**Rates per 100,000 population.** For explanation of age-adjustment, see text, Section 2. The standard population f is the total population of the United States enumerated in 1940. See also headnote, Table 138]

CAUSE OF DEATH	1990	1991	1992	1993	1994	1995
All causes	**520.2**	**513.7**	**504.5**	**513.3**	**507.4**	**503.9**
Major cardiovascular diseases	189.8	185.0	180.4	181.8	176.8	174.9
Diseases of heart	152.0	148.2	144.3	145.3	140.4	138.3
Rheumatic fever and rheumatic heart disease	1.5	1.4	1.3	1.3	1.2	1.1
Hypertensive heart disease [1]	4.8	4.7	4.8	4.9	5.0	5.1
Hypertensive heart and renal disease	0.5	0.5	0.5	0.5	0.5	0.4
Ischemic heart disease	102.6	99.1	95.7	94.9	91.4	89.5
Other diseases of endocardium	2.5	2.5	2.6	2.6	2.6	2.6
Acute mycocardial infraction	53.7	51.5	49.1	47.5	45.6	43.8
Old mycocardial infraction and other	47.8	46.6	45.7	46.5	45.0	44.9
Hypertension [1]	1.9	1.9	2.0	2.2	2.2	2.3

Source: 1998 US Statistical Abstract. Data for 1996 omitted to improve visibility of title.

Given these probabilities by race of murder,

the relative risk (1.13) is quite small.

Given these probabilities by race of victim,

the relative risk (2.6) is much larger.

But the telling condition is the fact that the high and low percentages of the death penalty by race of victim are outside the high and low percentages of the death penalty by race of murderer. This is what makes a Simpson's Paradox reversal likely in comparing the probability of the death sentence by the race of the murderer.

Recall, this was exactly what happened in our previous slide.

03/17/2000

Conditional Thinking

Statistical Literacy Conditional Thinking

15

Students have difficulty with conditional probability.

- Hypothesis tests and p-values:
  P(z>k | H_o is true) with P(H_o is true | z>k)
- Confidence Intervals:
  P(sample mean will be in interval | mu) with
  P(mu will be in the interval | sample mean).

•David Moore "What is Statistics?" MAA Notes #21

Garfield and Ahlgren, 1988. "Difficulties in Learning Basic Concepts in probability and Statistics…"

Difficulties with conditional probabilities reflect two causes:

1. Difficulties dealing with probability
2. Difficulties dealing with conditional thinking.

Having taught Critical Thinking for many years, I am strongly convinced that the 2nd element is at least as problematic as the 1st.

03/17/2000

Conditional Thinking

Arithmetic Comparisons:
Base Indicator Underlined

16

1. Simple Difference: Test minus Base.

Test is ___ **bigger/smaller** than Base

2. Simple Ratio: Test / Base

Test is __ **times** as big/large as Base

Test is __% **of** Base ["of" indicates whole]

3. Relative Difference: (Test - Base)/Base

Test is # **% bigger/smaller** than Base

Test is # **times** bigger than Base

Students who can compute the mean, median and even the standard deviation

but cannot evaluate the use of a statistic in an argument

are not statistically literate.

Students who understand probability, sampling distributions, confidence intervals and hypothesis tests

but cannot distinguish association from causation

are not statistically literate.

Students who can calculate anything

but cannot express themselves

are not statistically literate.

03/17/2000

Conditional Thinking

Grammar of Percentages Problem of Ambiguity

17

As a phrase, "percent(age) of" is ambiguous:

% OF runners

The percentage OF runners

Without a complete description, the reader doesn't know if *males* is part or whole.

This is a problem in reading tables and graphs.

03/17/2000

Conditional Thinking

Grammar of Rates Exception #1

18

"Rate of" normally indicates the part (rate of births).

But if a modifier of 'rate' indicates a part (birth rate) then 'rate of' indicates the whole (birth rate of teens).

- 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