

V1 Schield: 2019 NNN Palomar Workshop 1

Statistical Literacy: An Introduction

by
Milo Schield, NNN VP

*NNN Workshop
Palomar CC Sept 20-21, 2019*

*www.StatLit.org/pdf/
2019-Schild-NNN-Palomar-Slides.pdf*

V1 Schield: 2019 NNN Palomar Workshop 2

What are Statistics?

Statistics:

- Traditional): numerical properties of samples
- (Stat Lit): numbers in context (in reality)

Signs of context: time, place and units



- 7.7 billion is a number.
- 7.7 billion people on earth today* is a statistic.

* Sept, 2019.

V1 Schield: 2019 NNN Palomar Workshop 3

What is the Most Important Thing to Know about Statistics?

“All statistics are socially constructed.”
Joel Best, author *Lies, Damned Lies & Statistics*
This doesn't mean that there is no reality.
Means that people create statistics like diamonds

V1 Schield: 2019 NNN Palomar Workshop 4

What is the Most Important Thing to Know about Statistics?

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.

V1 Schield: 2019 NNN Palomar Workshop 5

What is the Most Important Thing to Know about Statistics?

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.

Joel Best, Sociologist. www.StatLit.org/Best.htm

V1 Schield: 2019 NNN Palomar Workshop 6

What follows from being Socially Constructed?

Numbers can't be influenced. $1 + 1 = 2$

Statistics can be influenced. $1 + 1$ may equal 2

One gallon of antifreeze and one gallon of water do not yield two gallons. The combination of large and small molecules takes up less space.

V1 Schield: 2019 NNN Palomar Workshop 7

Critical Thinking Inferences

Association is Not Causation.

V1 Schield: 2019 NNN Palomar Workshop 8

Association is not causation

This statement is ambiguous. It can mean:

- 1 Association is not sufficient to prove causation
- 2 Association provides no evidence for causation.

Teachers may intend #1; students often hear #2.

A better statement would be: *Association is typically evidence of causation somewhere.*

V1 Schield: 2019 NNN Palomar Workshop 9

A-B-C Words: A = Association

Statistical association: an observable connection.

Association:

- Height is *associated* with age in children
- Obesity is *correlated* with (related to) diabetes.

Prediction:

- Graduating from high school *predicts* success in life.

V1 Schield: 2019 NNN Palomar Workshop 10

A-B-C Words: C = Causation

Causation: Lightning *caused* (resulted in) the fire.
Insomnia is a side *effect*.

Sufficient: The more X you do, the more Y *you will get*.
*Prevent, stop, end, start, kill, produce, cure, avoid, ban, quit, block, ward off, stave off, cancel, hinder, or eliminate.*⁶

Contra-factual: Those who do X *will* get more Y *than if they had not done X*.

Schild and Raymond (2009). www.StatLit.org/pdf/2009SchildRaymondASA.pdf

V1 Schield: 2019 NNN Palomar Workshop 11

A-B-C Words: B = Between

Between words describe association but imply causation

Verbs: Red wine *cuts* cancer risk. TV *ups* kids' risk of flunking.
Gene X *increases* health risk. Smoking *raises* asthma risk.

Connectors: Nuts *linked to* cancer. Trauma *tied to* heart disease.

Contributor Diet *contributes to* diabetes. Age is *factor in* infertility

Nouns: Spinach is *asthma protector*. Bad water is a *killer*.

Logicals: Anxiety increased *due to* (because of) high stake testing

*Compare: People who take antidepressants have fewer migraines
Asthma attacks more likely for smokers *than* non-smokers.

*Covariation: As teacher pay *increases*, student scores increase.
The more hours worked, the *more likely* a promotion

*Manipulation is possible, and treatment and outcome are repeatable.

V1 Schield: 2019 NNN Palomar Workshop 12

A-B-C Words: Distribution in Headlines

Of the 2,000 news headlines analyzed⁶,
71% involved A, B or C.


Of those headlines involving A, B or C,

- **86% were "between" claims,**
- **11% sufficiency, 3% causation, 3% association.**

6. Schild and Raymond (2009).

V1 Schield: 2019 NNN Palomar Workshop 13

**Prediction: Its hopeless!
This bear is faster than we are**

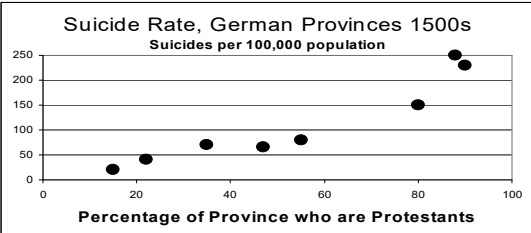


2nd hunter: No, it's not hopeless.
I don't have to outrun the bear.
I just have to outrun you!

V1 Schield: 2019 NNN Palomar Workshop 14

**Specification:
Cross-Level Inference**

Protestants more likely to suicide than Catholics?



No. Reverse was true here. Catholics more likely.

V1 Schield: 2019 NNN Palomar Workshop 15

**Statistical Studies:
A New Kind of Argument!**

Since 1900, a 100-fold increase in combined prevalence of "Studies say" and "Studies show".



Source: Google nGram:
Combined Count: "Studies say" and "Studies show"
Ignore capitalization

V1 Schield: 2019 NNN Palomar Workshop 16

**Statistical Literacy
Studies Statistics in Arguments**

The point of the argument

The more disputable the point, the stronger the evidence must be

Statistics as Evidence

"All Statistics are Socially Constructed"
So, "Take CARE"!!
Statistics may be influenced by:

C	A	R	E
Confounding	Assembly	Randomness	Error

V1 Schield: 2019 NNN Palomar Workshop 17

**Take C.A.R.E
Four Influences on Statistics**

Statistics are influenced by

C = *Confounding*: By related factors.
A = *Assembly*: By definitions and presentation.
R = *Randomness*: By uncertainty or chance
E = *Error*: By mistakes or bias.

Assembly is the etcetera category.

V1 Schield: 2019 NNN Palomar Workshop 18

**Confounding
Without Statistics**

A father and his children were on a subway.
The children were out of control: jumping on seats, yelling, and throwing things.
The father did nothing.
He slumped forward looking down at the floor, his head between his hands.
Finally an unhappy onlooker called on the father to take control of his kids.

V1 Schield: 2019 NNN Palomar Workshop 19

Confounding Without Statistics

The father looked up sadly and said:
We just left the hospital where their mom died.
 Immediately the negative judgments were transformed into pity for this family.

The onlookers were confused – confounded – by a confounder: the death of the kids’ mom.

V1 Schield: 2019 NNN Palomar Workshop 20

Statistical Influences: Confounding #1

Adults who shave their faces tend to be taller than Adults who shave their legs.

People that read home and fashion magazines are more likely to get pregnant than people that read car and sport magazines.

V1 Schield: 2019 NNN Palomar Workshop 21

Assembly Fueled Brexit? Gross vs. Net (50%):

V1 Schield: 2019 NNN Palomar Workshop 22

Statistical Influences : Assembly #1

Claims about college students:
 Administration: 80% of are ‘satisfied’.
 Students: 70% are not ‘satisfied’.

Same data:
Happy (30%), OK (50%), Unhappy (20%)

Q. Who is correct?
 A. Both are. Different definitions of ‘satisfied’.

V1 Schield: 2019 NNN Palomar Workshop 23

Statistical Influences : Assembly #2

Living with AIDs			
All (1,000)	White (non-Hispanic)	Black (non-Hispanic)	Hispanic
434	150	186	78

Two claims about groups living with AIDS:
 1. More blacks than whites.
 2. More whites than blacks.

V1 Schield: 2019 NNN Palomar Workshop 24

Assembly: Presentation

Child Abuse Statistics

Each year, more than 7,000 children in Minnesota are confirmed to be victims of physical or sexual abuse, emotional maltreatment, or neglect.

V1 Schield: 2019 NNN Palomar Workshop

Stat Influences : Randomness

Sport Illustrated Curse

Those featured as *best* by Sports Illustrated seem cursed!

They don't do as well the next year. Why not?

Being the best one year is often coincidence.

V1 Schield: 2019 NNN Palomar Workshop 26

Statistical Literacy deals with Ignorance

How was data collected?

C: What factors not taken into account?

A: How were things counted, measured, grouped?

R: How small is the group?

E: Are subjects telling the truth?

Often we don't know! We are ignorant!
We are not omniscient.

The solution? Think hypothetically! Plausible?

V1 Schield: 2019 NNN Palomar Workshop 27

Statistical Literacy is Quantitative Rhetoric

Deals with statistics in arguments.

Much of today's 'fake news' involves the use or misuse of statistics in arguments.

Students need statistical literacy in order to understand and evaluate the claims being made.

Statistical literacy is a new discipline.
Welcome aboard.

V1 Schield: 2019 NNN Palomar Workshop 28

Associations: Two Kinds

Two-group comparisons:

- Women live longer than men

Two-factor covariation: ordinal & quantitative

- The more height, the more weight
- As height increases, weight increases

- As height increases by x , weight increases by y
- For every additional x in X , Y increases by y .

V1 Schield: 2019 NNN Palomar Workshop 29

Associations: Snapshot vs. Movie

Two-factor covariation:

- As weight increases, height increases
- For every additional pound, height increases by a fifth of an inch.

Two interpretations:

- Snapshot: Change in focus (other people)
- Movie: Internal change (eat more pizza)

V1 Schield: 2019 NNN Palomar Workshop 30

Two-Group Comparisons: Math vs. Ordinary English

Arithmetic (Assembly):

- Eight is four times two. [Math speak]
- Eight is four times *as much as* two. [English]
- Eight is 300% more than two.
- *Eight is three times more than two.*

An exception:

- 8% is 6% more than 2%? [$8\% - 2\% = 6\%$?]
- 8% is 300% more than 2%. [$(8\% - 2\%)/2\%$]
- 8% is 6 percentage points more than 2%.

V1 Schield: 2019 NNN Palomar Workshop 31

Two-Group Comparisons: Math vs. Ordinary English

Another exception:
Two is four times less than eight. [Sometimes]

- Ok when the subject cannot go negative:
 Revenues, incomes, sizes, weights, prices of houses/groceries,
- Ambiguous when the values can go negative:
 Profits, temperatures, bank balances

V1 Schield: 2019 NNN Palomar Workshop 32

From Comparisons to Ratios: Using Prepositions

ARITHMETIC COMPARISONS
 Using Conjunctions or 'Change -By'

Difference: <i>more (greater) than</i> <i>increase by #</i>	Ratio: <i>times [as much as]</i> <i>increase by a factor of</i>	Relative Difference: <i>% (times) more than</i> <i>increase by X%</i>
---	---	---

↓

RATIOS (Using Prepositions)

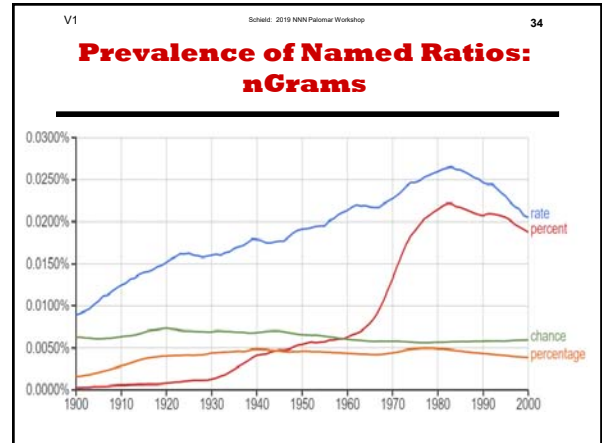
Common Prepositions: <i>Of, in, for, To [4 to 3; 4-3; 4:3]</i> <i>4 out of [every] 5; cut in half</i>	Per Grammar: <i>miles per gallon; mph</i> <i>deaths per 1,000 men</i>
--	--

V1 Schield: 2019 NNN Palomar Workshop 33

RATIOS (Using Prepositions)

Common Prepositions: <i>Of, in, for, To [4 to 3; 4-3; 4:3]</i> <i>4 out of [every] 5; cut in half</i>	Per Grammar: <i>miles per gallon; mph</i> <i>deaths per 1,000 men</i>
Named-Ratios	Named-Ratios
Ratio Grammar: <i>ratio of women to men</i> <i>student-teacher ratio</i>	Percent Grammar: <i>85% of military personnel are men</i>
Chance Grammar: odds/risk/probability <i>chance of [our] winning;</i> <i>chance that we will win</i> <i>chance to win; chance for a win</i>	Percentage Grammar: fraction/share <i>percentage of men who bet</i>
	Rate Grammar: prevalence, incidence <i>rate of n per d</i> <i>Men died at a rate of n per d</i>

Light-edge boxes need clause for part and whole (cannot compare ratios).
 Dark-edge boxes have part and whole in phrases (can compare ratios)

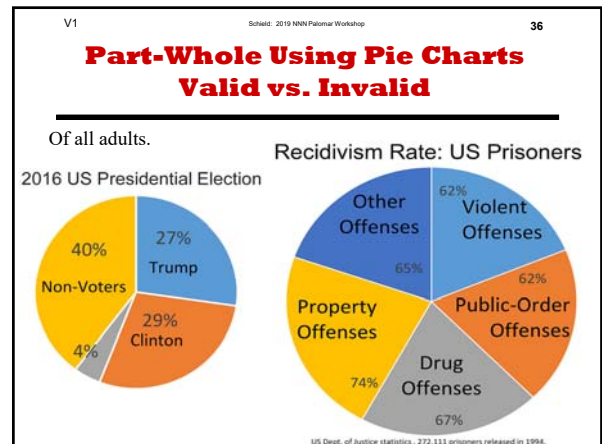


V1 Schield: 2019 NNN Palomar Workshop 35

Two Kinds of Percentages: 'percent of' and '% more'

Which kind of percentages are these:
 part-whole or percent compare?

1. One child's share of the candy.
2. Lifespan 100% longer: US than Swaziland
3. Advertisement: "40% off"
4. Mafia interest rate: 10% per month



V1 Schield: 2019 NNN Palomar Workshop 37

**Four Different Grammars:
Percent, Percentage, Rate, Chance**

- 40% of adults did not vote [for president].
Among adults, 40% did not vote [for President].
- The *percentage* of adults who didn't vote was 40%.
The percentage of non-voters among adults was 40%.
- The non-voter *rate* of adults was 40%.
The rate of non-voters among adults was 40%.
- There is a 40% *chance* that an adult was a non-voter.
Adults had a 40% chance of not voting.

V1 Schield: 2019 NNN Palomar Workshop 38

**Converting:
From Percent to Percentage**

- 60% of adults voted.
The percentage of adults who voted is 60%.
- 60% of male adults voted.
Percentage of male adults who voted is 60%.
- 60% of adults who are men voted.
Percentage of adults **who are men who voted** is 60%.
[Convert #3 to #2 first.]

V1 Schield: 2019 NNN Palomar Workshop 39

**Confusion of the Inverse:
Exchanging Part with Whole**

- Simple reversal:
"The percentage of men who are in the military" versus
"the percentage of the military who are men".
- Tricky grammar reversal:
"The percentage of smokers who are women" versus
"The percentage of smokers among women".
- Plausible claim, but the inverse is what is needed:
"Most CEOs had a pet as a child" versus
"Must children who had a pet become CEOs" or
"Children who had a pet are more likely to become CEOs".

V1 Schield: 2019 NNN Palomar Workshop 40

**Why teach Grammar?
Grammar indicates Control**

- Study design indicates "control of"
- Comparisons and ratios indicate "control for"

Control for:

- Comparisons control for a relevant basis
- Ratios control for size of group
- Comparisons of ratios control for both

V1 Schield: 2019 NNN Palomar Workshop 41

Confounding

What things block or negate confounders?

- Large effect size; large arithmetic comparison
- Study design
- Ratios
- Comparison of ratios.
- Selection and stratification
- Standardizing

V1 Schield: 2019 NNN Palomar Workshop 42

#1 Effect Size

- Does the association involve an effect size?*
If not, then no reason to think it is large
- Is the effect size material?* For example,
a factor of 10 increase in 1 chance in 10,000.
- Is the effect size statistically significant?*
- Is the effect size large enough to ward off confounders?* A: RR>4, B: RR > 3, C: RR>2, D: RR > 1.5. Schield (2018, ICOTS).

V1 Schield: 2019 NNN Palomar Workshop 43

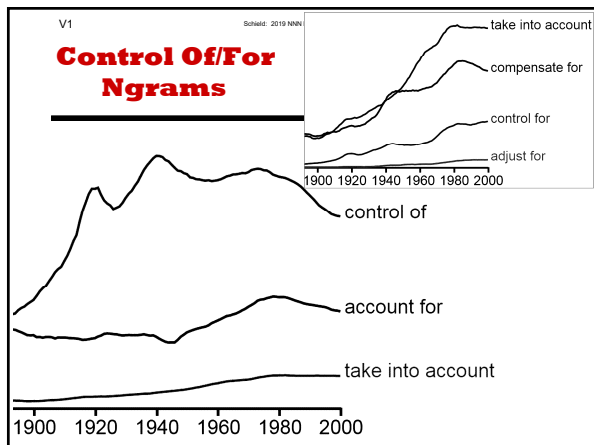
Controlling Confounding: Control Of

CONTROL OF CONFOUNDERS	
Physical Control (Grade = Quality)	
Experiment	Observational Study
A+ Scientific	C Longitudinal
A- Random Assign	D Cross-sectional
B Quasi-Exper	F Anecdotal story

V1 Schield: 2019 NNN Palomar Workshop 44

Controlling Confounding: Control For

CONTROLLING FOR CONFOUNDERS	
Take into account (mental)	
<i>Can do by hand</i>	<i>Calculator/Computer</i>
1 Select/Stratify	4 Linear Regression
2 Form Ratios	5 Logistic Regression
3 Standardize	6 Multivariate Regress



V1 Schield: 2019 NNN Palomar Workshop 46

Crude Associations

A **crude association** is an association in which nothing else has been taken into account.

More likely to get pregnant: Younger adults

- that are shorter
- that don't shave daily
- That have longer hair

What one takes into account is an assumption.
Teachers should say, "Check your assumptions."

V1 Schield: 2019 NNN Palomar Workshop 47

Conclusion

Students need a better understanding of the words and ideas involving statistics in arguments.

Statistical Literacy should be taught across the curriculum.

Learning this takes time – lots of time
Teaching this is not easy, but it is important!
Literacy is at least as important as the math!

V1 Schield: 2019 NNN Palomar Workshop 48

Statistical Literacy is Quantitative Rhetoric

Statistical Literacy studies statistics in arguments.

Much of today's 'fake news' involves the use or misuse of statistics in arguments.

Students need statistical literacy in order to understand and evaluate the claims being made.

Students need statistical literacy to become critical thinkers in a complex modern democracy.