V0G 2016 IASE-1 1

Teaching Social Statistics Association & Assembly

Milo Schield, Augsburg College
Member: International Statistical Institute
US Rep: International Statistical Literacy Project
VP. National Numeracy Network

IASE Roundtable in Berlin
July 20, 2016
www.StatLit.org/pdf/2016-Schield-IASE-1Slides.pdf

V0F 2016 IASE 1 2

Overview Teaching Social Statistics

Part 1: Stat Ed should offer 3 intro stat courses.
- Stat 100: Statistical Literacy in the Media
- Stat 101: Traditional Research Statistics
- **Stat 102: Statistics for Decision Makers**

Augsburg offers all three: 100@20 yr, 102@4 yr

Part 2: Teach multivariate thinking & confounding

Part 3: Teach Inference and confounder influence.

V0F 2016 IASE 1 3

What are Social Statistics?

This is [absolutely] the wrong place to start.

One must be very careful in making the first few steps in any journey.

The proper first question is "What are statistics?"

Different answers lead to different courses!
"Different answers" is the biggest – the most fundamental – problem in statistical education.

V0F 2016 IASE 1 4

What are Statistics? Two Definitions

1. Quantitative data from random samples – samples created by random selection (surveys) or by random assignment (clinical trials).
2. Numbers in context where the context matters. Counts and measures of real things.

This choice determines the nature of the course.

The first leads to a "Math-Stats" course;
the second leads to an "Applied" course.

V0F 2016 IASE 1 5

Statistics (#2) is Different from Mathematics

Math ignores or abstracts out the context.
a. There are no natures in mathematics
b. Math deals with variables and values
c. Math deals with associations and co-variates
d. Math has no operator for "causes"

Statistics (#2) deals with entities that have natures
a. Statistics deals with subjects & their characteristics
b. Statistics deals with "causes" and "confounders"
c. Numbers are statistics without their context
d. Mathematics is really a branch of statistics

V0F 2016 IASE 1 6

Math-Stats vs. Statistics

#1: Statistics studies variability (based on data).
#2: Statistics studies variability in context.

Certainty ------- Variability -------

Math Calculus	Probability; Discrete; Math-Stats	Statistics; Data-analysis

------- Ideal world --------- Context** Reality**

** where context (reality) matters

V0F 2016 IASE 1 7

What are Social Statistics? Two definitions

1. Random-sample data involving social conditions or activities. Typically surveys by government agencies. Focus on sampling, margin of error and bias.
2. All data involving social conditions or activities. Much – if not most -- of this data is:
   - population data (administrative systems)
   - longitudinal (time-series)
   - observational (susceptible to confounding)

V0G 2016 IASE-1 8

Teaching Statistics

"We teach the **wrong stuff**; We teach it the **wrong way**; We teach it in the **wrong order**." de Veaux

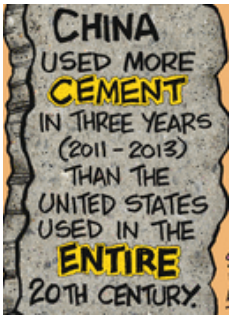
Consider teaching "Association is not causation"
- 1973 Berkeley sex discrimination case
- Ice cream sales and burglaries

Problem: These involve confounding – not chance.

Solution: Chance-based associations.
- Who gets longest run in 10 flips of a coin?
- How can we distinguish luck from skill?

V0G 2016 IASE-1 9

What are Social Statistics?



a. Statistics involving people
b. Statistics obtained from observational studies.

V0F 2016 IASE 1 10

How Should We Teach a Social Statistics Course?

Wrong question!

First answer these questions:
- Who are the students in Introductory Statistics?
- What are their goals and attitudes?
- What aspects of statistics will help them in their major?

Then answer this:
- What are the primary contributions of statistics to human knowledge?

V0F 2016 IASE 1 11

Goals of the Students; Perspectives of the Teachers

Students' majors. Teachers' disciplinary home

	Stat 101	ASA		Stat 101	ASA
Major/Home	Students	Teachers		Students	Teachers
Business-Econ	38%	7%			
Social Science	19%	2%		70%	16%
Health	13%	7%			
Psychology	10%	0%			
Engineering	9%	0%		28%	5%
Biological Sc.	9%	5%			
Math-Stats	1%	75%		2%	75%
Other	1%	4%			4%

V0F 2016 IASE 1 12

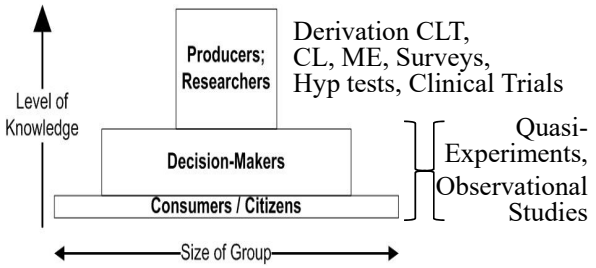
Stat 101 students: What are their Attitudes?

- Many (most?) see less value in statistics after the course than before.
- "Least valuable course in the Business-Econ core." Augsburg Business-Economics majors.
- Lost almost half of the course gain within 4 months Tintle et al, (2012) *SERJ*.
- Lost 33% of what they knew on their final within 12 month in an online course. Nadir (2004)
- Less than 0.2% will major in statistics (US nationwide).

www.amstat.org/misc/StatsBachelors2003-2013.pdf 1,135 stat majors in 2013 at 32 colleges

V0F 2016 IASE 1 13

Three Audiences: Three Courses



V0G 2016 IASE-1 14

We need to teach Statistics for Decision Makers

"One size fits all" doesn't work any more.
We should drop the idea of "*the course*" in intro stats.

We should support three algebra-intro statistics courses:

Stat 102: ***Statistics for Decision Makers.*** *Some Algebra*

Stat 101: *Traditional Research-Inference. Most Algebra*

Stat 100: *Statistical Literacy. Media-base, little Algebra*

At least half of all intro sections should be Stat 102.

V0F 2016 IASE 1 15

Association vs. Causation

Many (most?) students think that "association" is a collection of people with similar interests/goals, or a group of teams in a given sport or league.

1) Compare values of variables: As X ↑, Y ↓.
2) Compare averages of groups:
   Young adults with a bachelor's degree earned 62% more than high school completers.

   If you get a B.A., you can expect to earn 62% more than if you just complete high school.

V0F 2016 IASE 1 16

Social Statistics: Associations

Baseball players whose names begin with the letter "D" are more likely to die young

Asian-Americans are most susceptible to heart attacks on the fourth day of the month

---

Drinking a full pot of coffee every morning will add years to your life, but one cup a day increases the risk of pancreatic cancer.

Source: *Standard Deviations: Flawed Assumptions, Tortured Data, and Other Ways to Lie with Statistics* by Gary Smith (2015).

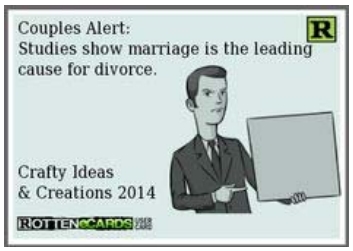
V0F 2016 IASE 1 17

Association vs. Causation Using Ordinary English

How Climate Change is Fueling Rise in Shark Attacks
www.yahoo.com/news/climate-change-fueling-rise-shark-145333862.html

V0F 2016 IASE 1 18

Association vs. Causation: Can be Tricky

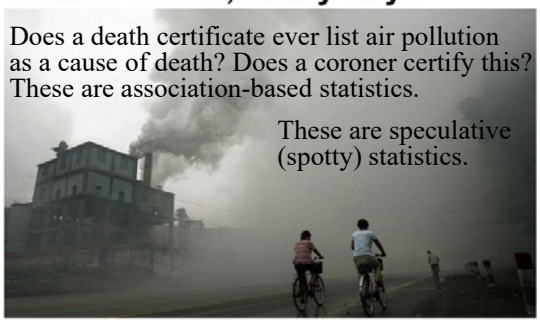


V0F 2016 IASE 1 19

Air Pollution Linked to 6.5 Million Deaths a Year, Study Says

Does a death certificate ever list air pollution as a cause of death? Does a coroner certify this? These are association-based statistics.

These are speculative (spotty) statistics.



Melissa Chan Time June 27, 2016

V0F 2016 IASE 1 20

Association vs. Causation Using Ordinary English

1. Study: 45,000 Uninsured *Die* a Year (CBS News)
2. 45,000 deaths *attributable to* uninsurance
3. 45,000 US deaths *associated with* lack of insurance
4. No health coverage *tied to* 45,000 deaths a year
5. Lack of insurance *linked to* 45,000 deaths
6. Study: 45,000 U.S. Deaths *From* Lack of Insurance
7. One death every 12 minutes *due to* no health insurance
8. 45,000 ... die *because of* lack of health insurance
9. Lack of Health Insurance *Kills* 45,000 a Year
10. Lack of Health Insurance *cause* 44,789 deaths
11. Lack of insurance *to blame for* almost 45,000 deaths

V0F 2016 IASE 1 21

Observational Statistics More Influences

Randomization eliminates many types of influence.
Inference models eliminate many others.
Teaching random-sample statistics is simpler.

Observational statistics have a host of influences.
Teaching observational statistics is harder.

Students *need a structure* that groups these influences into three or four categories.

V0F 2016 IASE 1 22

Observational Statistics are Influenced By:

Confounding: [See Part 2]
- what was – and was not – controlled for
- what kind of study was involved.

Assembly: [See Part 1]
- how the statistics were selected, collected, defined, grouped, summarized, compared and presented.

Randomness [See Part 3]

Error/bias

Statistical admonition: "Take **CARE**"

V0F 2016 IASE 1 23

Stats as Premise(Crit. Thinking) Stats as Conclusion (Stat Literacy)

The Point

An argument is like a house.
The more disputable the point,
the stronger the evidence
(the walls and floors) must be.

Statistics As Evidence

"All statistics are socially constructed
So, "Take CARE"!!
Statistics may be influenced by:

C	A	R	E
Context	Assembly	Randomness	Error

V0F 2016 IASE 1 24

Assembly: Defining Groups

US children: Elevated levels of lead:

- 27,000 in 2009
- 259,000 in 2010 [Almost a factor of 10]

In 2010, the CDC reduced the minimum for elevated levels of lead from 10 to five*.

* micrograms per dl of blood
www.cdc.gov/nceh/lead/data/StateConfirmedByYear1997-2011.htm

V0F 2016 IASE 1 25

Small Changes in Syntax; Big Changes in Semantics

V0F 2016 IASE 1 26

Assembly in Moore's Concepts & Controversies

Who is a smoker?

When estimating a proportion p, be sure you know what counts as a "success." The news says that 20% of adolescents smoke. Shocking. It turns out that this is the percentage who smoked at least once in the past month. If we say that a smoker is someone who smoked in at least 20 of the past 30 days and smoked at least half a pack on those days, fewer than 4% of adolescents qualify.

V0F 2016 IASE 1 27

Assembly in Presentation 7 nano grams/gram (7/Million)

V0F 2016 IASE 1 28

Why Teachers Don't Want to Teach Assembly

1. Ordinary English is too ambiguous.
2. Leave this up to subject-matter experts.
3. This is not really "statistics"
4. Teaching it requires subject-matter expertise.

1. Ordinary English is how statistics are communicated
2. If we don't teach it, students will never see it.
3. We define what is really "statistics".
4. We can teach it without subject-matter expertise:

Which is bigger in a class: (1) # of students, (2) # of male students, or (3) # of students in or waiting?

V0F 2016 IASE 1 29

Contributions of Statistics to Human Knowledge

Statistics are numbers in a context
Association is not causation

RANDOMNESS and CAUSATION	CONFOUNDING and CAUSATION
Chance, independence and sampling distributions	Comparisons, ratios, models and study designs
Margin of error, hyp tests & statistical significance	Epidemiological causation (Bradford-Hill)
Random assignment and causation (Fisher: RCT)	*Confounder conditions for nullification (Cornfield)*

V0.7 Conditional probability, medical tests and Bayesian reasoning
Coincidence, Simpson's Paradox and regression to the mean

V0F 2016 IASE 1 30

Most Important Topics/Ideas Augsburg StatLit Students

1 Classify different kinds of influence (Take CARE)
2 Confounding
2 Hypothetical thinking: Plausible confounders, plausible definitions [Assembly].
4 Statistics are more than numbers [Assembly]
5 Association-causation & Randomness (Luck vs. skill)
5 Bias: Placebo, single blind; double blind
5 Named Ratio grammar; Percent, Percentages, Rates

V0G 2016 IASE-1 1

Teaching Social Statistics Association & Assembly

Milo Schield, Augsburg College

Member: International Statistical Institute

US Rep: International Statistical Literacy Project

VP. National Numeracy Network

IASE Roundtable in Berlin

July 20, 2016

www.StatLit.org/pdf/2016-Schield-IASE-1Slides.pdf

V0F 2016 IASE 1 2

Overview
Teaching Social Statistics

Part 1: Stat Ed should offer 3 intro stat courses.

- Stat 100: Statistical Literacy in the Media
- Stat 101: Traditional Research Statistics
- **Stat 102: Statistics for Decision Makers**

Augsburg offers all three: 100@20 yr, 102@4 yr

Part 2: Teach multivariate thinking & confounding

Part 3: Teach Inference and confounder influence.

V0F 2016 IASE 1 3

What are Social Statistics?

This is [absolutely] the wrong place to start.

One must be very careful in making the first few steps in any journey.

The proper first question is "What are statistics?"

Different answers lead to different courses!
"Different answers" is the biggest – the most fundamental – problem in statistical education.

V0F 2016 IASE 1 4

What are Statistics? Two Definitions

1. Quantitative data from random samples – samples created by random selection (surveys) or by random assignment (clinical trials).
2. Numbers in context where the context matters. Counts and measures of real things.

This choice determines the nature of the course.

The first leads to a "Math-Stats" course;
the second leads to an "Applied" course.

V0F 2016 IASE 1 5

Statistics (#2) is Different from Mathematics

Math ignores or abstracts out the context.

a. There are no natures in mathematics
b. Math deals with variables and values
c. Math deals with associations and co-variates
d. Math has no operator for "causes"

Statistics (#2) deals with entities that have natures

a. Statistics deals with subjects & their characteristics
b. Statistics deals with "causes" and "confounders"
c. Numbers are statistics without their context
d. Mathematics is really a branch of statistics

V0F 2016 IASE 1 6

Math-Stats vs. Statistics

#1: Statistics studies variability (based on data).
#2: Statistics studies variability in context.

Certainty ------- Variability -------

Math Calculus	Probability; Discrete; Math-Stats	Statistics; Data- . analysis

------- Ideal world --------- Context** Reality**

** where context (reality) matters

What are Social Statistics? Two definitions

1. Random-sample data involving social conditions or activities. Typically surveys by government agencies. Focus on sampling, margin of error and bias.

2. All data involving social conditions or activities. Much – if not most -- of this data is:
   - population data (administrative systems)
   - longitudinal (time-series)
   - observational (susceptible to confounding)

Teaching Statistics

"We teach the **wrong stuff**; We teach it the **wrong way**; We teach it in the **wrong order**." de Veaux

Consider teaching "Association is not causation"

- 1973 Berkeley sex discrimination case
- Ice cream sales and burglaries

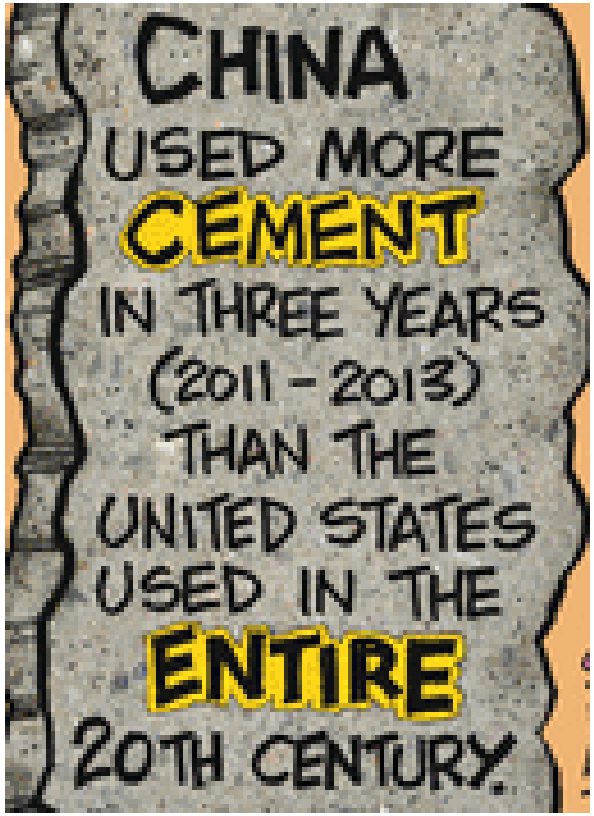
Problem: These involve confounding – not chance.

Solution: Chance-based associations.

- Who gets longest run in 10 flips of a coin?
- How can we distinguish luck from skill?

V0G 2016 IASE-1 9

What are Social Statistics?



a. Statistics involving people

b. Statistics obtained from observational studies.

How Should We Teach a Social Statistics Course?

Wrong question!

First answer these questions:

- Who are the students in Introductory Statistics?
- What are their goals and attitudes?
- What aspects of statistics will help them in their major?

Then answer this:

- What are the primary contributions of statistics to human knowledge?

V0F 2016 IASE 1 11

Goals of the Students; Perspectives of the Teachers

Students' majors. Teachers' disciplinary home

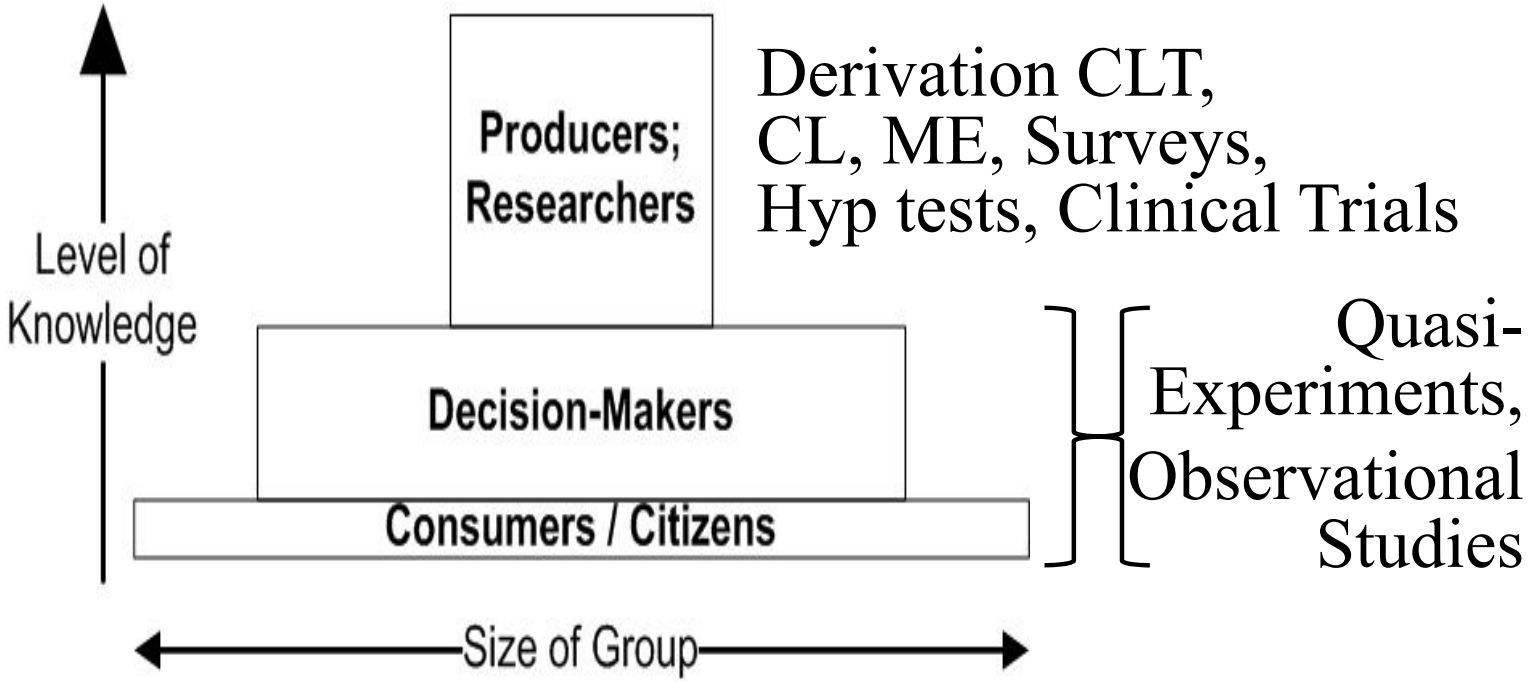
Major/Home	Stat 101 Students	ASA Teachers		Stat 101 Students	ASA Teachers
Business-Econ	38%	7%			
Social Science	19%	2%		70%	16%
Health	13%	7%			
Psychology	10%	0%			
Engineering	9%	0%		28%	5%
Biological Sc.	9%	5%			
Math-Stats	1%	75%		2%	75%
Other	1%	4%			4%

Stat 101 students: What are their Attitudes?

- Many (most?) see less value in statistics after the course than before.
- "Least valuable course in the Business-Econ core." Augsburg Business-Economics majors.
- Lost almost half of the course gain within 4 months Tintle et al, (2012) *SERJ*.
- Lost 33% of what they knew on their final within 12 month in an online course. Nadir (2004)
- Less than 0.2% will major in statistics (US nationwide).
  www.amstat.org/misc/StatsBachelors2003-2013.pdf 1,135 stat majors in 2013 at 32 colleges

V0F 2016 IASE 1 13

Three Audiences: Three Courses



We need to teach Statistics for Decision Makers

"One size fits all" doesn't work any more.

We should drop the idea of "*the course*" in intro stats.

We should support three algebra-intro statistics courses:

Stat 102: ***Statistics for Decision Makers***. *Some Algebra*

Stat 101: *Traditional Research-Inference. Most Algebra*

Stat 100: *Statistical Literacy. Media-base, little Algebra*

At least half of all intro sections should be Stat 102.

Association vs. Causation

Many (most?) students think that "association" is a collection of people with similar interests/goals, or a group of teams in a given sport or league.

1) Compare values of variables: As X ↑, Y ↓.

2) Compare averages of groups:

Young adults with a bachelor's degree earned 62% more than high school completers.

If you get a B.A., you can expect to earn 62% more than if you just complete high school.

Social Statistics: Associations

Baseball players whose names begin with the letter "D" are more likely to die young

Asian-Americans are most susceptible to heart attacks on the fourth day of the month

---

Drinking a full pot of coffee every morning will add years to your life, but one cup a day increases the risk of pancreatic cancer.

Source: *Standard Deviations: Flawed Assumptions, Tortured Data, and Other Ways to Lie with Statistics* by Gary Smith (2015).

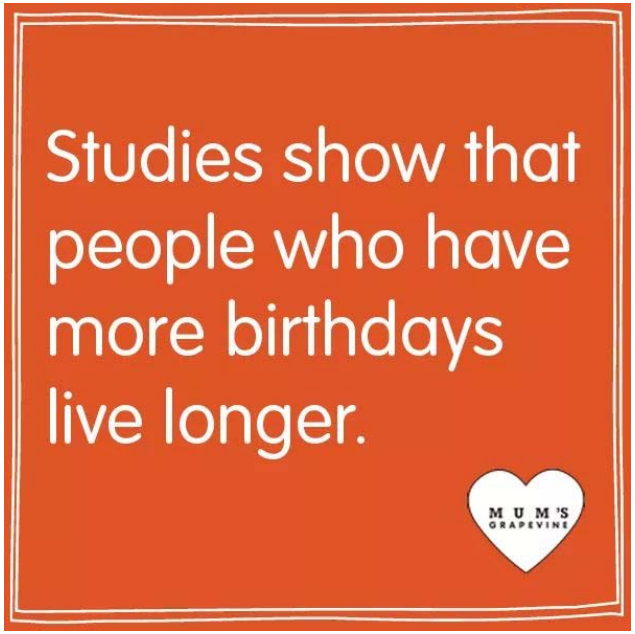
V0F 2016 IASE 1 17

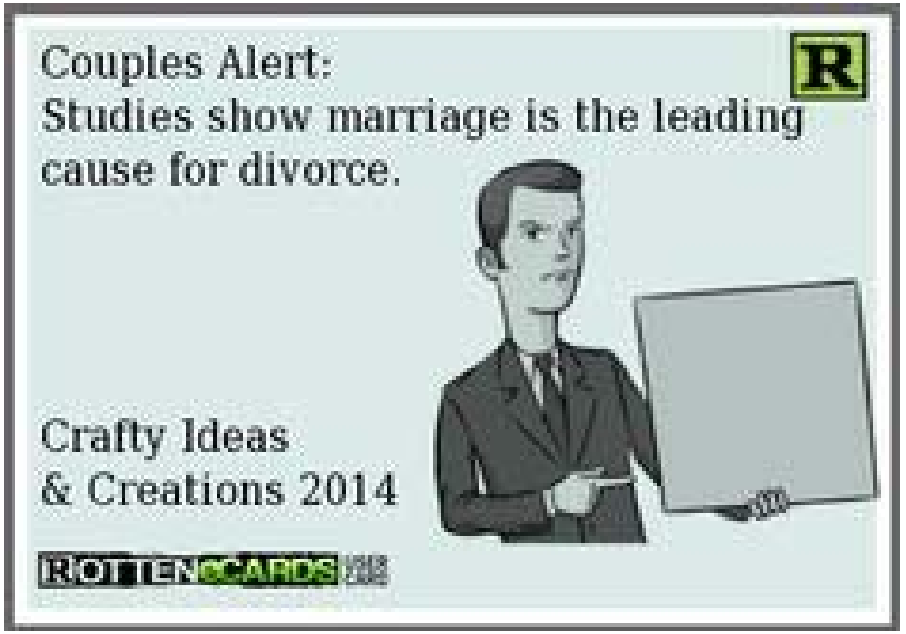
Association vs. Causation Using Ordinary English

How Climate Change is Fueling Rise in Shark Attacks

www.yahoo.com/news/climate-change-fueling-rise-shark-145333862.html

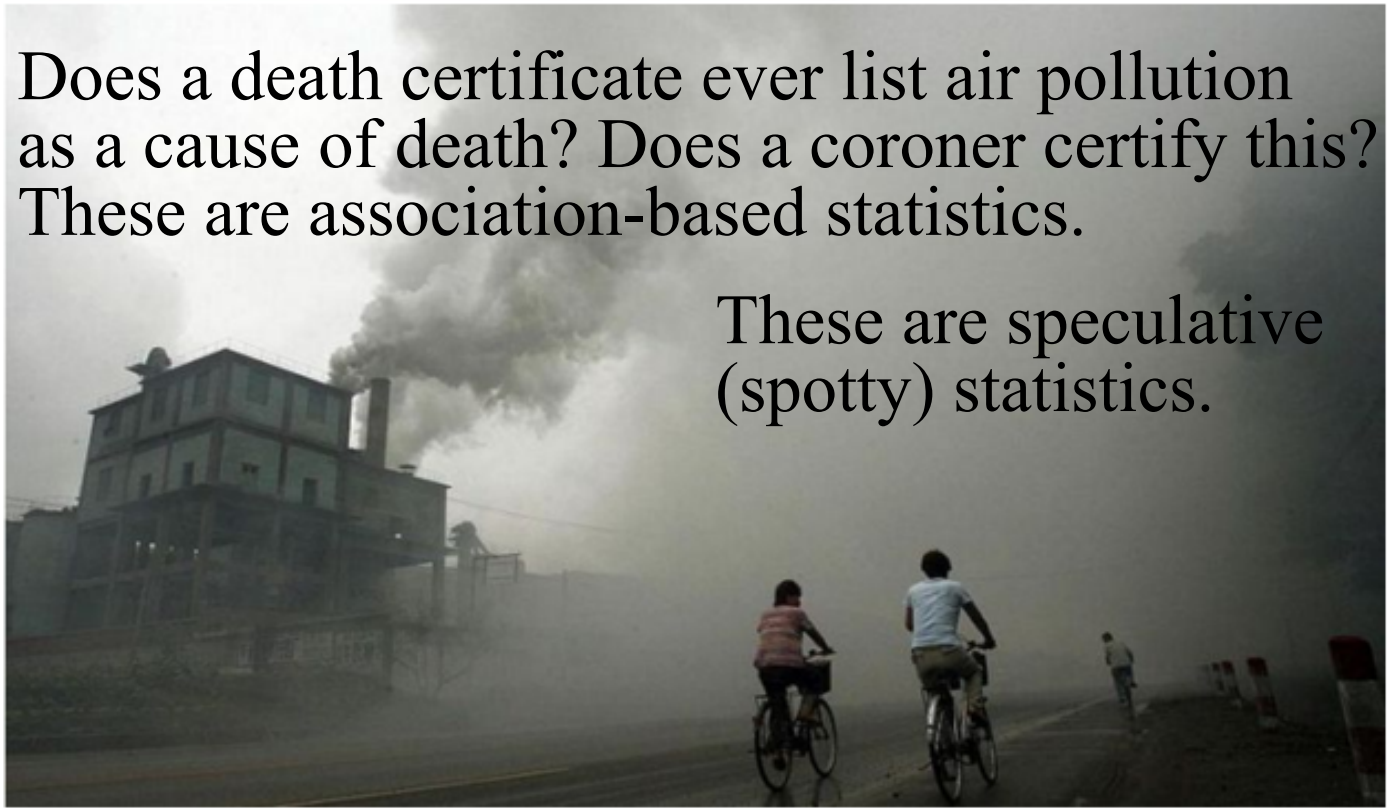
Association vs. Causation: Can be Tricky





V0F 2016 IASE 1 19

Air Pollution Linked to 6.5 Million Deaths a Year, Study Says



Melissa Chan **Time** June 27, 2016

V0F 2016 IASE 1 20

Association vs. Causation Using Ordinary English

1. Study: 45,000 Uninsured *Die* a Year (CBS News)
2. 45,000 deaths *attributable to* uninsurance
3. 45,000 US deaths *associated with* lack of insurance
4. No health coverage *tied to* 45,000 deaths a year
5. Lack of insurance *linked to* 45,000 deaths
6. Study: 45,000 U.S. Deaths *From* Lack of Insurance
7. One death every 12 minutes *due to* no health insurance
8. 45,000 ... die *because of* lack of health insurance
9. Lack of Health Insurance *Kills* 45,000 a Year
10. Lack of Health Insurance *cause* 44,789 deaths
11. Lack of insurance *to blame for* almost 45,000 deaths

Observational Statistics More Influences

Randomization eliminates many types of influence.
Inference models eliminate many others.
Teaching random-sample statistics is simpler.

Observational statistics have a host of influences.
Teaching observational statistics is harder.

Students *need a structure* that groups these influences into three or four categories.

Observational Statistics are Influenced By:

Confounding: [See Part 2]

- what was – and was not – controlled for
- what kind of study was involved.

Assembly: [See Part 1]

- how the statistics were selected, collected, defined, grouped, summarized, compared and presented.

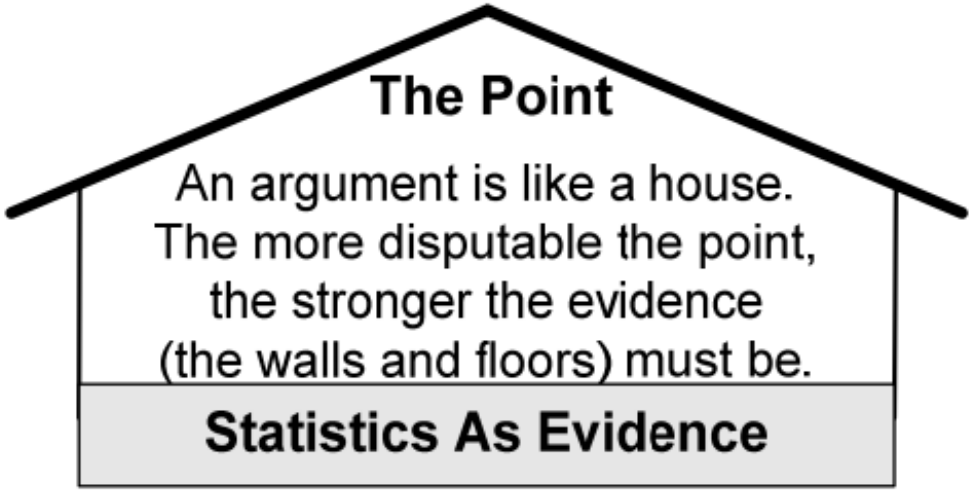
Randomness [See Part 3]

Error/bias

Statistical admonition: "Take **CARE**"

V0F 2016 IASE 1 23

Stats as Premise(Crit. Thinking)
Stats as Conclusion (Stat Literacy)



"All statistics are socially constructed
So, "Take CARE"!!
Statistics may be influenced by:

C	A	R	E
Context	Assembly	Randomness	Error

Assembly: Defining Groups

US children: Elevated levels of lead:

- 27,000 in 2009
- 259,000 in 2010 [Almost a factor of 10]

In 2010, the CDC reduced the minimum for elevated levels of lead from 10 to five*.

* micrograms per dl of blood
www.cdc.gov/nceh/lead/data/StateConfirmedByYear1997-2011.htm

V0F 2016 IASE 1 25

Small Changes in Syntax; Big Changes in Semantics

Assembly in Moore's Concepts & Controversies

Who is a smoker?

When estimating a proportion *p*, be sure you know what counts as a "success." The news says that 20% of adolescents smoke. Shocking. It turns out that this is the percentage who smoked at least once in the past month. If we say that a smoker is someone who smoked in at least 20 of the past 30 days and smoked at least half a pack on those days, fewer than 4% of adolescents qualify.

V0F 2016 IASE 1 27

Assembly in Presentation
7 nano grams/gram (7/Million)

Why Teachers Don't Want to Teach Assembly

1. Ordinary English is too ambiguous.
2. Leave this up to subject-matter experts.
3. This is not really "statistics"
4. Teaching it requires subject-matter expertise.

1. Ordinary English is how statistics are communicated
2. If we don't teach it, students will never see it.
3. We define what is really "statistics".
4. We can teach it without subject-matter expertise:

Which is bigger in a class: (1) # of students, (2) # of male students, or (3) # of students in or waiting?

V0F 2016 IASE 1 29

Contributions of Statistics to Human Knowledge

Statistics are numbers in a context
Association is not causation

RANDOMNESS and CAUSATION	CONFOUNDING and CAUSATION
Chance, independence and sampling distributions	Comparisons, ratios, models and study designs
Margin of error, hyp tests & statistical significance	Epidemiological causation (Bradford-Hill)
Random assignment and causation (Fisher: RCT)	*Confounder conditions for nullification (Cornfield)*

V0.7 Conditional probability, medical tests and Bayesian reasoning
Coincidence, Simpson's Paradox and regression to the mean

Most Important Topics/Ideas
Augsburg StatLit Students

1 Classify different kinds of influence (Take CARE)

2 Confounding

2 Hypothetical thinking: Plausible confounders, plausible definitions [Assembly].

4 Statistics are more than numbers [Assembly]

5 Association-causation & Randomness (Luck vs. skill)

5 Bias: Placebo, single blind; double blind

5 Named Ratio grammar; Percent, Percentages, Rates