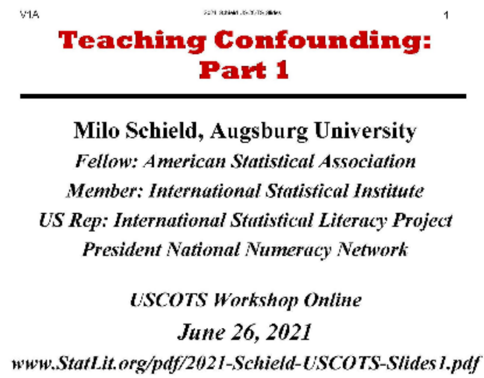
Hello. I am Milo Schield. I teach two introductory courses: traditional statistical inference and a confounder-based statistical literacy course. I've taught both for almost 25 years.



Slide 1.1 shows some of my background.
Here is the point of this workshop:

> *Confounding is the "elephant" in introductory statistics education. It is big. We know it is big in the everyday media. But we don't talk about it or teach it. I'm here to persuade you to teach confounding.*

First, you should learn more about confounding. You should devote one class to confounding in your statistical inference course. Then you should devote two. You should teach a separate confounder-based statistical literacy course as a topics course. Then you should add a confounder-based statistical literacy course to your catalog **as an alternative** to the statistical inference course and you should teach that course. Finally, you should help train others to teach this material.

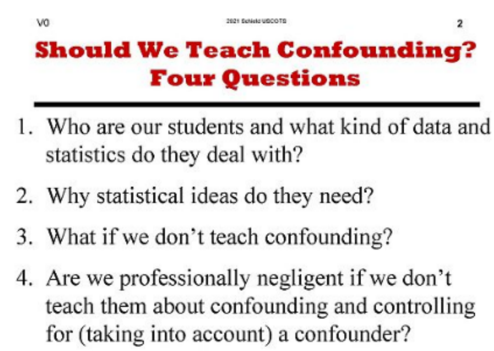
I'm sure you have other projects on your schedule. But I think teaching confounding is really important. Confounding is important to our students, and to the life of our society. Most of our students have never heard the word or know what it means. This is true before taking our class; it is true after taking our class. We know things about confounding that could really make a big difference to our students but we don't tell them anything. For me, this is professional negligence. I will revisit this point at the end of this workshop.

My goal for the next two hours is to talk about teaching confounding without using computer software. I have taught confounding for over two decades: trying things in the classroom. Here are my results.

The first part of this workshop is on confounding: its relationship to association and causation in our discipline and in my confounder-based statistical literacy course. The second part is on confounding: how do we adjust for (take into account) a confounder without using a computer. In the third part of this presentation, we will talk about what the University of New Mexico is doing: getting Statistical Literacy on the books, teacher training and the Cornfield conditions. Let's get started with the first part.

Slide 1.2: Should we teach confounding?
There are four things to consider:

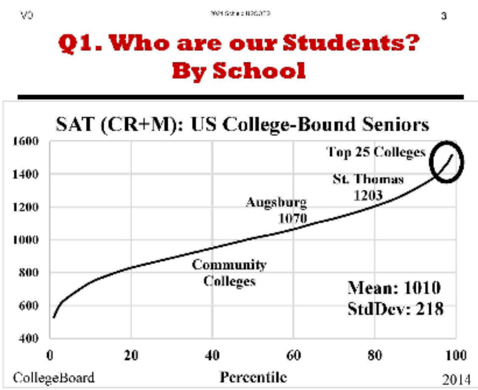
1. Who are our students and
2. What statistical ideas do our students need?
3. What will happen if we don't do teach confounding?
4. If we don't teach confounding are we professionally negligent?

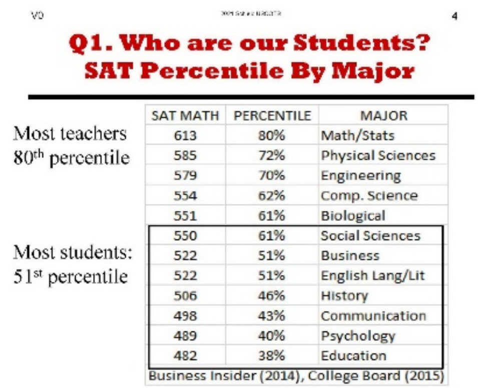


If we don't teach them about confounding, they don't know any more about what it means to "control for" something after taking our statistical inference course than they did before.

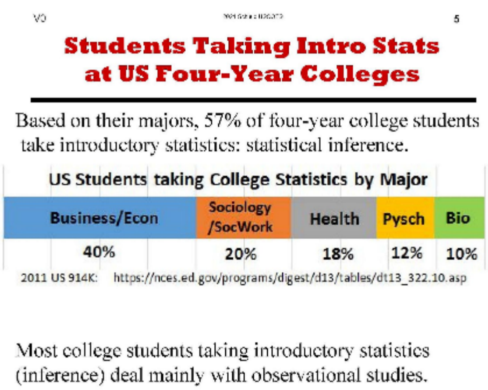
Slide 1.3 (left below). So, with that in mind, let's look at our students. Here they as high school seniors by percentile on the SAT. Many of the teachers at USCOTS are at the top 25 colleges. These are the elite college; they have very good students. Those students can handle a higher level of concepts than those [colleges] that have more intermediate students. Like those at St Thomas, and the University of New Mexico. I teach at Augsburg. Our average student is down close to the 60th percentile.

Statistical educators at the top schools are trying to set policy for what should be done at the low-end four-year colleges, community colleges, junior colleges, and even high schools. We need to think broadly about our students. We need to think about where they are by major.





Slide 1.4 (right above): Look at the college majors by SAT Math percentiles. What are the teachers? Mainly from majors near the 80th percentile: mathematics, the physical sciences, engineering, and computer science. Where are our students? They are in majors around the 50th percentile: business, the social sciences or psychology. The 80th percentile teachers who love math may not be the best for our 50th percentile students.



Slide 1.5: There's never been a poll of how many students take intro stats. But you can back into it by looking at what majors require introductory statistics.[1] On that basis, almost 60% of four-year college students take some form of introductory statistics.

Now let's look at those [students] by major. So this is 100% of all the students taking college statistics: 40% in business or economics, 20% in sociology or social work, 20% health and 10% psychology. All of these students will be influenced by the thinking that we do in USCOTS.

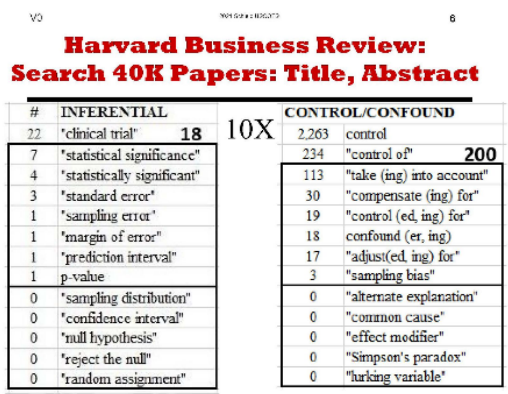
Look at what these majors are dealing with. Are they going to be dealing with clinical trials? Psychology is going to deal more with clinical trials than probably any other major in this group. Maybe health would be second; sociology and business are probably going to be last. Sociology, by its name, deals with social groups. You can't often run clinical trials on large social groups.

[1] Schield (2016). Stat 102: Social Statistics for Decision Makers. www.StatLit.org/pdf/2016-Schield-IASE.pdf

Based on students' majors, we need to start focusing on observational studies. This is really important,

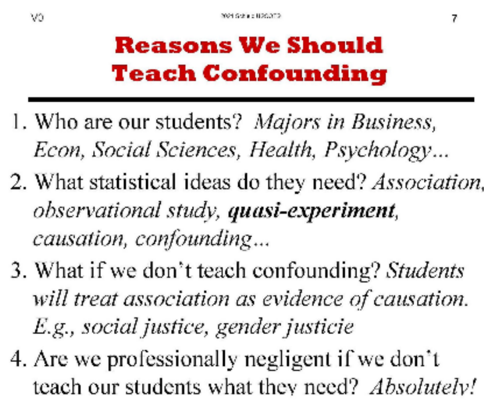
Slide 1.6 (left below): I teach business majors. I went to the Harvard Business cases and I looked to see how often inferential words appeared. Now this [search] is just in the title in the abstract, so these words may be in the body. But, as you can see, I had maybe 18 words and a lot of the good ones didn't appear: like random assignment or null hypothesis. P-value appears just once while 'statistically significant' and 'statistical significance' appear 11 times.

Whereas if I look at the words that I would connect with confounding, it's a different matter. Look at these words: 'taking into account', 'controlling for', 'confounding' and 'adjusting'. You get maybe 200 of those versus 20, on the inference side. Or if you include 'control of' and 'clinical trial', it's roughly 10 times as many words involving confounding as inference. This is very crude but it's data on what our students need.


V0 6

Harvard Business Review: Search 40K Papers: Title, Abstract

#	INFERENTIAL				CONTROL/CONFOUND	
22	'clinical trial"	18	10X	2,263	control	
7	'statistical significance"			234	"control of"	200
4	'statistically significant"			113	"take (ing) into account"	
3	'standard error'			30	"compensate (ing) for"	
1	'sampling error'			19	"control (ed, ing) for"	
1	'margin of error"			18	confound (er, ing)	
1	'prediction interval"			17	"adjust(ed, ing) for"	
1	p-value			3	"sampling bias"	
0	'sampling distribution"			0	"alternate explanation"	
0	'confidence interval"			0	"common cause"	
0	'null hypothesis"			0	"effect modifier"	
0	'reject the null"			0	"Simpson's paradox"	
0	'random assignment"			0	"lurking variable'	

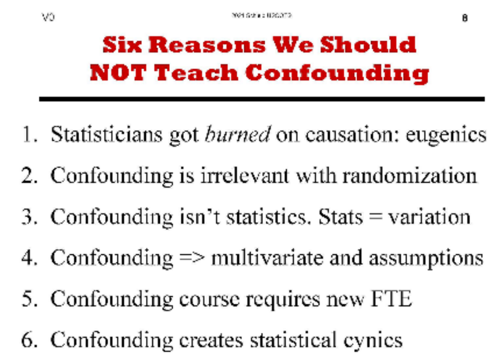


Slide 1.7 (above right): Let's summarize why we should teach confounding.
Q1. So who are our student? Business, Econ sociology health and psychology. You might not teach any of these because you're running a math or a statistics program. But in terms of the people that are being influenced by our thinking as statistical educators, here's where they are.

Q2. What ideas do they need? Association across the board; observational study is big. Normally we tend to just toss that to one side and say, you know that isn't really good quality stuff. Quasi experiment (which usually doesn't even come up), I would argue, is the most common form of study in business and in politics. The Covid19 lockdown has been a massive quasi experiment on a state by state basis. The states have control of something, but not a lot.

Q3. What happens if we don't teach confounding? What happens if you tell kids "Don't cross the street in the middle of the block"? They are going to do it anyway. What you should do is to say, "Look both ways before you cross the street in the middle of the block." The same thing happens when we don't talk about confounding. College graduates are going to do the same thing. They are going to treat association as evidence of causation and ignore confounding. I will come back and talk about that later because that's going to be central to my argument that teaching confounding is our job.

Q4. Are we doing professionally negligent if we don't teach our students what they need? I think there can only be one answer: yes! Absolutely! Doing this in the first semester is essential. Most of our students will only see us for one course (they are not coming back for a second course).



Slide 1.8: Now look at reasons NOT to teach statistical literacy. .

#1. Statistics as a discipline got burned on causation when our leaders supported the Eugenics movement.

#2. Confounding is irrelevant with randomization. That's what randomization does. I shouldn't say irrelevant it's statistically minimized.

#3. If we teach statistical literacy does statistics retain its core? This is what David Moore asked in his 1997 MSMESB talk in in Iowa City: "Is statistical literacy (lurking variable) really statistics?"[2] Of course, that depends on how you define statistics. We generally agree that statistics studies variation. But normally when we say variation we head right down the random side, and are into our intro inference course. We almost ignore the systematic side of variation. Confounding involves the systematic side of variation. So, teaching statistical literacy and confounding is definitely a part of statistics.

#4. If we are going to talk about confounding, we have to be multivariate. If we are going to introduce multivariate regression, there are assumptions. But, we can't deal with all the assumptions and diagnostics. And yet we don't want to be involved in anything that is any way unprofessional or less than professional.

#5. We are under pressure with cutting FTE or we're going to have to get new FTE for this course.

#6. We don't want to create statistical cynics.

I want to talk about each of these objections. I have written over 70 paper on why we should teach confounding.[3] This time, I want to focus on the reasons for **not** teaching confounding.

#1: Early statisticians supported Eugenics.

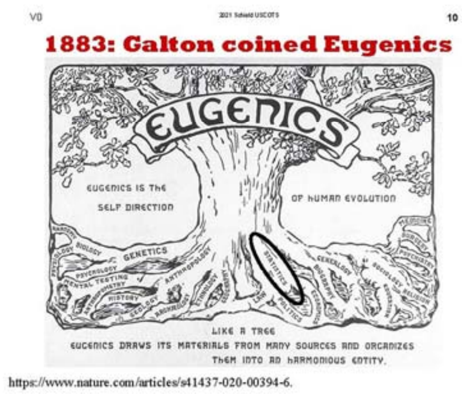
Slide 1.9: Let's start with the history of the Royal Statistical Society in 1834. Look that the two RSS logos. The left logo was the initial one. The right logo was introduced about 40 years later and is the current logo. What is the difference between the two? The little banner with a phrase in Latin. One interpretation is "let it be threshed out by others." We statisticians will gather and analyze the wheat (the data). But others should make the decisions. So why was that banner removed? Notice that 40 years after 1834 takes you into the 1880s.

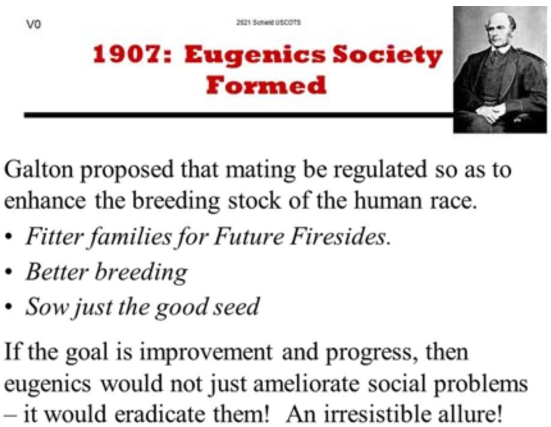
---

[2] Moore (1997). Statistical Literacy and Statistical Competence in the 21st Century. Copy of his slides at www.statlit.org/pdf/1997MooreASAslides.pdf

[3] Schield publications by topic at www.StatLit.org/Schield-Pubs.htm

Slide 1.10 (left below): In 1883, Charles Galton coined the word 'Eugenics'. Notice the tree of Eugenics; look at all the roots that are involved. Statistics is the most prominent of the roots. Yes, statistics and statisticians were involved in Eugenics big time.

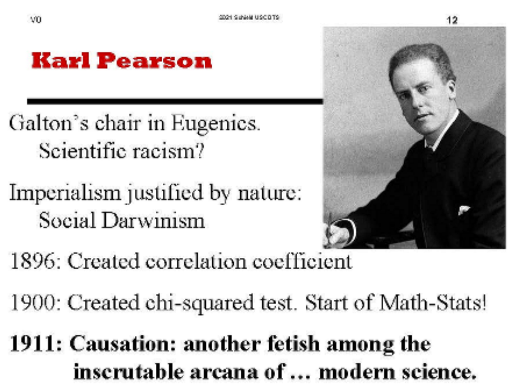




Slide 1.11 (right above): Eugenics was "fitter families, better breeding, sow just the good seed". The goal was improvement and progress. Eugenics wasn't just going to ameliorate social problems; it would eradicate them. Eugenics had an almost irresistible allure at the time.

Slide 1.12 (left below): Karl Pearson was a social Darwinist. He felt that races, are going to compete and that imperialism was justified by the nature of the white race. It was called scientific racism.

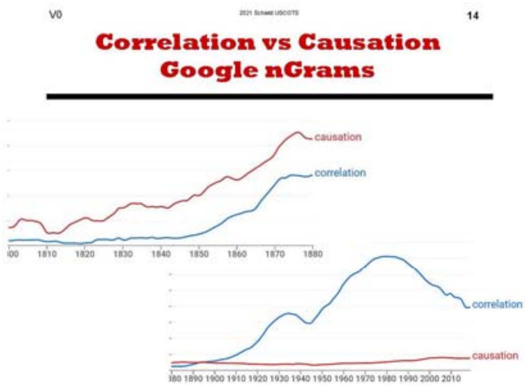
In 1896 Pearson created the Pearson correlation coefficient. In 1900 he created the Chi square test: the start of mathematical statistics. In 1911 he described causation as "a fetish among the inscrutable arcana of … modern science". That is a big change from supporting Eugenics.

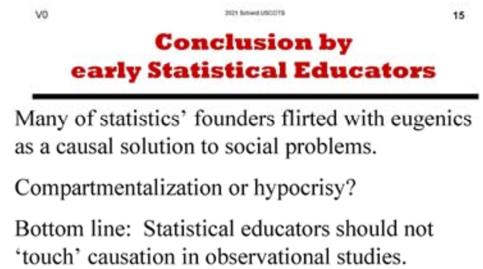


Slide 1.13 (right above): Ronald Fisher was also involved in Eugenics. At age 17 he attended a conference on Eugenics. He wrote a paper a few years later titled, "Some hopes of a Eugenist".

In 1935 he introduced the design of experiments, the null hypothesis, and random assignment. This was an incredible piece of work in history of our discipline. Yet, in 1938 he's talking about paying mothers for A-1 (healthy) babies.

Slide 1.14 (left below): Look at the prevalence of causation and correlation over time. Up until 1880, causation has always been the more popular according to Google Ngrams. But after 1900, correlation is massively more common than causation. --- Although, it has been dropping since it peaked in 1985.

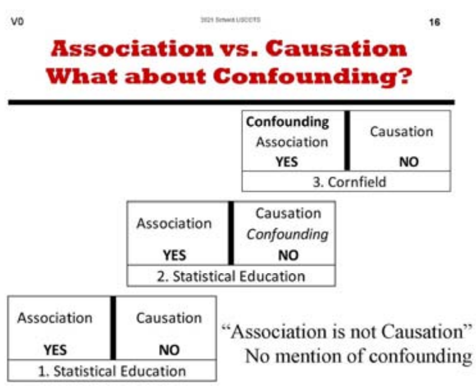


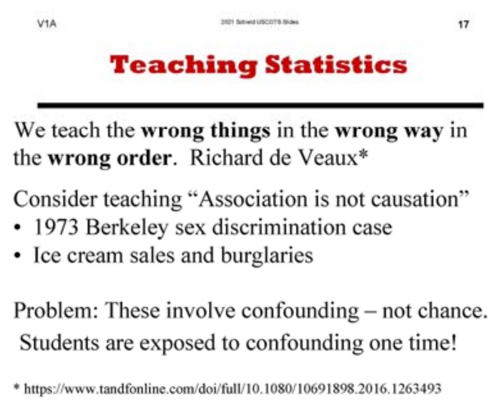


Slide 1.15 (right above): Many of the statistical founders flirted with Eugenics. Actually, they may have done more than flirt. Did they do this with hypocrisy or with compartmentalization: the left hand worked with math stats; the right hand worked with Eugenics? I don't have to go there. All I have to know is that our discipline decided that statistical educators should not touch causation in observational studies. Every time we say "association is not causation" I think we agree with that conclusion.

Slide 1.16 (left below): How do we as statistical educators handle confounding?

1. Ignore confounding entirely.
2. Put confounding alongside causation. Don't talk about either. I did a non-random sample of 100 of my intro textbooks. Less than 20 had confounding in the index. In most cases, it was the one line the entry for the example we used to distinguish association from causation. Examples include the Berkeley sex discrimination or ice cream sales and burglaries. The word confounding might appear for a moment, and then it would disappear. It would never appear in the textbook again. That's why I call confounding "the elephant in the room." We are all aware of it; we all know that it is there but it's not there for our students.
3. Put confounding next to association; on the opposite side of causation. This is how Jerome Cornfield handled confounding in the 1950s. [I will say this again and again. I think Jerome Cornfield is the most under-appreciated pioneer in statistics. He is crucial in understanding how to deal with confounding.]





Slide 1.17 (right above): Richard de Veaux made a very pithy comment, "We teach the wrong things in the wrong way in the wrong order."[4] I'm not sure that de Veaux ever told us the right things.

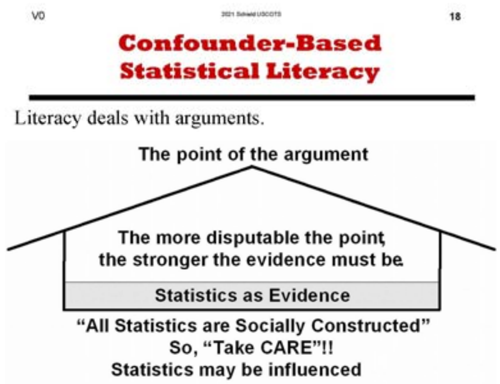
---

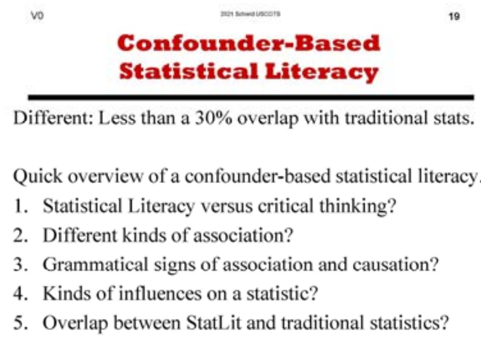
[4] www.statlit.org/pdf/2015-DeVeaux-USCOTS-Opening-Slides.pdf

I am more specific. I say we are teaching the wrong things by not teaching confounding.

So how do I teach confounding?

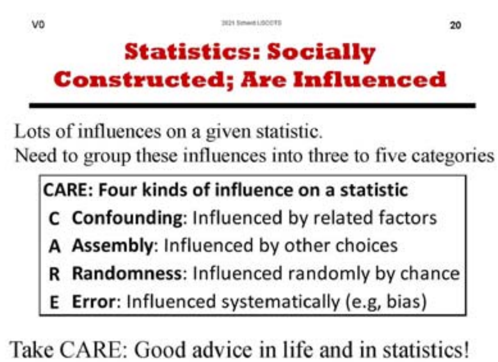
Slide 1.18 (left below): Statistics are primarily used as evidence in arguments. Statistics are typically in the basement of this model house. Typically, the statistics (the foundations) support the walls and the roof and they support the point of an argument. The best advice I can give to anybody dealing with statistics is this, "take care." Why? Social statistics are socially constructed. Statistics can be influenced.

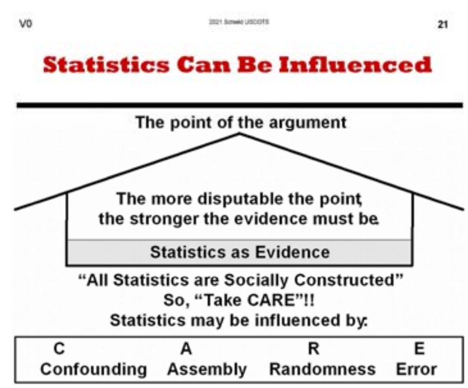




Slide 1.19 (right above). I say there is less than a 30% overlap between a confounder based statistical literacy course and a traditional inference-based statistics course. I want to go through some of the features.

Slide 1.20 (left below): There are a lot of influences on a statistic. My mind just doesn't hold much more than three, four or five things at one time. I have grouped these varied statistical influences into four categories. The four letters in CARE signify the four kinds of influence on a statistic. C stands for confounding. A stands for assembly or assumptions. R stands for randomness. E stands for error. Telling our students to "Take CARE" is good advice in statistics and in life.





Slide 1.21 (right above): Here is how I look at argument with statistics in the basement where they are influenced by the four areas. How do we untangle causation from statistics? This figure connects them. Isn't that just what we said we didn't want to do?

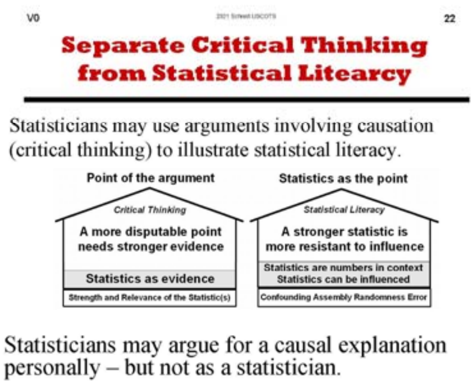
Slide 1.22 (left below): Here's how I separated statistics from causation. I think it is really important to have a strong bright line identifying what we do as statisticians and what we don't. On the left we've got critical thinking or decision making. Should we have a shutdown? Are the shutdowns effective? We are using statistics as evidence; we want to see how relevant they are to the point of the argument. I am saying the statisticians have no special expertise on that side.

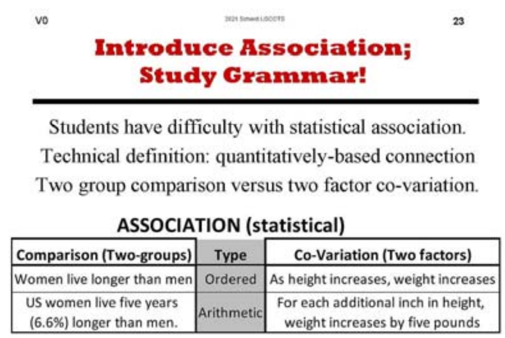
Statisticians have expertise on the argument on the right. On the right side, the statistic is the point. Some statistics are more resistant to influence than others. We want to look at how resistant the statistics are to Confounding, to Assembly, to Randomness and to Error.

I think that distinction between decision making and statistical literacy is important. Yes, I will go to the left side and show how changes on the right side change things on the left side.

But as statisticians, we need to say we handle the influences on statistics. We have no expertise on how strongly they support or don't support a particular decision on what we should or should not do.

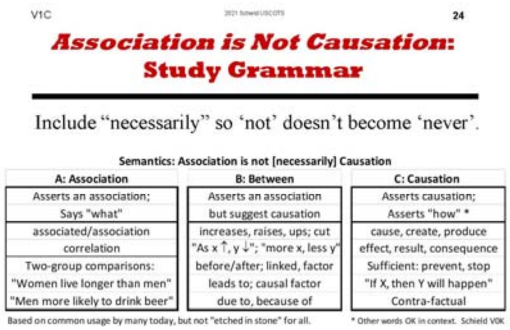
We may speak outside our area of professional competence, but we should tell others when we do so.





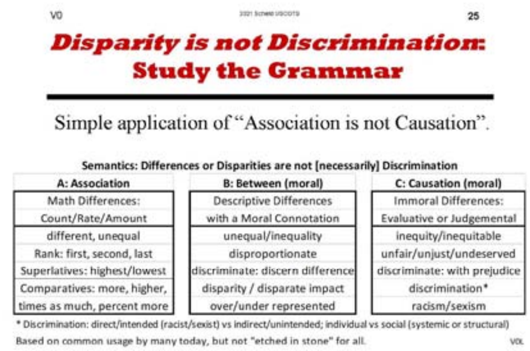
Slide 1.23 (right above): Association is a core idea in statistics. Students have difficulty with Association because we're using it to include comparisons and co variation. Students are not used to thinking that way, so it takes some time.

Slide 1.24 (left below): The distinction between association and causation is illustrated in terms of words. That is because that's how students typically are going to use distinguish these two groups. The association words (the A words) are on the left. The causation words (the C words) are on the right. In between are what I call the Between words (the B words). The Between words are the words the journalists love. For example, "eating nuts cuts cancer. *Cuts* and *ups* are two very short, very popular action verbs. As action verbs, they have the connotation of causation but they don't assert causation. *Before* and *after*, *leads to due to* and *because of* are related phrases. Between words are common. Students need to know that these do not assert causation.



A: Association	B: Between	C: Causation
Asserts an association;	Asserts an association	Asserts causation;
Says "what"	but suggest causation	Asserts "how" *
associated/association	increases, raises, ups; cut	cause, create, produce
correlation	"As x ↑, y ↓"; "more x, less y"	effect, result, consequence
Two-group comparisons:	before/after; linked, factor	Sufficient: prevent, stop
"Women live longer than men"	leads to; causal factor	"If X, then Y will happen"
"Men more likely to drink beer"	due to, because of	Contra-factual

Based on common usage by many today, but not "etched in stone" for all. * Other words OK in context. Schield V0K



A: Association	B: Between (moral)	C: Causation (moral)
Math Differences:	Descriptive Differences	Immoral Differences:
Count/Rate/Amount	with a Moral Connotation	Evaluative or Judgemental
different, unequal	unequal/inequality	inequity/inequitable
Rank: first, second, last	disproportionate	unfair/unjust/undeserved
Superlatives: highest/lowest	discriminate: discern difference	discriminate: with prejudice
Comparatives: more, higher,	disparity / disparate impact	discrimination*
times as much, percent more	over/under represented	racism/sexism

* Discrimination: direct/intended (racist/sexist) vs indirect/unintended; individual vs social (systemic or structural)
Based on common usage by many today, but not "etched in stone" for all. V0L

Slide 1.25 (right above): What about moral issues? As a statistician, I talk about inequality, disparity, disparate impact, disproportionate, and over/under represented. Occasionally, I introduce the old definition of 'to discriminate': to be a discriminating shopper is to discern the difference.

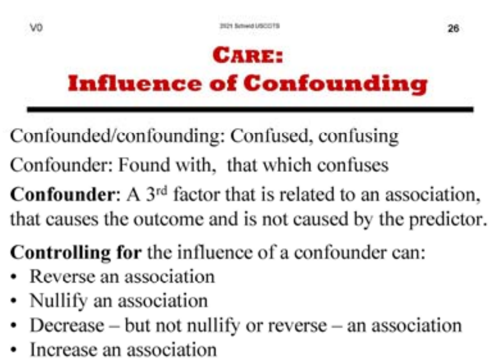
Normally, the things on the far right (inequity, unfair, discrimination) require an argument. However people sometimes reverse the order. If there is discrimination, then there is a disparity. If there is an inequity, then something is unequal. If something is inequitable, then there is an inequality. These relationships are generally true, but they reverse premise and conclusion.

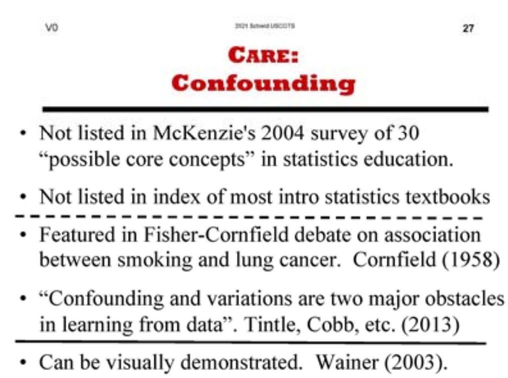
I generally note that there are many forms of discrimination: direct/intended vs. systemic or structural. However structural and systemic depend on social disparities as their evidence.

Slide 1.26 (left below): Students need to understand that controlling for a confounder can increase, decrease, nullify or reverse an association. They know this for counts: California has more unemployed people than Montana, but Montana may have a higher unemployment rate than California. Montana may have more unemployed people per worker than California.

Students have no idea of how they would control for (take into account) the influence of a confounder on a rate or percentage. Confounding wasn't listed in Mackenzie's list of important statistical terms. Confounding isn't listed in the index of most Introductory Statistics textbooks.

Technically, a confounder is any 3rd factor that is related to an association, that causes the outcome, and that is not caused by the predictor.

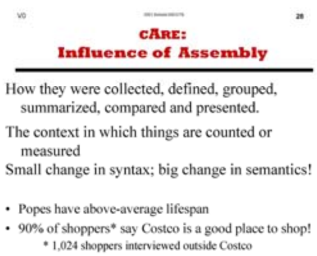


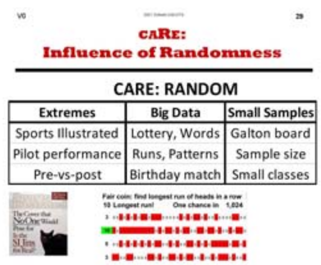


Slide 1.27 (right above). The C in 'Care' stands for confounding. **Confounders** are related variables that were not included in generating the association. Technically, a **confounder** is any third factor that is related to a predictor-outcome association, that causes the outcome, and that is not caused by the predictor.

Asa mentioned previously, controlling for (taking into account) the influence of a confounder can do any of four things to an association. It can reverse, nullify, decrease or increase an association.

Tintle, Rossman, Chance and Cobb[5] argued that confounding and variation are two major obstacles in analyzing data. I think confounding is critical. In 2003, Howard Wainer presented a simple graphical technique to control for (take into account) the influence of a binary confounder. I've used that method ever since. I will be showing that in the last half hour.





Slides 1.28 (above left), 1.29 (above center) and 1.30 (above right): I'm going to skip over these other three kinds of influence on a statistic or a statistical association for this workshop.

Are there any questions at this point? If not, I'll ask myself a question.

> Q. How long did it take you to go from teachings traditional statistics to teaching confounder-based statistical literacy? A. I would say 10 to 15 years and I came to statistics from an argument background: a critical thinking background.

Now it should take less to understand something that others have worked out, but it is going to take time. Think back to how long it took you to really get inside randomness (random selection and random assignment)? How long did it take for you to really understand the sampling distribution, confidence intervals or hypothesis testing?

And yet, I am trying to get our students to do it one semester. This is the same sort of thing. It is not quick; it is not easy; it takes time. Therefore, you really have to be motivated to take this on.

My purpose in giving this workshop is to motivate a few of you to stop and say, "You know, I thought I had projects that I wanted to do, but I think this is really important. I really want to try doing this."

I will be talking more on that transition because it's not an easy matter. I am currently working with two teachers at the University of New Mexico right now. We are finding out how difficult that is. There are a lot of concepts that you haven't run into before. For example, what is the difference between a confounder and a mechanism? As a math person I would rather talk about co-variation.

Ranjini Grove: Milo so I have a question, so I absolutely agree with you that this is something that gets overlooked and. I have actually been trying to do a lot more of this in my intro statistics class. We have

[5] Tintle et al. (2017). *Introduction to Statistical Investigations.*

a class at my university that does just statistical reasoning, so it seems like a good place to have a focus on this topic. However, what I have found is that for many of our students, for whom English is a second language, these words and these distinctions just confuse the heck out of them. They are ending up feeling frustrated that I even started talking about these things.

Milo Schield: Language is very important. Augsburg has a 60% minority population – mainly Somali. Larry Lesser at UT El Paso has been using some of my stuff because many – if not most – of his students are ESL (English as a second language).

Students have difficulty using these ideas in their writing using ordinary English. Let me give you one example. Consider "the percentage of women who are runners". Can you make this into a pie with a slice? Now suppose I change just one word. Consider "the percentage of women among runners". Is this the same pie with the same slice? No! This is just one example of how a small change in syntax (ordinary English) can make a big change in semantics (meaning).

Have you ever seen that written anywhere? I spent six months in England looking through used bookstores. I thought I'd find a book that would show me this simple example. I never found that book.

My textbook has more on ordinary English than I ever dreamed it would have. But for me, that's the job: teaching our students how to communicate using ordinary English. I've found that some native English speakers have weak English skills. Yes, language is critical.

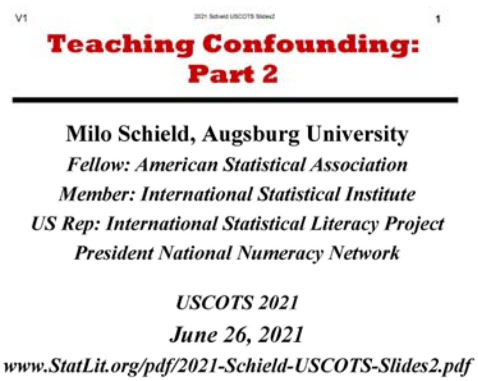
It takes my students probably half a semester to really get onboard with what a confounder is. This course is front loaded.

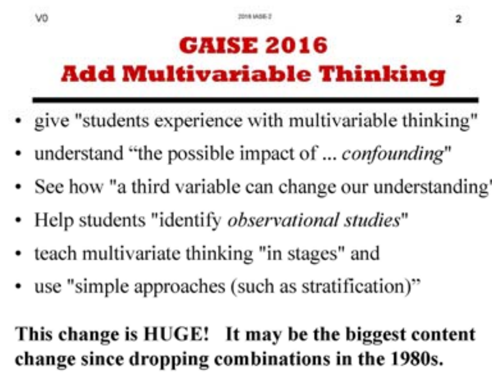
Slide 2.1(left below): Teaching confounding: Part 2. My name is Milo Schield. I have taught traditional statistics for over 30 years. I have taught confounder-based statistical literacy for over two decades. Let me repeat the point of this workshop:

Confounding is the "elephant" in introductory statistics education. It is big. We know it is big in the everyday media. But we don't talk about it or teach it. My goal is to persuade you to teach confounding.

Confounding may be something you may have no intention of doing, something you may not be interested in doing, something you were not trained to do, and something you may not enjoy doing at first.

I say this: "Teaching confounding is our job! It's what our students need". We need to offer it as a separate statistical literacy course: an alternative to traditional statistical inference. Having taught confounding for over 20 years, I have some ideas on what works with a wide variety of students

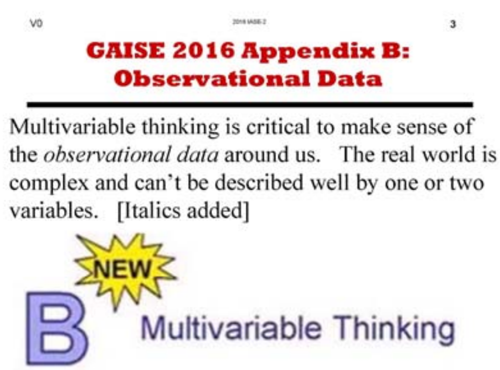


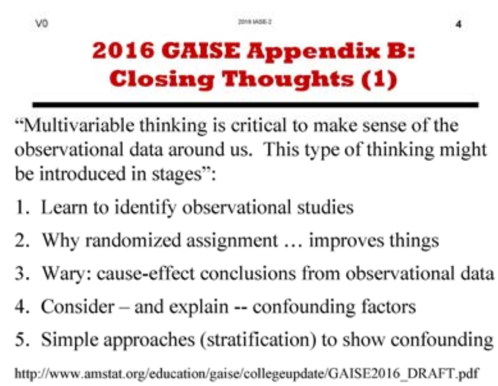


Slide 2.2 (right above): The 2016 update to the GAISE guidelines is the biggest content change in statistics education since we dropped combinations and permutations in the 1980s.

Slide 2.3 (left below): The 2016 update talks about observational studies and even mentions confounding: an entire appendix. Although multivariate thinking made it to the highest level, confounding did not.

Multivariate analysis is the door that allows you to talk about confounding. You cannot talk about a third variable if you're just dealing with bivariate data. You need multivariate.





Slide 2.4 (right above): Notice that they said, "Multivariable thinking is critical to make sense of the *observational* data around us." Normally they leave out the word 'observational' and say "to make sense of the data around us".

The guidelines are specifically mentioning 'observational data.' That's most of what students encounter in the everyday media. Even if you're reading JAMA, most of the articles are on observational studies (like the Harvard Nurses Study) – not clinical trials.

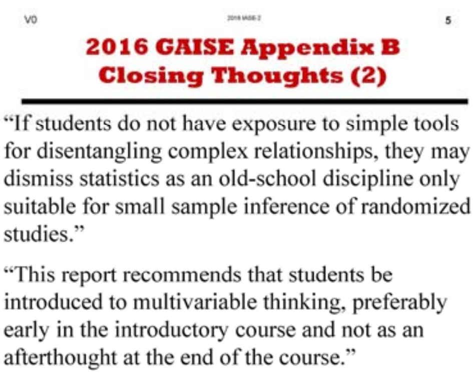
#4: Not just consider, but 'explain' confounding factors. Explaining is much more difficult than just showing. It is easy to consider or show examples of Simpson's paradox. But how many statisticians can explain it to someone without a mathematical background?

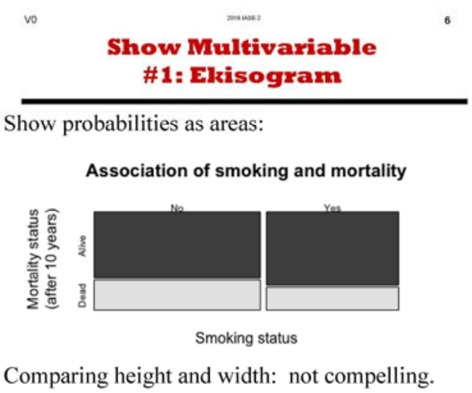
#5: Use simple approaches (stratification) to show confounding.
Simple approach like stratification can show the influence of confounding. But stratification does not help students understand what it means "to control for" (to take into account) confounding. That involves ratios when comparing counts, and standardization when comparing ratios.

Nevertheless, including multivariate thinking and observational studies is a huge improvement!

Slide 2.5 (left below). "If students don't have exposure to simple tools they may dismiss the small sample statistics as old school. Introduce multivariable thinking early in the introductory course."

Is this doable an intro course? In his 2007 USCOTS plenary address, Rossman said, "You simply can't achieve these [**GAISE** statistical literacy] goals in one course if you also teach a long list of methods."[6]

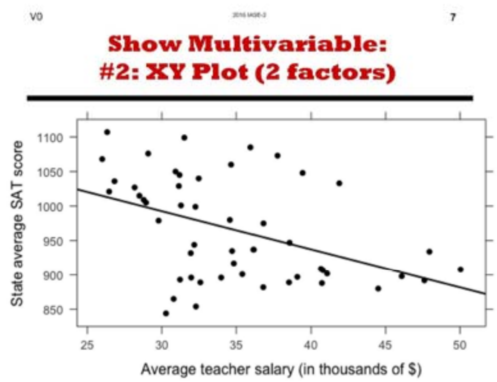


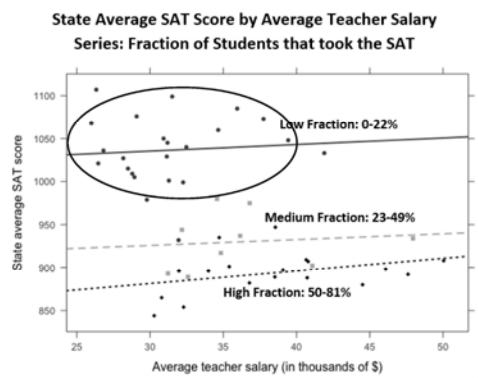


Slide 2.6 (right above): The 2016 GAISE update showed three ways of presenting confounding. The first was these Ekisograms. Comparing areas (height and width) is tough. For me, that approach is not compelling. I don't see how students can work any problems using that approach.

Slide 2.7 (left below): A second approach involves regression. This can be helpful depending on what your confounder is. If the confounder is quantitative, it is hard to see the relationships on a two-dimensional page. If it is categorical with more than three groups and if the relationships involved are complicated, then the confounding is doubly confusing.

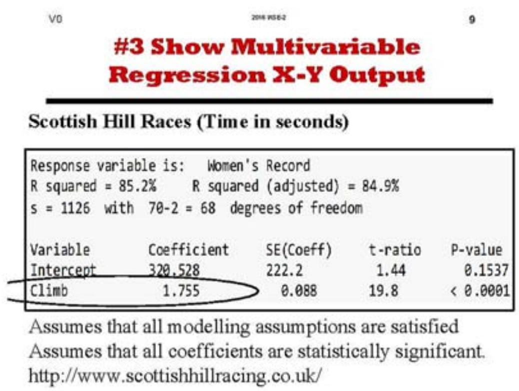
[6] Rossman (2007). Seven Challenges for the Undergraduate Statistics Curriculum in 2007. Slides at http://www.statlit.org/pdf/2007RossmanUSCOTS6up.pdf

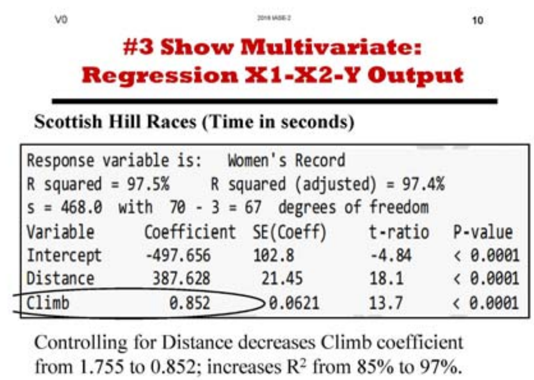




Slide 2.8 (right above): The percentage of each state's students that took the SAT exam are classified into three categories: a low fraction, a medium fraction and a high fraction. To understand this confounder, you have to know the geography: which states have low, medium and high SAT participation rates and which states have low, medium and high teacher salaries. This is a very complex confounder. I would never try and use this as a beginning example. If the confounder is binary (e.g., males and females), then it is easier for students to understand.

Slides 2.9(left below) shows bivariate regression with a Climb coefficient of 1.76



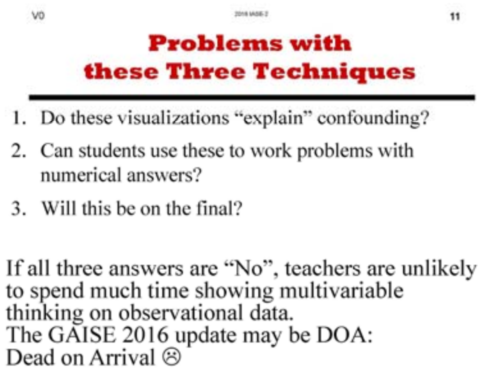


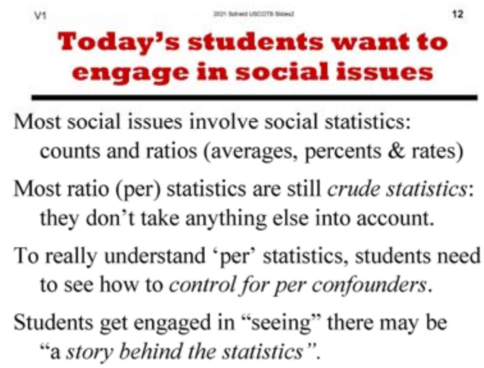
Slide 2.10 (right above) shows multivariate regression with two predictors and a Climb coefficient of .85. This clearly shows how controlling for (taking into account) a related factor can change the size of an association. But there are two problems. (1) Are we going to present the assumptions and diagnostics for multivariate regression; are we going to discuss whether these associations are statistically significant? If we don't, is this professional negligence? (2) Multivariate regression is a black box: it shows that controlling for a related factor can change an association, but it doesn't show how.

Slide 2.11 (left below): So how are we going to present confounding? We need visualizations that explain confounding. We need simple techniques that students can use to work, problems with numeric answers that don't require a computer: problems that can be asked on the final exam. If it's not on the final you can talk forever and the students say, "Whatever…" and don't pay attention.

I am saying that these three techniques are not adequate. If so, teachers are unlikely to spend time showing multi variable thinking on observational data. If I'm right, then that part of the GAISE 2016 update may be DOA: dead on arrival. Even without this, teaching multivariate may be too big of a change. We don't have time for it in the intro inference course.

I am proposing an even bigger change. We need a separate course: a confounder-based course. To see why, let regroup. Let's start with our students because our thinking should be based on that reality.





Slide 2.12 (right above):
Line 1: Today's students want to talk about social issues. Most social issues involve social statistics. These social statistics are generally of two kinds: counts or totals, and ratios.

Students know that a comparison of counts is oftentimes a crude association. There are more unemployed people in California, then in North Dakota. Duuh. To control for the size of the workforces, we need to create a count-per-worker. I call this a 'per' statistic. It converts a comparison of counts into a comparison of rates. Students know all about that.

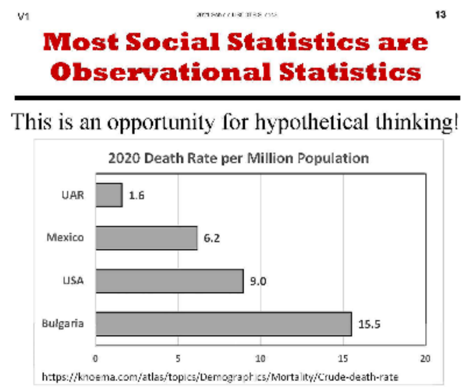
Line 2: What students don't know is that per-statistics (ratios, averages, percents and rates) can still be a crude statistic. An association of per-statistics can still be a crude association; it doesn't take anything else into account; it can still be confounded.

Line 3: To really understand these ratio or per-statistics, students need to know how to take into account control for these 'per' confounders.

Line 4: But once they realize there may be a story behind what was just a straight up statistic, their eyes open! They've never seen anything like this in any math course they've ever had. That's because this is not presented in any algebra or calculus course. They won't see it until they take differential equations. The total derivative can have a different value and sign from a partial derivative. Students come alive when they see that there may be a "story behind the statistics".

Slide 2.13: Supposing a student walks up to you, and shows you the 2020 death rates per million population for three countries: the US at nine, Mexico at six and the United Arab Republic (UAR) at less than two. If the death rate is one measure of health care quality, then don't these statistics support the idea that Mexico has better health care than the US (and that UAR has better health care than both)?

You may be thinking, "That's why I like teaching statistical inference. You never get this kind of a problem. This problem requires subject matter expertise." But you recall defining statistics as numbers in context. These are certainly numbers in context!



Here is where you have to talk to your students about hypothetical thinking. Hypothetical thinking is an important part of this process. So you say to your students, "Let's think about it together. What could cause death?"

Lifespan is a mathematical explanation. The death rate per year is roughly the inverse of the lifespan in years. If people live 100 years on average, and then they die, we would expect 10 deaths per thousand population every year. If people die at age 50, then we would expect 20 annual deaths per million. But if these death rates reflect lifespan, that means Mexico has a longer lifespan than the US and the UAR has a longer lifespan than both. If the US lifespan is 80, then the lifespan in the UAR should be at least five times 80: 400. That doesn't make sense. The explanation must be different than lifespan.

So you ask your students what else could influence death rates beside lifespan. Students know this subject; they are subject matter experts. They talk about exercise, diet, disease, availability of doctors, etc. That is hypothetical thinking. Finally, somebody will say, "Age." It could be the mixture of people by age. Old people are more likely to die than young people.

You say "Yes, the age mix is a plausible confounder. How can we get quick info on the age distribution? What about the median age?" That's an empirical question. Students ask Google. They find out the median age in the US is 39, Mexico is 29. The younger population in Mexico helps explains why Mexico has a lower death rate than the US.

What about the UAR? The median age in the UAR is 33. How can the median age be higher in the UAR than in Mexico, but the death rate is lower in the UAR than in Mexico? There will be times in teaching confounding, when you as must say, "I have no idea." Acknowledging ignorance may be appropriate

Hypothetical thinking is a big part of the course.

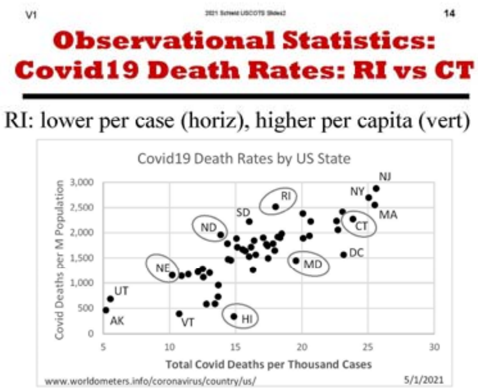
I don't put hypothetical thinking on the final I don't expect students to come up with the right answer on a final. But students need to be doing this in every class throughout the course.

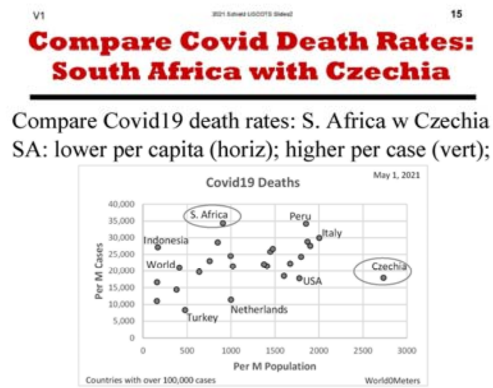
Slide 2.14 (left below): Let's talk about something simple: Covid death rates: rates based on total deaths to date. Which US state has the higher Covid death rate: Rhode Island or Connecticut? On the horizontal axis, Connecticut (CT) has a higher number of Covid deaths per thousand cases than Rhode Island (RI). Now, look at the vertical axis. We are still talking about the same deaths, but now they are per million population. Rhode Island has a higher total death rate than Connecticut.

What's going on here? This is what I call the diabolical denominator. The choice of denominator influences the direction of the association between two rates having the same numerator.

The simplest case is where two groups have the same numerator: number of deaths. The group with the smaller population will have the higher death rate *per person.* The group with the smaller number of cases will have the higher death rate *per case*. If the group with the higher population has the smaller number of cases, there will be a reversal. The group with the highest death rate *per capita* will have the smallest death rate *per case*.

This is certainly confusing; it certainly confounds; it is confounding. But does this involve a confounder? This is new territory for me. In the video, I said "Yes". As of this writing, I will set it aside.



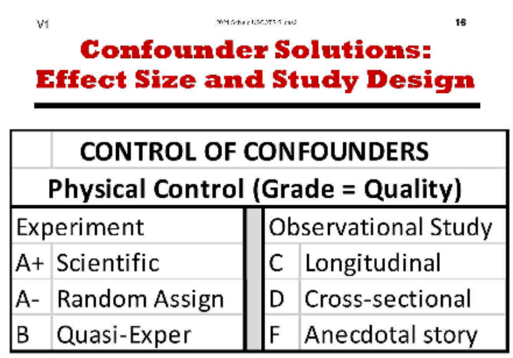


Slide 2.15 (right above): Compare the death rates for South Africa and Czechia -- what used to be the Czech Republic. On the horizontal axis, the death rate *per person* was higher in Czechia (152K/M population) than in South Africa (26.4K/M population). But on the vertical axis, the death rate *per case* is higher in South Africa (34K/M cases) than in Czechia (18K/M case). Once again the choice of the denominator changes the direction of the association between two rates.

What's going on here? The ratio of the two denominators varies by group. As of 5/1/2021, cases per million population were 53% higher in RI (145K) than in CT (95K). Cases per million population were 5.8 times as many in Czechia (152K) as in South Africa (26.4K).

What are the ways we have in dealing with confounders? The biggest one is effect size. That's number one. The factor of 10 effect size is what made the association between smoking and lung cancer death almost immune to the influence of any known confounder.

Number two is study design.
Slide 2.16: Study design physically takes control of certain kinds of confounders. Students need to be exposed to the full spectrum of study designs. In teaching inference, we focus on random selection and random assignment. Random selection controls for sampling bias, but it doesn't control for any confounder. Whereas random assignment statistically controls for any pre-existing confounders.



Experiments have various degrees of control of confounders; observational studies lack any control of confounders. Thus experiments have more control of confounders than do observational studies.

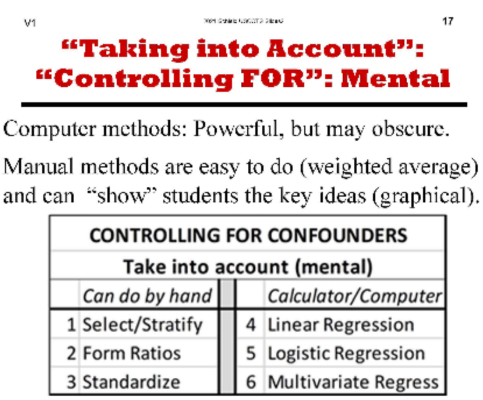
There are distinctions within experiments. I argue that quasi experiment are the most common experiment for students in business or in political science. In quasi experiments someone has control of something. Businesses control the price of their goods or services. City, states and countries implement policies for different regions or groups. The different lockdown strategies for different US states is a massive quasi-experiment.

On the observational side, cross-sectional studies are more common than longitudinal. One reason is that they take less time and money.

Note that each of these is given a letter grade. Students need some way to organize them hierarchically. Now these grades are just starting grades. As more is known about a study, its control of confounders may change. Consider the association between smoking and lung cancer. It started as cross-sectional study based on a two group comparison. But with more data, it progressed and became a cross-sectional study based on a two-factor co-variation: the amount a person smoked. In the end the large effect size (smokers are 10 times as likely to die from lung cancer as are non-smokers), warded off all known confounders and left this observational study as strong evidence.

The opposite can happen. Cold fusion was supposedly observed in a scientific experiment. Some confirmed it; others did not. It may be either randomness or error was responsible for the varied results.

Number 3 is controlling for confounders.

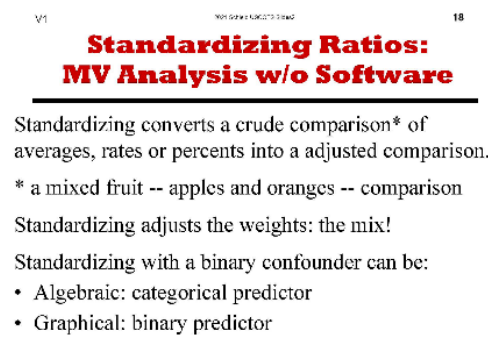


Slide 2.17: Effect size and study design are already 'baked into' the statistics that most students see. Students need to think about them, but they can't change them. Students can mentally control for (take into account) confounders. These mental methods of controlling for confounders have been divided into two groups: those that you can manually, and those requiring a calculator or computer. We are all familiar with the methods on the right side: linear regression, logistic regression and multivariate regression. But as I said before, these black boxes may actually obscure what it means to control for (take into account) the influence of a confounder. In trying to help students understand what is it means to control for (take into account) the influence of a confounder, the manual methods have some advantages.

Number one is selection or stratification. It is simple to do; it is simple to understand. As mentioned before, it shows the effect of confounding, but it doesn't show what it means to control for (take into account) the influence of a confounder.

Number two is to form ratios. College students are very familiar with forming ratios. Students recognize that a comparison of counts can be confounded by the size of the groups. They know that ratios take into account the size of related factors. So, a comparison is ratios may be more relevant than a comparison of counts. But they have no idea that a comparison of ratios can be confounded.

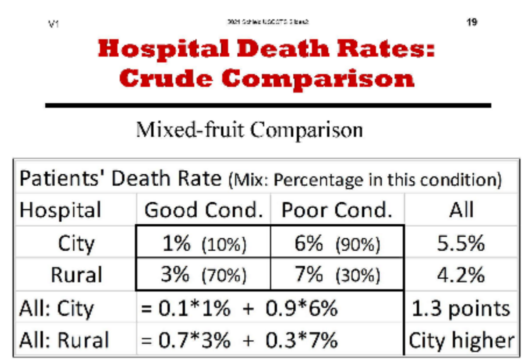
Slide 2.18: Number three is standardizing. Standardizing is a simple technique to control for (take into account) the influence of a confounder on a comparison of ratios. For many of us in statistical education this isn't something we've encountered. If you're into population demographics or into life insurance, you might have. But if you look at the history of our discipline in the 50s and 60s before they had faster computers, standardization was a popular technique. So, we are going back in time here.



What does standardizing do? Standardizing converts a crude comparison of averages, rates or percentages into an adjusted comparison. That still pretty abstract. I say, "a crude association of ratios can still be a 'mixed fruit comparison', an 'apples and oranges' comparison". Standardizing adjusts the weights to make it an 'apples and apples comparison.' "Standardizing adjusts the mix!" I like really short words, phrases and sentences. They help make things memorable. I'm hoping those words, phrases and sentences stay with a student after the classes over.

To keep things very simple, consider cases where the confounder and predictor are both binary. There are two approaches: arithmetic and graphical

Now, you may be thinking that I'm doing multivariate regression without any mention of assumptions or diagnostics. You may wonder if I'm leading you down some primrose path that's going to lead us into statistical negligence. But when you use regression with a binary predictor and binary confounder, it is a fully saturated model: the assumptions are automatically satisfied or irrelevant. You can check that out. [Even if not true, when this technique is applied to big data sets, everything is statistically-significant.]



Slide 2.19: First, the arithmetic technique. This involves seventh grade math: the weighted average. Consider the patient death rates at two hospitals. The percentages in the far right column show that the patient death rate is higher for City hospital (5.5%) than for Rural (4.2%).

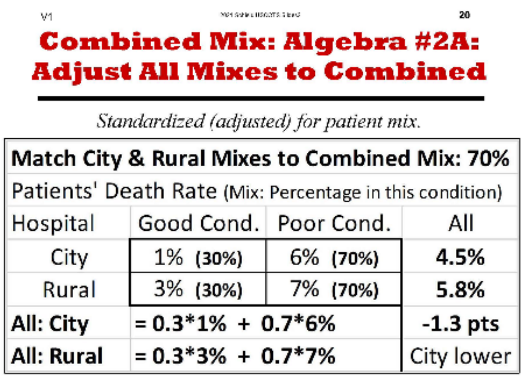
Suppose that you don't want to die, and you have to go to one of these hospitals which one do you choose? You may say "Rural". You may think that the city hospitals are more likely to have weird diseases; you can come out with something you didn't have when you walked in the door. Rural is more likely to have friendly nurses that don't wake you up in the middle of the night. As humans we are experts at creating stories that explain things – even if the stories aren't true.

Now look at the data by patient condition. Patients in good conditions are ones that walked in the door; patients in poor condition are ones that were carried in the door.

If you look at the patients in good condition (2nd column), the patients' death rate is lower at City (1%) than at Rural (3%). If you look at patients in poor condition (3rd column), the patients' death rate is lower at City (6%) than at Rural (7%). If you are in good condition you should chose City; if you are in poor condition you should chose City. Anyone can see the problem.

Statistical educators know this is Simpson's paradox. It's a difference in the mix of patient types. The weights (the mix) of patients is shown in the parenthesis. Of the City patients, 90% are in poor condition (only 30% for Rural). The math for calculating the weighted averages is shown in the bottom two lines.

The weighted average death rate for City is the fraction of patients that are in good condition (0.1) times the associated death rate (1%) plus the fraction of patients that are in poor condition (0.9) times the associated death rate (6%) which gives the 5.5%. The same procedure gives the 4.2% for the average patients' death rate at Rural.

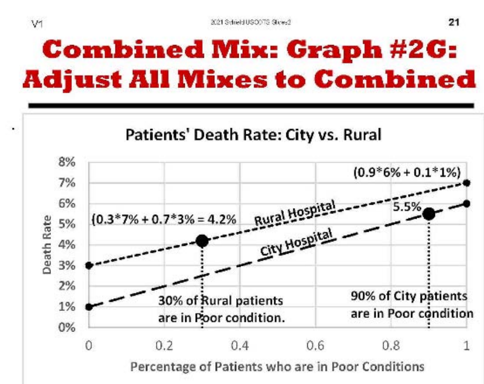


Match City & Rural Mixes to Combined Mix: 70%			
Patients' Death Rate (Mix: Percentage in this condition)			
Hospital	Good Cond.	Poor Cond.	All
City	1% (30%)	6% (70%)	4.5%
Rural	3% (30%)	7% (70%)	5.8%
All: City	= 0.3*1% + 0.7*6%		-1.3 pts
All: Rural	= 0.3*3% + 0.7*7%		City lower

Standardizing involves giving both groups the same mixture of a related factor: both hospitals the same mix of patients.

Slide 2.20: We want to standardize. Suppose we adjust on their combined mixture. One way is to combine the patients of both hospitals together and then apply the resulting mix to each hospital. . We are not changing the death rates for any of the four groups

Suppose that if we combine the patients at both hospitals, we find that 70% are in poor condition (30% in good condition). Now you recalculate the average rates for each. You get a city rate that's lower than the rural rate. This is Simpson's paradox. Students can do the math, they can work out the numbers. But I'm not certain that they really see what's going on.



Slide 2.21: To show students how standardizing works, I use a graphical technique. This graph is more complex; it takes more time to go over than I have in this workshop. Howard Wainer introduced this graph.[7]

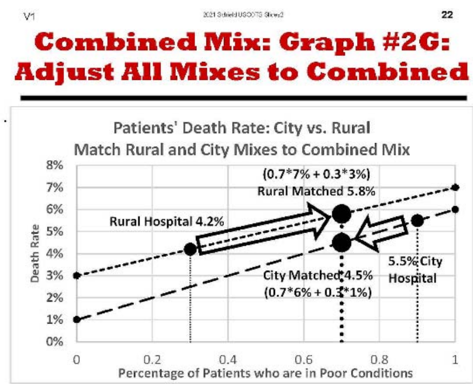
On the right side, we have the patients in poor condition. They have the high death rates in the upper-right corner: 6% for City and 7% for Rural. On the left side, we have the patients that are in not poor condition: in good condition. Remember, the confounder and the predictor must both be binary for this technique. Those in good condition have death rates of 1% at City and 3% at Rural. Draw diagonal

[7] Wainer, H. (2002). The BK-Plot: Making Simpsons' Paradox Clear to the Masses. *Chance* 2002 15(3).

lines between those for each hospital. These diagonal lines are the weighted average lines. The particular value of a weighted average is determined by the mixture: the percentage of each hospital's patients that are in poor condition.

City hospital had 90% of its patients in poor condition. Rural had 30%. That is how we got our initial values of 5.5% and 4.2%.

Slide 2.22: What happens if we standardize where 70% are in poor condition? Rural goes up to 5.8%; City goes down to 4.5%. The higher death rate at the Rural hospital was confounded by the low fraction of their patients in poor condition. Now the students can see Simpson's paradox as the mixture is changed. Peter Holmes said that seeing this graph was the first time he really 'saw' Simpson's paradox.



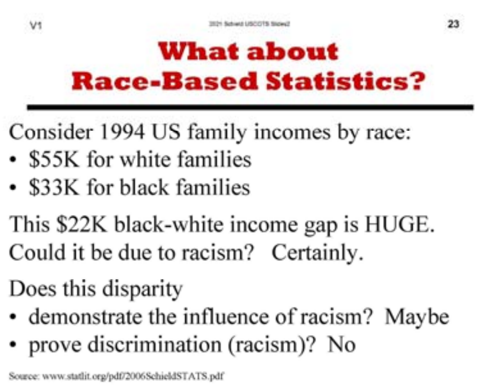
I have taught the graphical technique for over 15 years. I taught the arithmetic for the last three. I have taught the combo (both together) for the last year and a half. I am sold on teaching the combo.

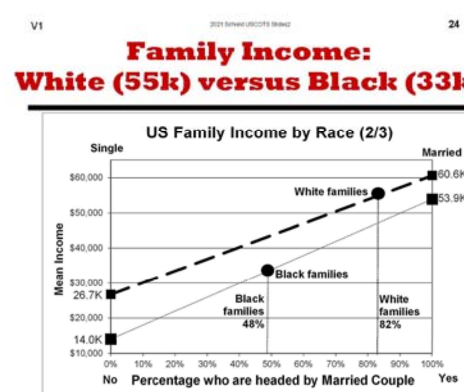
How does teaching confounding relate to understanding the story behind race-based disparities?

Slide 2.23 (lower left): Suppose that mean family incomes were $55 K for whites, $33 K for blacks. This $22 K gap is huge. Could it be due to racism? Certainly. Does this disparity show that racism is the cause? Maybe.

Does this disparity prove race-based discrimination? I say "No". Observationally-based statistics don't prove discrimination. Our job is to ask whether this crude association could they be influenced by related factors.

I ask students, "What could influence income?" They say education, your job or occupation, age, where live in the country, etc. Finally somebody will say, "You are talking about family incomes. Maybe it is family structure. Two adults can earn more than one." I say, "We just happen to have the data on that".





Slide 2.24 (upper-right): Consider the same graph. On the right, we have families headed by married couples with incomes of $61k for whites; 54k for blacks. On the left, we families that are headed by a single parent. I've been a single parent, so I know something about this group. Single parent families have average incomes of $27K for whites; 14K for blacks.

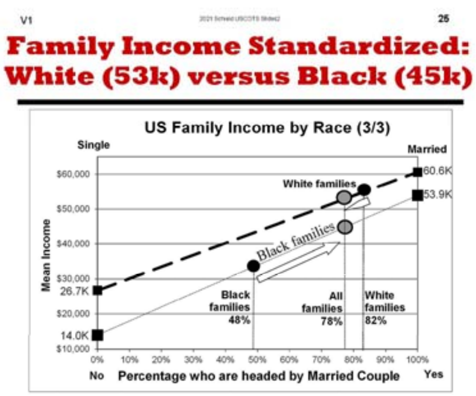
You see the diagonal lines weighted-average lines for white families and for black families. The average values depend on the mixture. We need the percentages of white families that are headed by a married couple (and that percentage for black families).

Suppose the percentage for white families is 82% (48% for black families). That gives us an average income of around $55K for white families, around $33K for black families: for a black-white income gap of around $22K. This is where we started.

This $22K black-white income disparity is a fact. It is not a false statistic; it is not fake news. But it is a crude statistic: a crude association of statistics. It doesn't take into account related factors like family structure. If we want to control for (take into account) family structure, we need to standardize.

Slide 2.25 (lower left): Suppose we put white and black families together and found that about 78% of all US families are headed by a married couple. After standardizing, white family income drops to around $53K; black family income increases to around $45K. After controlling for family structure, the black-white income gap is $8K.

The graphical approach doesn't have the accuracy of the arithmetic approach. In multiple choice problems, I give answers every $5K and ask them to pick the closest. They don't have to be real accurate (or they can do it arithmetically).



V1 26

Family Income Gap: "Explained by"

68% of black-white family income gap is *explained* by family structure

Controlling for Marital Status	Crude Before	Adjusted After	Change
Whites	55K	53K	-2K
Blacks	33K	45K	+12K
BW Income Gap	22K	8K	-15K
Percentage of gap explained: 15K/22K = 68%			

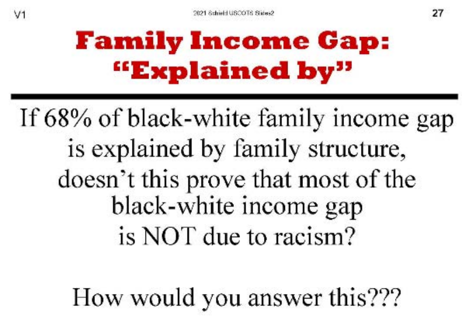
Slide 2.26 (upper-right): Here is a summary of the average incomes before and after the adjustment. The crude statistics ($55K and $33K) had a black-white income gap of $22K. The adjusted statistics ($53K and $45K) had a black-white income gap of $8K.

The $22K was what I call "journalistically significant". It makes a great headline. Whereas, $8K is not as journalistically significant. It's not likely to be featured in the headline. But, both numbers are true.

Look at the $15K reduction in the income gap. That $15K is 68% of the original $22K income gap. Of the original black-white income gap, 68% has been eliminated by controlling for (taking into account) family structure. By controlling for (taking into account) family structure, 68% of this black white income disparity is explained.

You might note that this percentage explained is not based on a reduction in variance. That is true. But I think both are important.

Slide 2.27: One of my Somali students asked me, "What does this say about race-based discrimination?" I presumed he was asking, "Does this prove that most of the black-white income gap is not based on race?"



I had not considered that question before. I felt little alarm bells going off inside my head. I was certain that observationally-based statistics never prove the presence or absence of causation or discrimination. So I said "No". But I felt he wasn't satisfied.

How would you answer this question if you were standing in front of a class? I'm guessing you would say, "Let me teach statistical inference. It's math! I can prove it! You never get this kind of question!

My point is this: it takes time to handle a statistical literacy course because it is a critical thinking course. But I am saying, "Teaching confounding – teaching confounder-based statistical literacy: that's our job!"

Back to my student. My answer was "No." This percentage reduction doesn't prove it." I felt that he wasn't satisfied. I decided I had to think of a way that racism could explain family structure.

So I said, "Suppose that our criminal justice system is racist. Suppose that black men are more likely to be sent to prison than white men for the same crime, same circumstances, same everything else.

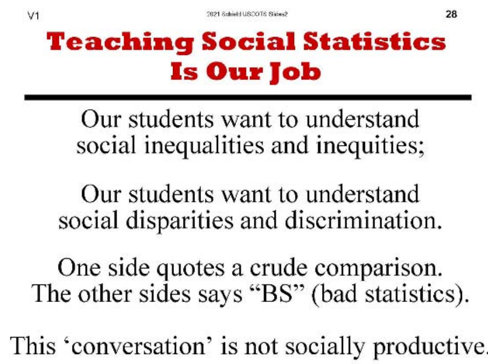
Black men in prison are not likely to get married. Black wives may be more likely to ask for a divorce if their husband is in prison. Race-based sentencing may help explain the disparity in family structure.

Just because we've taken one thing into account doesn't eliminate the possibility that the factor taken into account is confounded by something else. The bottom line is. You have to argue about these things. It's not just a simple matter of making a statement and saying, "Gotcha. This proves it."

We want our students to become critical thinkers. Right now, too many of them are what I would call naive thinkers. The statements they are making may be true statements. But they don't realize the statistical disparities are usually crude associations. Crude association can generally be influenced by something else.

Conclusion: Our students want to understand the difference between inequalities and inequities, between disparities and discrimination. And we want them to recognize that inequalities and disparities may involve crude associations: mixed fruit comparisons. We can – and should – help them think more clearly, more deeply and more productively about statistics as evidence in arguments.

Slide 2.29: A social justice warrior may quote a crude comparison. The other side may say, "That's 'BS' " which in our world stands for "bad statistics". But crude statistics are not bad statistics; they are not good statistics. Things are good and bad only in relation to some standard: to some value, end, purpose or function.

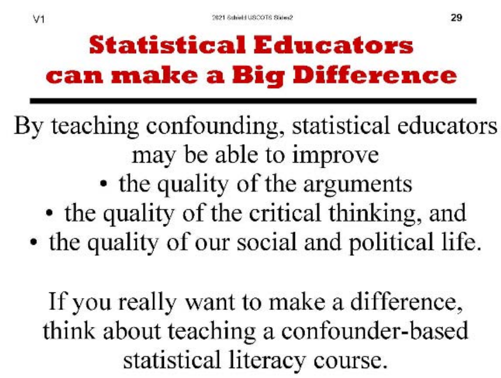


A crude comparison can be a true statement but it's still a crude comparison. Students need to understand that distinction. Students may understand that after you take something into account, it can change the comparison, it can change the amount, and it can change the direction.

Much of today's arguing is not socially productive. That is sad because statistical educators can bring something to the argument that might help both sides move forward. What can we bring? The idea of confounding. The understanding of what it means to control for something; what it means to take something into account.

Teaching confounding won't settle the dispute. But, it may help them sharpen their arguments, and be more socially productive in their conversation than they are now.

Slide 2.29: My bottom line is this. I want you to consider teaching confounder-based statistical literacy course *as an alternative* to a traditional statistical inference course. . By teaching confounding we may be able to improve the quality of the arguments, the quality of the critical thinking, and the quality of our social and political life.



This is massive! Aside from philosophy, there is no other discipline that could make the impact that statistics could. If you really want to make a difference, think about teaching a confounder based statistical literacy course. Yes, I'm in sales mode. But I believe that our students and our society need what we could – and should – offer.

Milo Schield
Let me repeat the point of this workshop:

Confounding is the "elephant" in introductory statistics education. It is big. We know it is big in the everyday media. But we don't talk about it or teach it. I'm here to persuade you to teach confounding.

Before I start the third part of this workshop, I want to make two recommendations.

First, I want to recommend a book by a friend of mine that was published earlier this year. My friend is Donald Macnaughton; the title of his book is *The War on Statistical Significance: The American Statistician vs. the New England Journal of Medicine.*"

Macnaughton's thesis is that journal editors need statistical significance as a threshold. On pages 62-63, Macnaughton gives three reasons. (1) The concepts (when used properly) are objective; science values objectivity. (2) If we abandon the concepts, then more false-positive errors will appear and the "replication crisis" will get worse. (3) Keeping the concepts enables journals to save time.

I strongly recommend this book to everyone who teaches or uses statistical inference. Macnaughton is a very thoughtful author. His reasoning is worth studying. I extend what Macnaughton said to include journalists and their readers. Just like journal editors, journalists and their readers don't have time to study the full context of an experiment or study. Statisticians must recognize that the phrase, "statistically significant" is no longer a phrase they can control. "Statistically significant" is too convenient; it is going to be used regardless of what statisticians do or say.

Second, I want to announce the publication of my Statistical Literacy textbook by Wiley scheduled for 2022. This textbook deals with a lot more than confounding. It is different. It will give you some ideas on how you might teach a confounder-based statistical literacy course.

Chat: Someone asked about defining confounding. In my textbook, I have a functional description and two traditional definitions. Here is the functional definition: A **confounder** provides an alternate explanation for an association. Here is the negative definition: **Confounders** are related variables that were not included in generating the association. Here is a positive definition: A **confounder** is a third factor that is related to an association, causes the result, and is not caused by the predictor.

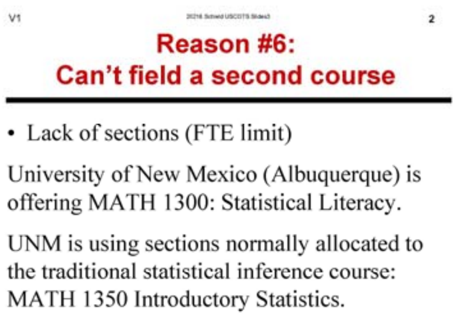
Instead of talking about confounding, I want to talk about mechanism. I never intended to mention mechanism in my class or my textbook. I define a mechanism as a related factor the causes the result in an association and is caused by the predictor.

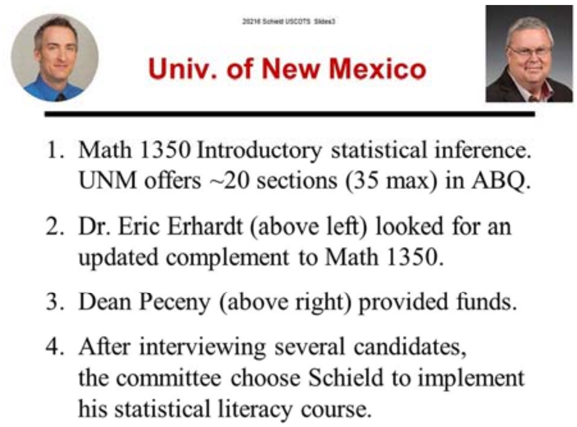
My students kept coming up with mechanisms. Each time, I had to tell them, "That's a mechanism."
So, much to my surprise, I teach them about mechanisms as well as confounders.

I know you are probably over-loaded. A mind can only take on so many new ideas and concepts. But we need to move on. How does this confounder-based statistical literacy course work? It is one thing to talk about it, in theory. What topics do you cover? How do you assess computational matters? How do you assess the hypothetical thinking and critical thinking?

I want to talk how this confounder-based statistical literacy course is being implemented at the University of New Mexico (UNM). UNM has put this Statistical Literacy course in their catalog and will be teaching it this fall as MATH 1300. It is offered *as an alternative* to their traditional statistical inference course. They are taking some of the sections from their traditional inference course and converting those to statistical literacy. I really think UNM is on the cutting edge in teaching confounding.

Slide 3.2: How did they fund this course? The allocated some sections of traditional statistical inference (MATH 1350) to the new statistical literacy course (MATH 1300).





Slide 3.3 (Upper-right): Dr. Eric Erhardt was the lead on this project. His job was to look for a compliment to MATH 1350: Introduction to Statistics. He attended USCOTS and JSM looking for ideas. Dean Peceny provided funds to get it started. I was chosen to implement my statistical literacy course.

Slide 3.4 (lower left): Implementing a new course at a large institution is really complex. We are not talking about a one-off topics course. We are talking about getting it through all the steps. It must satisfy a math requirement in the UNM core curriculum. It must satisfy a math requirement in the New Mexico state higher education general education curriculum. A given step may require several documents

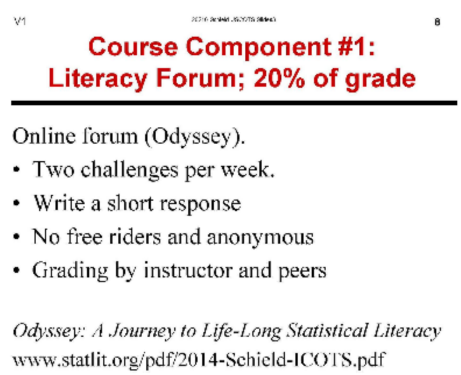
Slide 3.5 (upper right): The application must satisfy the requirements of the Registrar and the UNM budgetary and load implications along with the new course sign off by a lot of people.

Slide 3.6 (lower left): The application must provide the New Mexico Higher Education Department with the student learning outcomes (SLOs), how you are going to assess these SLOs, and how these SLOs related to the desired outcomes of the state generation education curriculum. There's a lot of paperwork involved and a number of committees. I am not going to cover any of that here. I will be providing that in my paper for the JSM in August. You can study the details when the paper is published. See Statistical Literacy Approved for General Education at the University of New Mexico: www.StatLit.org/pdf/2021-Schield-ASA.pdf

Slide 3.7 (upper right): All of that has been accomplished. MATH 1300 Statistical Literacy is in the UNM Albuquerque catalog. It may be slightly different at your school, but it can be done. I would be happy to work with anybody that wants to take on that project. Here is the catalog entry:

> Statistical Literacy (MATH 1300): *Participants will study the social statistics encountered by consumers. Investigate the story behind the statistics. Study the influences on social statistics. Study the techniques used to control these influences. Strong focus on confounding*.

That is a radically different course. The University of New Mexico should be acknowledged for pioneering in this area. We all know that new things never work out quite as well as we hope. So there will be some uncertainty and problems. You will see why, when we talk about teacher training. But UNM is pursuing a worthy goal.



Slide 3.8: Are students evaluated on their thinking and reasoning based on their writing? Yes, this is a literacy course; they have to write. This course is designed to be scalable. I have taught large auditorium classes with at one hundred students or more. Students use an online forum. This is 20% of their grade.

You may say, "Online forums don't work. Some students wait till the very end; they read everybody else's stuff and then they submit the best of what they read. They are free riders."

(1) This Odyssey forum doesn't allow free writers. A student doesn't see what anybody submits does until they have posted their answer.
(2) Everybody's anonymous. Some, like that; some don't, Some Somali women have said, "I like it. I would never post in a forum if people knew who I was." Some have said they don't like anonymous. They want to know who they are talking to.
(3) The challenges are usually quite short: a paragraph to a page. It may be a current news story, an interesting graph or table. No research required. Typically the topics allow good answers pro

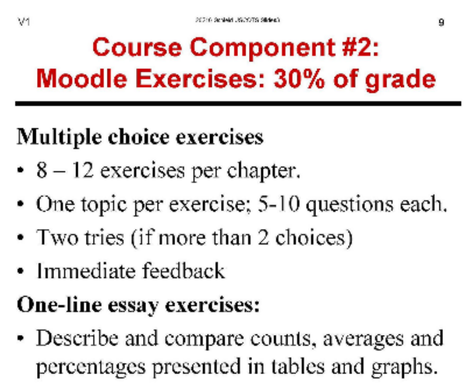
and con. The goal is to state your choice and give good reasons to support your choice. For example, their first challenge is decide how much math do college students need. They read an article by a math teacher saying they really don't need that much math unless required by their major. They can agree or disagree, but they have to define exactly what they mean by 'math', and they have to give their reasons.

(4) There are two challenges per week. You may be saying, "I am teaching 20-40 students per class; I am teaching two or three sections a semester. You are running two challenges, a week on 20-40 students per section. That is 40-80 papers, a week per section. You have got to be kidding!
(5) Students are grading each other. Students aren't used to this. Some don't like it saying "Why am I paying big bucks for a college education when I may be graded by one of the worst students in the class?" I note that after graduation, they may be evaluating others. As a supervisor or manager, you need to get used to giving reviews.
(6) The teacher can do as much or as little as they want. The teacher's posts are also anonymous.
(7) The Odyssey program computes an up-to-the minute score for each students based on all the grades each student receives from their reviewers.

Yes, this peer review does offload some of the grading from the teacher to the students. But the goal is to help the students review their own writing before they submit.

Choosing one side is different for them they're used to doing compare and contrast. Now they pick just one side and give arguments for it. They soon learn the goal is not to write lots of words. After doing it week after week after week, they get better at sizing up a story or argument. What kind of study is involved, what kind of statistics are there, what is taken into account, and what isn't taken into account.

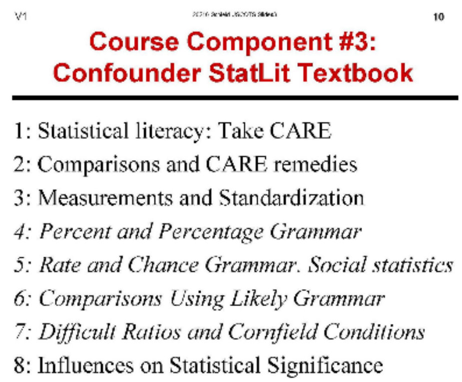
Slide 3.9: Online multiple choice right-wrong exercises with a few one-line essay questions are 30% of the grade. In Moodle, there are 8-12 exercises per chapter with one topic per exercises: 5-10 questions each. Two tries (generally) with immediate feedback. The one-line writing exercises require the student to describe and compare counts, average and percentages presented in tables, graphs and statements. These require teacher grading.



Some students do this as though it were a game. They go through it once without much study. If they get a high score, they are done. If they get a low score, they read the material and try again.

The one-line exercises are the hardest for most students. They are not used to using ordinary English with such precision: especially when dealing with percentages. With part-whole percentages, the confusion of the inverse becomes a major problem. They get things backwards. They get part and whole reversed in the exercise, in the chapter quiz, in the second quiz, and even in the final.

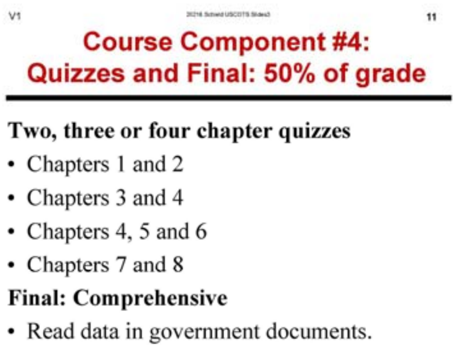
I just finished teaching my summer class. It's not just the ESL students. Some of the native English speakers have almost as much of a problem.

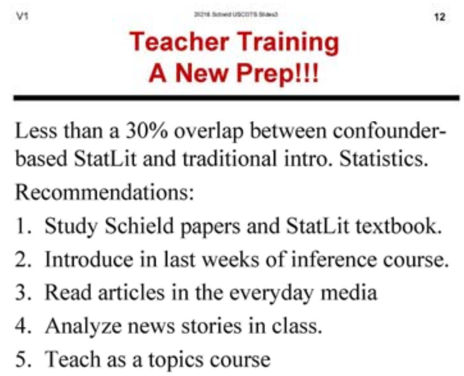


Slide 3.10: The course textbook has eight chapters: four of them on part-whole percentages. Initially you might just do the first three chapters. They present all the big ideas needed to analyze news stories. Then add more chapters one at a time. I'd recommend adding 7 and 8 first; chapter 4 second with chapters 5 and 6 last.

Slide 3.11 (left below): Each teacher will decide how to handle the chapter quizzes, tests and final. Most of the quizzes and tests are multiple choice with right-wrong answers. The writing is limited to a few one-line statements that describe or compare some statistics presented in a table, graph or statement.

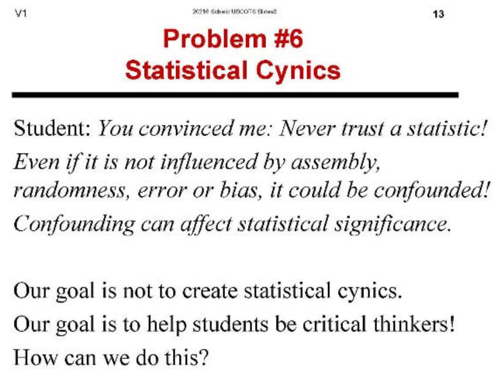
Part of their final involves something I got from my colleague, Marc Isaacson. Students have to answer questions based on the data presented in a one-page government info-graph. Short tables; lots of charts.





Slide 3.12 (right above): So what is the biggest problem in teaching this? This is a new prep – a very new prep. There is less than a 30% overlap with a traditional statistical inference course. It takes time to prepare yourself. This course is a wicked combination of words and numbers in ways that students have never seen before. Ideally a new teacher would have studied some of my papers, they would have gone through the textbook once or twice just let it sink in. They would have introduced confounding in the last week or two of their inference class before the final. Meanwhile, they are actively reading the everyday media looking for interesting articles to evaluate. They might bring a new story to class and work with the class in analyzing it. Eventually, they may teach it as a topics class.

It took me five to 10 years to make the transition and I came from a critical thinking background. With a textbook, it should take others less time.

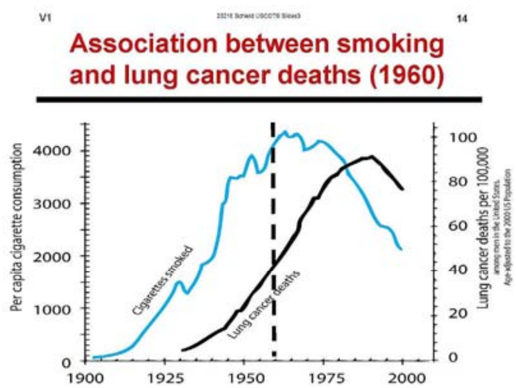


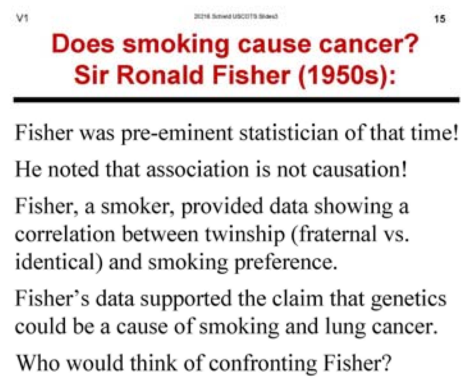
Slide 3.13: Here is another reason not to teach confounding. It was brought up by a colleague of mine: a fellow of the American Statistical Association. He said, we wouldn't teach anything that creates statistical cynics. Certainly we don't want that. We don't want our students to be naïve; we don't want them to be cynics. We want our students to be critical thinkers.

One student said to me, "You have convinced me. I will never trust a statistic now that I realize how easily they can be influenced or manipulated."

I knew there was no test for confounding or bias or for how things were assembled. In 1998, I heard Paul Rosenbaum talk about the Fisher-Cornfield debate at a national statistics conference. Most statisticians don't know much about Jerome Cornfield even though he created the odds ratio and relative risk, and he was a president American Statistical Association.

Slide 3.14 (left below): To learn about Cornfield, we need to go back to the 1950s when statisticians were seeing a strong association between smoking and lung cancer deaths. They could only see the part before 1960.



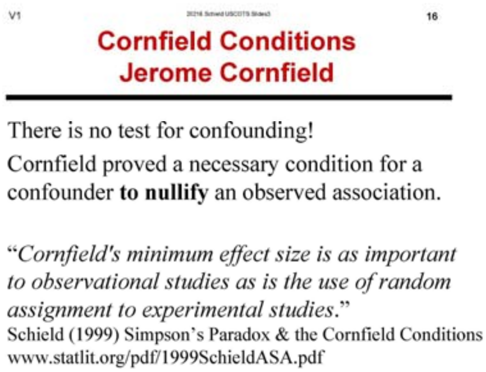


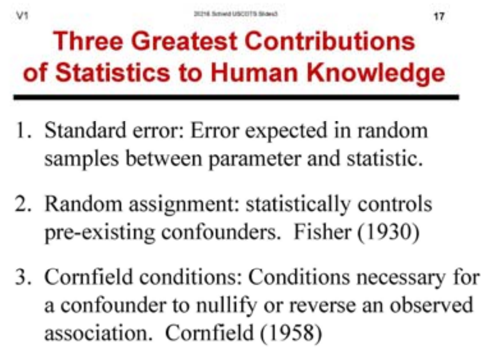
Slide 3.15 (right above): The pre-eminent statistician of that time, was Sir Ronald Fisher: a smoker. He reminded everyone that association was not causation in observational studies. Fisher may have been a smoker but he was a very smart smoker. He had data from a German twins study showing there was a correlation between the kind of twinship (fraternal versus identical) and smoking preference.

Now, if the world's number one statistician says "Association is not causation" (and you agree) and if he has data showing that there could be a genetic factor involving smoking preference, most statisticians would back off – just walk away. Who would confront Fisher and his most-excellent argument?

Slide 3.16 (left below): But Cornfield did confront Fisher. Cornfield knew there was no test for confounding. But Cornfield identified a necessary condition – a minimum effect size – for a confounder to nullify or reverse an observed association.

I argued that "Cornfield's minimum effect size is as important to observational studies as is the use of random assignment to experimental studies."





Slide 3.17 (right above): Cornfield's condition is 'huge'. I believe that Cornfield's conditions are one of the three biggest contributions of statistics to human knowledge along with random selection and assignment, and standard error. Standard error in random sampling is the doorway to margin of error, confidence intervals and hypothesis testing. Random assignment statistically controls for all pre-existing confounders and allows one to make statistically-strong statements about causation. The Cornfield conditions identify the conditions under which a confounder can nullify (or reverse) an existing association. These are three things I think all our students should know and understand after taking any introductory course.

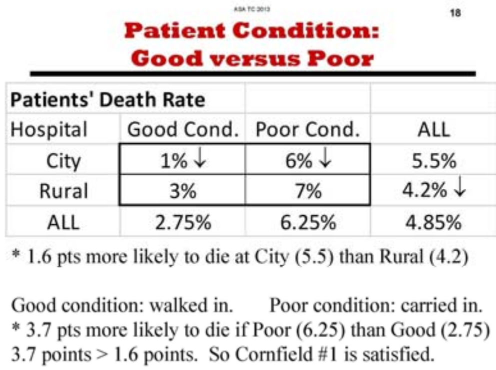
Slide 3.18 (left below): Since most statistical educators, have never heard of the Cornfield conditions let's talk about them. In this two-by-two table we have patient death rates for two hospitals: City and Rural. On the far right column are the average patient death rates: 5.5% for city and 4.2% for rural.[8]

But if we look in the bottom row we see these average patient death rates: 2.75% for patients in good condition and 6.25% for patients in poor condition.

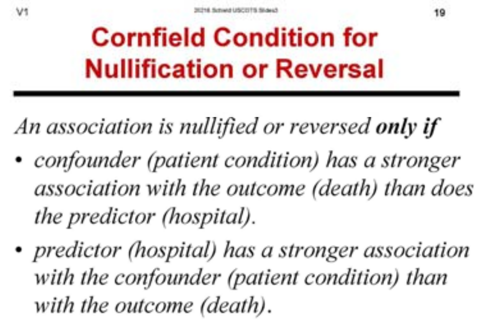
The difference in patient death rates for those City hospital versus Rural is 1.6 percentage points. The difference in patient death rates for those in good versus poor condition is 3.5 percentage points.

The difference in patient death rates by patient condition (3.5) is bigger than the difference in patient death rates by hospital (1.6). This satisfies Cornfield's condition.

[8] The rates in 3.18 and 3.20 differ from those shown in the video which didn't match those in 2.19 – 2.22.

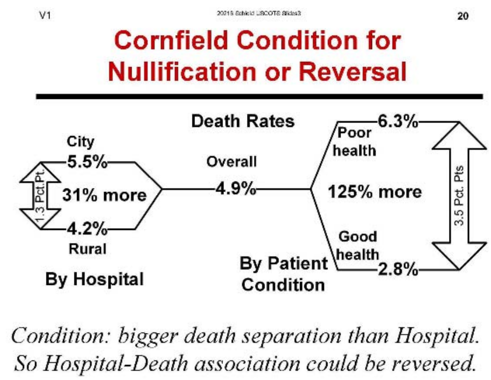


Patients' Death Rate			
Hospital	Good Cond.	Poor Cond.	ALL
City	1% ↓	6% ↓	5.5%
Rural	3%	7%	4.2% ↓
ALL	2.75%	6.25%	4.85%



Slide 3.19 (right above): Cornfield actually identified two necessary condition. I used to teach both, but now I just teach the first. I want to give students just enough so they avoid becoming statistical cynics.

Slide 3.20: This is how I present the Cornfield condition graphically. On the left side, we have the patient death rates by hospital: 5.5% versus 4.2%.[8] On the right side, we have the patient death rates by patient condition: 6.3% versus 2.75%.

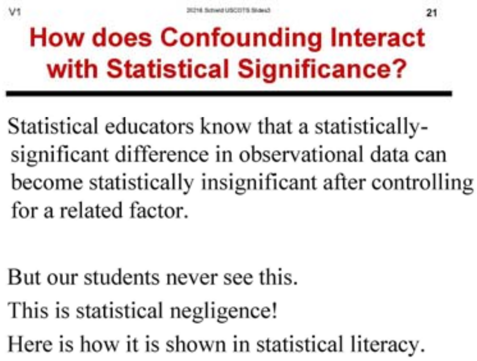


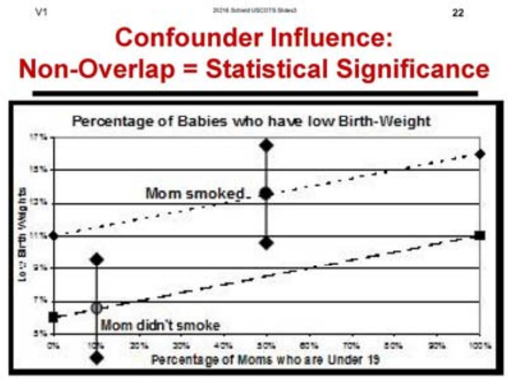
Students can see that percentage point difference is bigger on the right than on the left. .This satisfies the first Cornfield condition. Controlling for (taking into account) patient condition could actually reverse the association between hospitals – which it does in this case. This is Simpson's paradox.

Slide 3.21 (left below): If a confounder can reverse the direction of an association, can it transform statistical significance into statistical insignificance? As teachers, we know it can for observational data.

But our students never see this. I don't know of any introductory statistics textbook that mentions this. I think our students go away thinking that once you have statistical significance it's etched in stone.

I am on record for a long time in saying this is statistical negligence! We know it, but we don't show it. We unintentionally lead our students to a conclusion that we know is false. So how can we show it?





Slide 3.22 (upper-right): How do we show it in confounder-based statistical literacy, well? We use the same type of graph. Here we have the percentage of babies that have low birth weight by the mom's age and by her smoking status.

On the right, we have the low birth-weight rates for the younger moms (those under 19). On the left, we have the same rates for older moms (those at least 19). Connecting those points gives the diagonal weighted average lines.

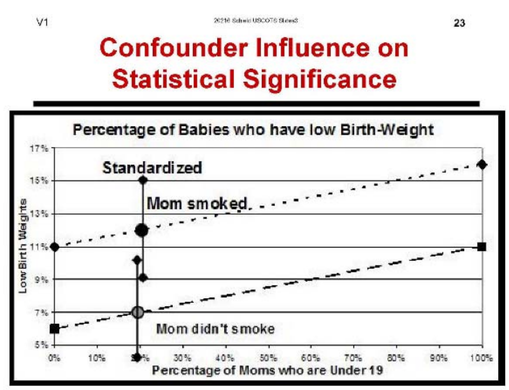
The percentage of moms who are young is 50% among those who smoke, 10% among those who didn't. This gives us the weighted average for each group. This is the same approach as for the hospitals.

This graph shows the confidence interval for each group. Notice that these confidence intervals do not overlap. That is sufficient to say that the difference in these sample statistics is statistically significant.

Yes, this is a crude approach to statistical significance. But is shows a very big idea. And that is our goal.

We notice the big difference in the mixture of younger moms. To control for (take into account) the age of our moms, we need to standardize. With parallel weighted average lines, it doesn't matter which kind of standardization is done.

Slide 3.23: W ere we standardize assuming 20% of all moms are young. Now the confidence intervals overlap. IN this statistical literacy course, we say the difference in the sample percentages is not statistically significant. Students can see the change. They can see how controlling for (taking into account) the influence of a confounder can change statistical significant into statistical insignificance.

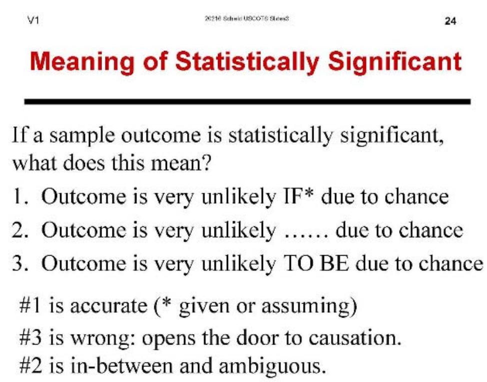


Ideally, there would be an example that shows the opposite.

This non-overlap criteria for statistical significance is a very crude test. We know this presentation makes several assumptions. (1) The margin of error is the same for equal sized samples even though they have different proportions. (2) The margin of error does not change even though the sample proportions do change in standardizing. (3) If the confidence intervals overlap then the difference in sample proportions is statistically insignificant.

As statisticians, we know that all three of the assumptions are false. What justifies making such blatant errors in teaching our students? Do we want to teach bad practice? No. But we must not let the perfect become the enemy of the good. Our primary goal is not to train – or attract – future statisticians. Those students can learn more precise methods in later courses.

Our primary goal in the introductory course is to introduce the big ideas of statistics. And to introduce them in a way that is simple and memorable. Furthermore the lack of overlap in 95% confidence intervals is sufficient to conclude that the difference in sample statistics is statistically significant.



Slide 3.24: This statistical literacy course uses 'statistically significant' instead of p-values. This avoids having to teach a number of ideas including the binomial distribution, the normal distribution, Z-values, etc. Statistical significance is a big idea – one of the biggest in statistical inference. But, there is still a lot of misunderstanding about the meaning of 'statistically significant' beyond mistaking it for 'important'.

This slide presents three different interpretations of statistically significant: (1) unlikely If due to chance, (2) unlikely due to chance, and (3) unlikely to be due to chance.

Here is another case where small changes in syntax can create big differences in semantics.

The first one is accurate. Statisticians may use other words such as 'given', 'when' or 'assuming' in place of 'if'. All these uphold the idea that chance in sampling from the null population is the premise.

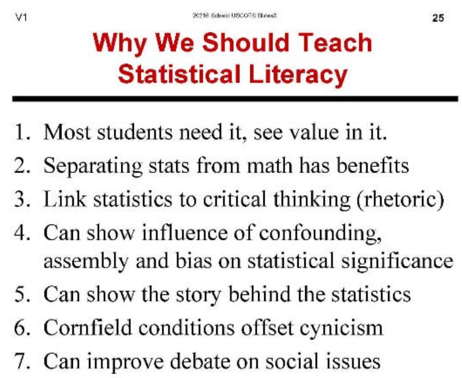
The third is wrong because as frequentists we never say anything about the validity of the premise.

The second is ambiguous. The second one is the one that many people love, because it can be read either way. Note that these three statements are similar to the the A-B-C grammar in distinguishing association from causation.

I know these are picky distinctions. But as statistical educators we know they are important distinctions. We need to sensitize our students to these small grammatical distinctions. They may not remember the details of these subtle differences, but at least they'll be more likely to "Take CARE" when making that kind of a statement or hearing it.

This concludes my overview of statistical literacy: specifically the confounder based portion of statistical literacy. Let me repeat my opening statement:

Confounding is the elephant in introductory statistics education. It is big. We know it is big in the everyday media. But we don't talk about it or teach it. I'm here to persuade you to teach confounding.



Slide 3.25: So why should we teach confounder-based statistical literacy? Here are my seven reasons:

#1: Most of our students need to understand confounding because they deal primarily with observational studies. But the need is bigger than just the students whose majors require them to take statistics. Most students deal with observational studies, so they all need to understand confounding.

Many, if not most colleges are requiring a quantitative literacy quantitative reasoning requirements. In many cases, students majoring in English, art, music, communications, political science, history, religion and philosophy are taking statistics. They need a confounder-based statistical literacy course. And they see value in it.

#2: Offering this course will separate us from mathematics. It will separate us more concretely than just saying, "Statistics are numbers in context." All too often, we are pretty vague on this context stuff.

This course takes no prisoners. It includes the entire context, all the influences on a statistic.

#3: This statistical literacy course links statistics to critical thinking by teaching quantitative rhetoric.[9] As teachers, we all want to encourage students to become better critical thinkers. In today's world, there are numbers involved in many – if not most – of our social and political arguments.

#4: This presentation focused on confounding. Traditional statistical inference focuses on randomness and bias. In this presentation we haven't talked about Assembly. That's another workshop. Assembly or assumptions is big: it's huge! This course shows how confounding, assembly and error or bias can influence statistical significance.

#5: Showing the story behind the statistics is big. They've never seen that statistics may be closer to words than they are to numbers. Once they can see stories behind the statistics, they can begin to think about what might be the story.

#6: This confounder based course includes the Cornfield conditions. These are the conditions that help prevent statistical cynicism. Only the biggest confounders can nullify or reverse an association.[10] I've never heard of any statistics course in the world that includes the Cornfield conditions. This statistical literacy course is not just stat-light. It has new content, new methods, and new assessment. It is different: very different!

#7: For me, the biggest reason it this last one. Teaching confounder-based statistical literacy can help improve the quality of our debate on social and political issues. No, you don't have to teach this course. But for me, teaching confounding in an introductory course: that's our job!

[9] Schmit, John (2010). Teaching Statistical Literacy as a Quantitative Rhetoric Course. Copy at www.statlit.org//pdf/2010SchmitASA6up.pdf
[10] Schield, M. (2017). Confounding and Cornfield: Back to the Future. www.statlit.org/pdf/2018-Schield-ICOTS.pdf

I'm looking for a handful of people that are going to take on this challenge. They will (1) read more on Schield's approach to statistical literacy.[11] (2) Buy Schield's Statistical Literacy textbook and study how it handles various topics. (3) Introduce some confounder-based topics into their teaching of traditional statistical inference. (4) Try teaching a confounder-based workshop to colleagues, friends or students. (5) Schedule and teach a *separate* confounder-based statistical literacy course. And (6) schedule and teach a catalog-approved confounder-based statistical literacy course as an alternative to a traditional statistical inference course.

For me, confounder-based statistical literacy is the next big thing in statistics. This change in content is much bigger than the many changes in pedagogy: inverted or flipped classrooms, using real data instead of synthetic or toy data, or replacing distribution-based concepts with simulation-based results.

None of these pedagogical changes can improve the quality of our political and social arguments. I think those few people that take this challenge will be known as the first pioneers. I'm happy to talk with you and work with you. I can help you go through the steps of getting a new course approved.[12]

Slide 3.26: As for resources, start any of my 70+ papers on statistical literacy. The list of my papers at www.StatLit.org/Schield-pubs.htm is organized by topic, so you pick the ones you find interesting.

My Wiley textbook is scheduled will be out in 2022.

Contact me if you want to discuss things, teach specific topics or an entire course. I am available.
> Milo Schield <Schield@Augsburg.edu>
> Milo Schield <SchieldMilo@UNM.edu>

Are there any questions? [Silence]

Milo: Let me ask a hard question: "What is the downside of using your course? Milo answering: "Wiley is not putting the exercises up in their grading system. For New Mexico, I am putting them up in an offsite Moodle site."

Marian Frazier (she/her)
You teach this as at Augsburg. You have a traditional inferences course and then this Statistical Literacy course. Does it [the Statistical Literacy course] serve as a prerequisite to a statistical modeling course?

Milo Schield
I wouldn't recommend it. I'd have to think about whether I would allow it.

Marian Frazier (she/her)
So this course is for students who may never take another statistics course. "Hey, I'm an English major and I just want my quantitative reasoning requirement so I can graduate."

[11] Schield, M. (2021). Schield publications arranged by topic. See www.statlit.org/Schield-Pubs.htm
[12] Schield, M. (2021). Univ. of New Mexico

Milo Schield
Or I'm a business management major (I'm not looking to be a statistician); I'm a marketing major.

Marian Frazier (she/her)
Which of course is the majority that we have. Right?

marina
Milo, I have a question: Do you have any prerequisites for this class?

Milo Schield
None aside from the algebra -math requirement for all courses at this level. Parts of this class could easily be done at the junior college, community college or high school level. My students have said that some of the grammar stuff could and should be done in middle school.

Sylvia Kuzmak
I do have a question too. It's clear that understanding the domain is critical in discussing confounding. What do you think about just sticking within one domain and let people get into that domain. If it's going to be healthcare in hospitals, stick with that for a while and let them.

Milo Schield
Good idea. In writing a textbook I admit I tried to jump domains: do crime, do hospital, etc., etc. As a teacher, I might stick closer to one domain. That way students could get more an idea of what the relevant confounders are. On the other hand, they're going to be reading things from a lot of domains in the real world. I think we want to expose them to a little bit of variety.

Sylvia Kuzmak
You spend a lot of time with the graphs. What about just having at least one slide that gives a layout of knowledge in the domain so, for example, with the healthcare one. [For example] When you have people a poor condition, you know for some ailments they're all going to end up in the city hospital. They're not going to end up in the rural hospital because they [the rural hospital doesn't have the facilities.

I would even just have a graphic of this: what this domain is like; this is how things happen in the domain. It would sort of show you that these are going to be candidates for confounding factors when you try to analyze the data.

Milo Schield
Makes sense. I'd leave that up to the individual teacher.

Unknown Speaker
These poll results are very subject to sampling bias.

Milo Schield
Absolutely. I didn't cover error and bias at all in this workshop. Not enough time. But sampling bias is just rampant in anything observational. Every time we do anything observational, I ask, "What's the possibility of sampling bias?" Is sampling bias possible? Absolutely.

As an author I have tried to provide a framework for students to think of all these things. As long as they can say under Error, Yes that's bias, and under bias they see sampling bias, then they can see where that fits in. I've asked students what concepts, they found to be most valuable. I didn't I gave them a list of like 30 to choose from. I summarized their ideas into a list of topics. I asked them to rank them.

"Take care" was the one that they found most valuable for organizing the information in their minds, I was very surprised. I didn't expect that. I'm not saying that proves that "take CARE" is the best. It says that students like a framework. You could maybe use three groups or five or something else. Students need a framework to put all these ideas onto.

komaroff
I have a question. I use the David Moore's book, *The Basic Practice of Statistics*. They make a distinction between confounding variables and lurking variables. And now I also have hidden variables in my mind and I tried to explain it. And then I realized I got confused and I stopped trying to explain it. But I was hoping that maybe you can clarify it for me. Do you see a distinction between lurking variables, confounding variables and hidden variables?

Milo Schield
Aside from mechanism, No. I need to look back at David More's book and see. [After the workshop: According to Moore, a confounder is a measured variable; a lurking variable is an unmeasured variable.[13] Milo: I use confounder to indicate either measured or unmeasured.]

komaroff
In education they've published articles about lurking variables. We use that term so it's rather popular in the social sciences. That's what I'm working in right now.

Milo Schield
I will be holding an USCOTS Birds of a Feather on confounding. You might think about this and maybe have some questions. If you want to talk more about the teaching of confounding I will be there.

In conclusion, let me repeat the point of this workshop:

Confounding is the "elephant" in introductory statistics education. It is big. We know it is big in the everyday media. But we don't talk about it or teach it. My goal is to persuade you to teach confounding.

Thank you for attending. I really appreciate having this group to talk to. Best wishes and Take CARE!

ACKNOWLEDGEMENTS:
Thanks to Cynthia Schield for reviewing the first part of this document and for recommending that I insert the summary in boxed italics. This text is loosely based on a machine transcript of Schield's oral presentation in his 2021 USCOTS two-hour workshop. The slides are those presented in the workshop with corrections to slides 3-18 and 3-20.

REFERENCES:
Paper: www.StatLit.org/pdf/2021-Schield-USCOTS.pdf
Slides: www.StatLit.org/pdf/2021-Schield-USCOTS-Slides.pdf

[13] https://www.coursehero.com/file/7936174/Lurking-Variables-and-Confounding-Variables/

Attendees were invited into breakout rooms to discuss the advantages and disadvantages of teaching confounding. They then recorded their own conclusions.

1st Breakout: What are the advantages to teaching confounding?

1	forces you to think about multiple variables, shows that stats isn't a "traditional" math class, it requires reading and writing (communication!), it breaks the habit of just memorizing a bunch of tools like oh here, I use a t-test here
2	Widens the view beyond just this test with these variables.
3	Makes them more likely to understand the things they see in the world?
4	Better statistical literacy
5	Students will encounter it, so we should help them reason about it.
6	What good is it to do statistics if they haven't learned about confounding?
7	I think students may feel like stats is too clean-cut to be really practical without discussing the messiness of data involved with confounding.
8	Necessary for real decision making
9	It is important to help students understand the important issues and associated language so that they really do become statistically literate.
10	Getting things right!
11	It helps students understand that conclusions in statistics are not reliable.
12	many students who are not as interested in the math-focused procedural knowledge may be more engaged, there is a focus on the data generating process, forces multivariable thinking, and a look at association and real-world connections.
13	It's a necessary part of teaching reasoning with statistics -- that is, actually addressing finding answers to real world problems/ assessing scientific "truth". It can be motivating for students to learn a more complete set of knowledge and skills (including confounding) to do reasoning with statistics.

1st Breakout: What are the disadvantages to teaching confounding?

1	intro stat/AP Stat has a lot to cover and set expectations
2	That students may be convinced that they cannot trust any statistical result. It is hard to teach well.
3	What do we drop?
4	Time to fit it into a traditional course.
5	It is hard to do. Students might go overboard on thinking that they can ignore anything they don't like because "maybe there confounding that explains away what I don't like."
6	It is no good to do it until you get it right!
7	Takes away from time that can be used for teaching other important intro concepts
8	Hard and takes a lot of time
9	It is more information to include in a course and it takes time to make the change.
10	It takes time and sometimes you also have to have less emphasis on other "required" topics. It also seems like it is sometimes hard to come up with relevant examples for students to better understand the concept.
11	Hard for first semesters to understand confounding when they are just beginning to understand statistical language.
12	time to cover other material, student buy in from those who wanted or prefer math as opposed to writing, available material
13	It's a hard concept to learn and takes time, as is true for many statistics concepts.

2nd Breakout: What are the advantages to teaching confounding?

21	More wide view of what statistics does and doesn't show
22	feel the weight of choices you make as a statistician, show students that sometimes we have to stop without having a final answer
23	Mostly, we discussed how we thought the race and income example did so many extremely useful things for teaching stat literacy and confounding topics.
24	I think this has the power to really make the students see the relevance, importance, and LIMITS of statistics.
25	can get deeper into a subject/content area as you consider multiple variables
26	Providing a truly critical thinking based course.

2nd Breakout: What are the disadvantages to teaching confounding?

21	Students believe everything can be explained by confounding.
22	sometimes you need domain expertise to be able to do the hypothetical thinking required, do we have it as teachers? will students come in with it?
23	Difficulty of including new material in an already jam-packed course.
24	I still don't know what to cut to make room for this stuff! :(Or is a whole new course?
25	data limited to what is collected and might not include confounding variables
26	Changing what we are used to doing. Deciding what to keep and what to cut.