UNM Statistical Literacy: Design and Rationale

Milo Schield

Statistical Literacy Coordinator: Univ. of New Mexico
Fellow: American Statistical Association
Elected Member: International Statistical Institute
US Rep: International Statistical Literacy Project
President: National Numeracy Network

February 18, 2022

www.StatLit.org/pdf/2022-Schield-UTEP-Slides.pdf

Today's student need to study Statistics

Disparities in
- Education, suspensions and graduation
- Policing, crime, sentencing and prison
- Wages, income, assets, loans and wealth
- Health, health care, homicides and deaths

Disparities by
gender, race, ethnicity, religion, politics, age, etc.

All of these rely on statistics: social statistics.

University of New Mexico is offering a new course!

Taught 4 sections in fall 2021

Statistical Literacy UNIVERSITY CATALOG

MATH 1300 (3)
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.

Meets New Mexico General Education Curriculum Area 2: Mathematics and Statistics.

Math1300: Statistical Literacy

Less than 30% overlap with traditional statistics

Math1300: Statistical Literacy

STATISTICS STUDIES VARIATION

Two kinds of variation

SYSTEMATIC	RANDOM
Confounding (control for: selection, ratios, regression)	**Probability**
Assembly (define, present)	**Sampling error**
Error (Bias)	**Confidence intervals**
Systematic Influences	**Test of Hypothesis**
	Statistical Inference

Math1300: Statistical Literacy

TWO KINDS OF INTRODUCTORY STATISTICS COURSES

Statistical Literacy. *More confounding than statistical inference.*	Quantitative Methods. *More statistical inference than confounding.*
UNM Math 1300	**Utts, Bennett, Stat 101**

By what standard is the content selected?

Math1300: Statistical Literacy

Statistical literacy is designed for the consumers of statistics: students, citizens, and decision makers.

- Students in non-quantitative majors.

But this statement of audience is still ambiguous.

Who decides what the content will be?

- Statisticians: GAISE 2016 update?
- The statistics in research articles?
- The statistics in the everyday media?

Math1300: Statistical Literacy Statistics in the everyday Media

This is what these students "NEED".

Statistical content analysis:

- hundreds of news stories containing statistics
- 46,000 articles in the Harvard Business Review
- 4.5 billion words in UK Cobuild corpus
- 1 billion words BYU corpus (COCA)

Statistical Literacy: Social Statistics

Statistics are different from numbers

Statistics are numbers in context (in reality)

Statistics can be influenced by reality:

- In arithmetic, 1 plus 1 is always 2.
- In reality math:
  1 bunny plus 1 bunny can give three bunnies
  1 ice-cube plus 1 ice-cube can give zero ice-cubes

Statistics are *Socially Constructed*

Joel Best, author of "Lies, Damned Lies and Statistics" identified this as fact as *the most important, the most fundamental* aspect of *all* reality-based statistics.

He didn't mean all of reality was a mental construct.

He meant that statistics, just like words, are created by people: people with motives, values and goals.

Each statistic embodies a particular view of reality.

Statistics can Be Influenced

Q. Best advice when dealing with statistics?

A. "Take CARE". Statistics can be influenced.

All influences are grouped into four categories:

C: Confounding

A: Assembly (how things are defined, counted, etc.)

R: Randomness

E: Error (including bias)

Admonition: "Take CARE"

Students like "CARE". It gives them a structure.

Without a unifying structure, statistical literacy is just a collection of disparate influences.

Everything they learn about influences fits.

When asked to rank what they considered most valuable, they chose "CARE".

If they remember just one thing from the course, it should be "Take CARE".

V1 2022 Schield UTEP Symposium (slides) 13

Seven Student Failures: Failure to recognize that…

1. **Association is not causation**
2. Statistics can be manipulated
3. Statistics can be confounded: Simpson's paradox
4. Bigger data, the more likely an unlikely statistic
5. Ratio statistics are ordered: confusion of the inverse
6. Spotty statistics are modeled
7. Statistical significance can be influenced

V1 **Failure #1** 2022 Schield UTEP Symposium (slides) 14

Association is Not Causation

Semantics: Association is not [necessarily] Causation

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

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

V1 **Failure #1** 2022 Schield UTEP Symposium (slides) 15

Disparity is Not Discrimination

Semantics: Differences or Disparities are not [necessarily] Discrimination

A: Association	B: Between (moral)	C: Causation (moral)
Math Differences: Count/Rate/Amount	Descriptive Differences with a Moral Connotation	Immoral Differences: 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, times as much, percent more	disparity / disparate impact over/under represented	discrimination* 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

V1 2022 Schield UTEP Symposium (slides) 16

Seven Student Failures: Failure to recognize that…

1. Association is not causation (eg., disparity…)
2. **Statistics can be manipulated**
3. Statistics can be confounded: Simpson's paradox
4. Bigger data, the more likely an unlikely statistic
5. Ratio statistics are ordered: confusion of the inverse
6. Spotty statistics are modeled
7. Statistical significance can be influenced

V1 **Failure #2** 2022 Schield UTEP Symposium (slides) 17

Statistics can be Manipulated Assembly/Assumptions

The number of children killed by gunfire has doubled **each year** since 1950;

The number of children killed by gunfire **each year** has doubled since 1950

30% of middle school students are **bullied**.
Define bullying. Increase or decrease the percentage

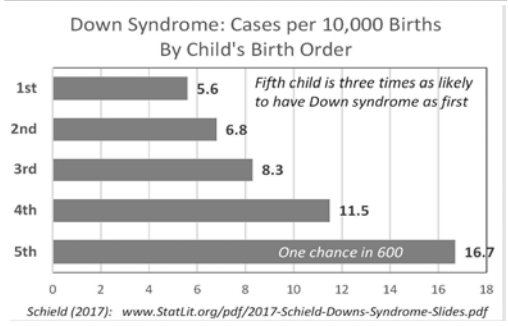
V1 2022 Schield UTEP Symposium (slides) 18

Seven Student Failures: Failure to recognize that…

1. Association is not causation
2. Statistics can be manipulated
3. **Statistics can be confounded: Simpson's paradox**
4. Bigger data, the more likely an unlikely statistic
5. Ratio statistics are ordered: confusion of the inverse
6. Spotty statistics are modeled
7. Statistical significance can be influenced

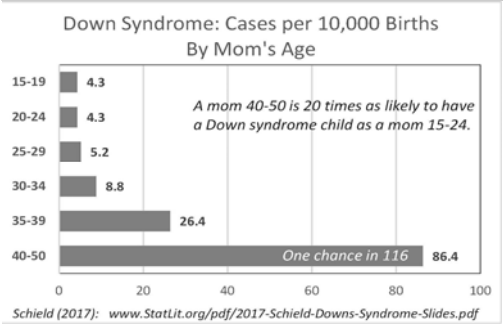
V1 **Failure #3** 2022 Schield UTEP Symposium (slides) 19

Statistics can be Confounded Downs Syndrome



V1 **Failure #3** 2022 Schield UTEP Symposium (slides) 20

Statistics can be Confounded Down Syndrome



V1 **Failure #3** 2022 Schield UTEP Symposium (slides) 21

Statistics can be Confounded Vaccinated: More Likely to Die

In the UK, vaccinated cases are 2.4 times as likely to die from Covid as [are] the unvaccinated cases.

National Health Service. 268,169 cases.

A crude comparison (mixed fruit comparison).

What could confound this association?

Age!

V1 **Failure #3** 2022 Schield UTEP Symposium (slides) 22

Statistics can be Confounded Vaccinated: More Likely to Die

UK 260,000 cases May-Sept 2021

Covid Death Rates per Case			Crude	Fraction of Cases			Adjusted
	<50	50+	All	<50	50+	All	Standard
Un-vac	0.03%	5.96%	**0.17%**	0.977	0.023	1.000	**0.71%**
Vaccinated	0.02%	1.68%	**0.41%**	0.767	0.233	1.000	**0.21%**
0.17% = 0.977*0.03% + 0.023*5.96%				0.885	0.115	1.000	
0.41% = 0.767*0.02% + 0.233*1.68%				**0.21%**=.885*.02%+.115*1.68%			
		Ratio	2.4	Ratio	10.2	Ratio	3.4

After standardizing, unvaccinated are more likely to die than [are] the vaccinated.

V1 **Failure #3** 2022 Schield UTEP Symposium (slides) 23

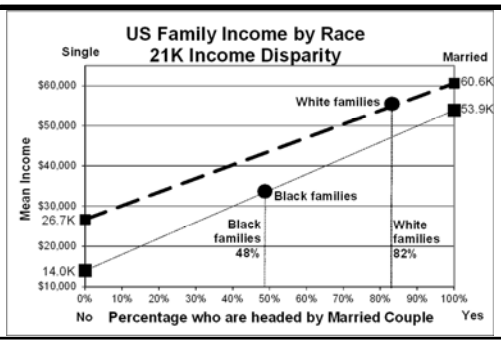
Black-White Income Gap 22K Income Disparity

Family Income			
Total	Race	Single	Married
$55,000	Whites	$26,700	$60,600
$33,000	Blacks	$14,000	$53,900

Distribution	
Single	Married
18%	82%
52%	48%

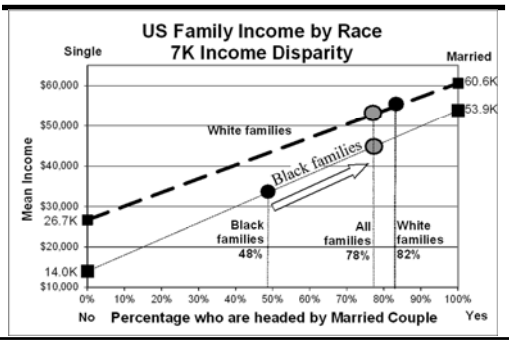
V1 **Failure #3** 2022 Schield UTEP Symposium (slides) 24

Standardization Graphically: 21K Income Disparity



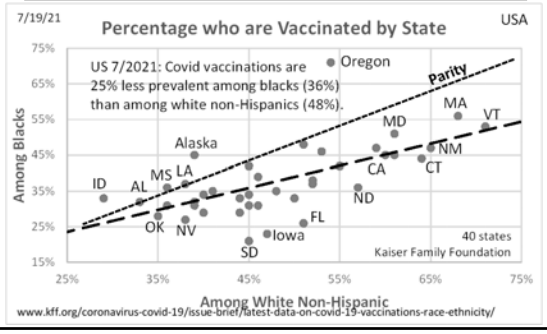
V1 **Failure #3** 2022 Schield UTEP Symposium (slides) 25

Standardization Graphically: 7K Income Disparity



V1 **Failure #3** 2022 Schield UTEP Symposium (slides) 26

Blacks less likely to be Vaccinated than White NH



V1 **Failure #3** 2022 Schield UTEP Symposium (slides) 27

Students Need Practice with Confounding

Understand how a crude association can be a:

- Mixed fruit comparison
- An apples and oranges comparison.

Students need to really understand what it means:

- To take something into account
- To control for something.
- To standardize an association:
  To give both groups the same mix!

V1 2022 Schield UTEP Symposium (slides) 28

Seven Student Failures: Failure to recognize that…

1. Association is not causation
2. Statistics can be manipulated
3. Statistics can be confounded: Simpson's paradox
4. **Bigger data, the more likely an unlikely statistic**
5. Ratio statistics are ordered: confusion of the inverse
6. Spotty statistics are modeled
7. Statistical significance can be influenced

V1 **Failure #4** 2022 Schield UTEP Symposium (slides) 29

Randomness and Big Data

The bigger the data, the more likely a rare event.
The Law of Truly Large Numbers.

Suppose an unlikely event occurs one chance in N.
Given N tries,

- one such event is *expected*, and
- at least one such event is more likely than not.

Schield (2009).

V1 **Failure #4** 2022 Schield UTEP Symposium (slides) 30

The bigger the data, the more likely an unlikely outcome

V1 2022 Schield UTEP Symposium (slides) 31

Seven Student Failures: Failure to recognize that…

1. Association is not causation
2. Statistics can be manipulated
3. Statistics can be confounded: Simpson's paradox
4. Bigger data, the more likely an unlikely statistic
5. **Ratio statistics are ordered: confusion of the inverse**
6. Spotty statistics are modeled
7. Statistical significance can be influenced

V1 **Failure #5** 2022 Schield UTEP Symposium (slides) 32

Distinguish Numerator and Denominator

1. The percentage of smokers who are men.
2. The percentage of men who are smokers.
3. The percentage of men among smokers.

Confusion of the inverse

1. Guys are more likely to smoke than gals.
2. Guys are more likely among smokers than gals.

V1 **Failure #5** 2022 Schield UTEP Symposium (slides) 33

Distinguish Percent and Percentage Grammar

If "20% of guys are smokers",
then 20% is *the percentage of guys who smoke*.

So if 20% of guys who run are smokers,
then 20% is
the percentage of guys who run who are smokers.
This last phrase is ambiguous!
What is the status of *run*?

V1 2022 Schield UTEP Symposium (slides) 34

Seven Student Failures: Failure to recognize that…

1. Association is not causation
2. Statistics can be manipulated
3. Statistics can be confounded: Simpson's paradox
4. Bigger data, the more likely an unlikely statistic
5. Ratio statistics are ordered: confusion of the inverse
6. **Spotty statistics are modeled**
7. Statistical significance can be influenced

V1 **Failure #6** 2022 Schield UTEP Symposium (slides) 35

Distinguish Counted Counts from Modeled Counts

V0A **Failure #6** 2021 Schield USCOTS

Responsible for …

Study blames diesel for deaths

By Jon Brodkin / Daily News Staff
Wednesday, February 23, 2005

Diesel pollution is responsible for more deaths than drunk drivers and homicides, according to a new study that estimates how many premature deaths, asthma attacks and heart attacks are caused by diesel pollution in every U.S. county.

Nationwide, diesel pollution causes 21,000 premature deaths each year, including 475 in Massachusetts and 81 in Middlesex County, robbing those who die of an average of 14 years of their lives.

V0A **Failure #6** 2021 Schield USCOTS 37

2. Epidemiological statistics are common!
US Annual Deaths **Attributable To**:

Smoking: 467,000	Blood pressure: 395,000
Overweight: 216,000	Inactivity: 191,000
Blood sugar: 190,000	LDL cholesterol: 113,000
Dietary salt: 102,000	Low omega-3 : 84,000

High dietary trans fatty acids: 82,000
Alcohol use: 64,000 (90,000 less 26,000 averted)
Low intake of fruits and vegetables: 58,000
Low poly-unsaturated fatty acids: 15,000

www.emaxhealth.com/2/24/30740/smoking-high-blood-pressure-obesity-top-preventable-death-causes.html

V1 **Failure #6** 2022 Schield UTEP Symposium (slides) 38

Distinguish Counted Counts from Modeled Counts

Counted deaths are coroner certified.
Some causal deaths are coroner certified: alcohol.

Modeled deaths are statistical deaths.
Smoking, second-hand smoke, obesity, etc.
These counts are speculative (spotty) statistics.

Schield (2009): Confound those Speculative Statistics
www.StatLit.org/pdf/2009SchieldASA.pdf
Schield (2011): Epidemiological Models and Spotty Statistics
www.StatLit.org/pdf/2011SchieldISI.pdf

V0A **Failure #6** 2021 Schield USCOTS 39

Epidemiological statistics encourage seductive grammar

Consider these titles of news stories:
- 45,000 deaths *attributable to* uninsurance
- 45,000 deaths *associated with* lack of insurance
- Lack of insurance *linked to* 45,000 deaths
- 45,000 die … *because of* lack of health insurance
- Lack of Health Insurance *Kills* 45,000 a Year
- Lack of Health Insurance *cause* 44789 deaths
- Lack of insurance *to blame* for almost 45,000 deaths

Source: www.StatLit.org/pdf/2010SchieldICOTS.pdf

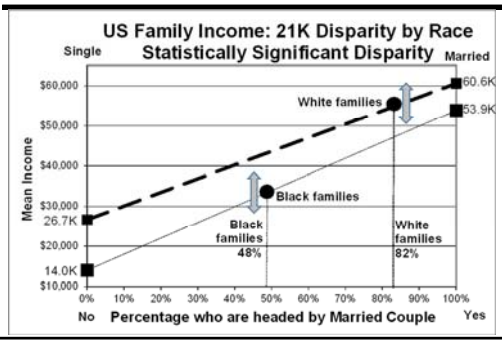
V1 2022 Schield UTEP Symposium (slides) 40

Seven Student Failures: Failure to recognize that…

1. Association is not causation
2. Statistics can be manipulated
3. Statistics can be confounded: Simpson's paradox
4. Bigger data, the more likely an unlikely statistic
5. Ratio statistics are ordered: confusion of the inverse
6. Spotty statistics are modeled
7. **Statistical significance can be influenced**

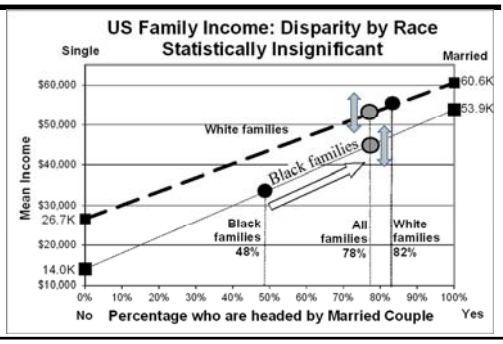
V1 **Failure #7** 2022 Schield UTEP Symposium (slides) 41

Statistical significance: Influenced by Confounding



V1 **Failure #7** 2022 Schield UTEP Symposium (slides) 42

Statistical Insignificance: Influenced by Confounding



V1 2022 Schield UTEP Symposium (slides) 43

Result: Students should:

Know that statistics can be influenced.

Understand "control for" and "take into account."

Know that standardizing converts a mixed fruit comparison into an apples and apples comparison.

Recognize the possibility of a Simpson's paradox.

V1 2022 Schield UTEP Symposium (slides) 44

Math1300 Highlights

Asserts that *Association is Not Causation*
Asserts that *Disparity is Not Discrimination*

Focus on *The Story Behind the Statistics*

Shows how a *crude association* (mixed fruit comparison) may conceal the real story!

Shows students how to *control for* confounders

Shows students these things *without computers*

V1 2022 Schield UTEP Symposium (slides) 45

Anonymous Student Survey

Before finals, students are asked three questions.

1. Did this course improve your critical thinking?
2. Would you recommend this course to a friend?
3. Should all students be required to take this course?

When I teach traditional statistics I get 'Yes' (Agree or Strongly agree) from 15% to 25% of the students.

When I teach Statistical Literacy to art, music and management majors, I get 'Yes' from at least 50%.

V1 2022 Schield UTEP Symposium (slides) 46

Anonymous Student Comments (Fall 2021)

I like the content and critical thinking aspect of the class. As someone who had to drop the regular stats class, I was very happy to have this class as an option.

This course is an answer to my prayers, I am a music major and horrible at math so fulfilling my math requirement has been hard. This is the first math class I actually liked. … the material is about things I can apply to everyday life. …I would recommend this class for anyone.

V1 2022 Schield UTEP Symposium (slides) 47

Study Confounder-Based Statistical Literacy

Statistical Literacy: What Students Like about the Course
www.statlit.org/pdf/2021-Fall-UNM-MATH1300-S1.pdf

Statistical Literacy: The Diabolical Denominator
www.StatLit.org/pdf/2021-Schield-MathFest.pdf

Statistical Literacy: Teaching Confounding
www.StatLit.org/pdf/2021-Schield-USCOTS.pdf

University of New Mexico Offers Math 1300
www.StatLit.org/pdf/2021-Schield-ASA.pdf

Schield's papers: www.StatLit.org/Schield-Pubs.htm

UNM Statistical Literacy: Design and Rationale

Milo Schield

Statistical Literacy Coordinator: Univ. of New Mexico

Fellow: American Statistical Association

Elected Member: International Statistical Institute

US Rep: International Statistical Literacy Project

President: National Numeracy Network

February 18, 2022

www.StatLit.org/pdf/2022-Schield-UTEP-Slides.pdf

Today's student need to study Statistics

Disparities in

- Education, suspensions and graduation
- Policing, crime, sentencing and prison
- Wages, income, assets, loans and wealth
- Health, health care, homicides and deaths

Disparities by

gender, race, ethnicity, religion, politics, age, etc.

All of these rely on statistics: social statistics.

University of New Mexico is offering a new course!

Taught 4 sections in fall 2021

Statistical Literacy

UNIVERSITY CATALOG

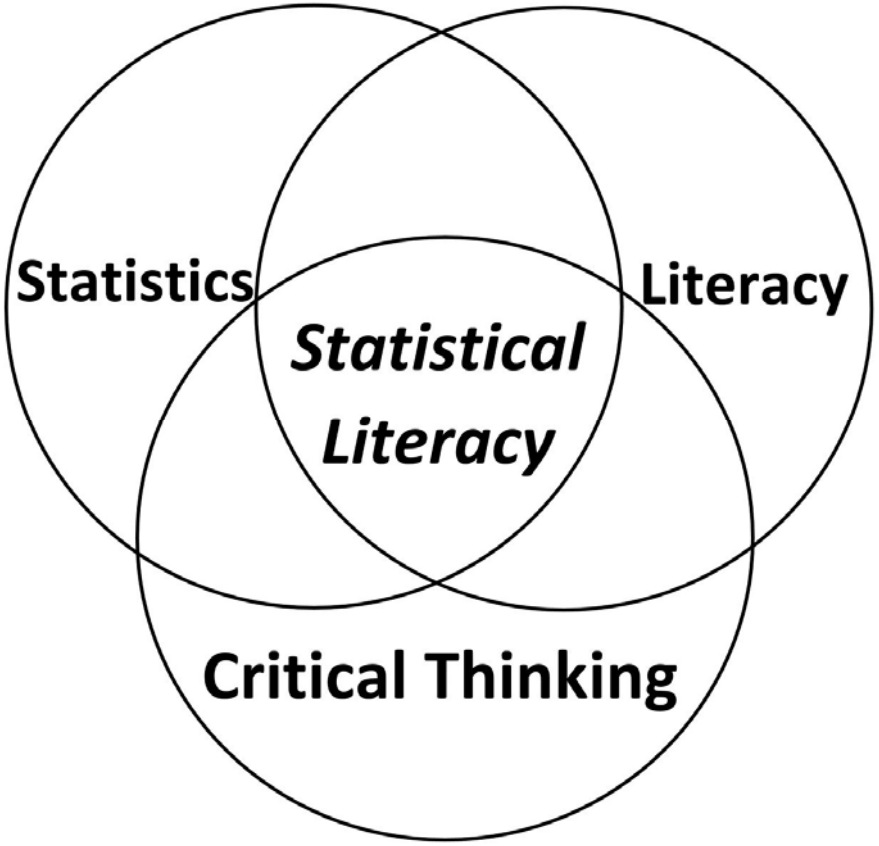
MATH 1300 (3)

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.

Meets New Mexico General Education Curriculum Area 2: Mathematics and Statistics.

Math1300: Statistical Literacy

Less than 30% overlap with traditional statistics



Math1300: Statistical Literacy

STATISTICS STUDIES VARIATION

Two kinds of variation

SYSTEMATIC	RANDOM
Confounding (control for: selection, ratios, regression)	**Probability**
Assembly (define, present)	**Sampling error**
Error (Bias)	**Confidence intervals**
	Test of Hypothesis
Systematic Influences	*Statistical Inference*

Math1300: Statistical Literacy

TWO KINDS OF INTRODUCTORY STATISTICS COURSES

Statistical Literacy. *More confounding than statistical inference.*	**Quantitative Methods.** *More statistical inference than confounding.*
UNM Math 1300	**Utts, Bennett, Stat 101**

By what standard is the content selected?

Math1300: Statistical Literacy

Statistical literacy is designed for the consumers of statistics: students, citizens, and decision makers.

- Students in non-quantitative majors.

But this statement of audience is still ambiguous.

Who decides what the content will be?

- Statisticians: GAISE 2016 update?
- The statistics in research articles?
- The statistics in the everyday media?

Math1300: Statistical Literacy Statistics in the everyday Media

This is what these students "NEED".

Statistical content analysis:

- hundreds of news stories containing statistics
- 46,000 articles in the Harvard Business Review
- 4.5 billion words in UK Cobuild corpus
- 1 billion words BYU corpus (COCA)

Statistical Literacy: Social Statistics

Statistics are different from numbers

Statistics are numbers in context (in reality)

Statistics can be influenced by reality:

- In arithmetic, 1 plus 1 is always 2.
- In reality math:
  1 bunny plus 1 bunny can give three bunnies
  1 ice-cube plus 1 ice-cube can give zero ice-cubes

Statistics are *Socially Constructed*

Joel Best, author of "Lies, Damned Lies and Statistics" identified this as fact as *the most important, the most fundamental* aspect of *all* reality-based statistics.

He didn't mean all of reality was a mental construct.

He meant that statistics, just like words, are created by people: people with motives, values and goals.

Each statistic embodies a particular view of reality.

Statistics can Be Influenced

Q. Best advice when dealing with statistics?

A. "Take CARE". Statistics can be influenced.

All influences are grouped into four categories:

C: Confounding

A: Assembly (how things are defined, counted, etc.)

R: Randomness

E: Error (including bias)

Admonition: "Take CARE"

Students like "CARE". It gives them a structure.

Without a unifying structure, statistical literacy is just a collection of disparate influences.

Everything they learn about influences fits.

When asked to rank what they considered most valuable, they chose "CARE".

If they remember just one thing from the course, it should be "Take CARE".

Seven Student Failures: Failure to recognize that…

1. **Association is not causation**
2. Statistics can be manipulated
3. Statistics can be confounded: Simpson's paradox
4. Bigger data, the more likely an unlikely statistic
5. Ratio statistics are ordered: confusion of the inverse
6. Spotty statistics are modeled
7. Statistical significance can be influenced

Association is Not Causation

Semantics: Association is not [necessarily] 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

Disparity is Not Discrimination

Semantics: Differences or Disparities are not [necessarily] Discrimination

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

Seven Student Failures: Failure to recognize that…

1. Association is not causation (eg., disparity…)
2. **Statistics can be manipulated**
3. Statistics can be confounded: Simpson's paradox
4. Bigger data, the more likely an unlikely statistic
5. Ratio statistics are ordered: confusion of the inverse
6. Spotty statistics are modeled
7. Statistical significance can be influenced

Statistics can be Manipulated
Assembly/Assumptions

The number of children killed by gunfire has doubled **each year** since 1950;

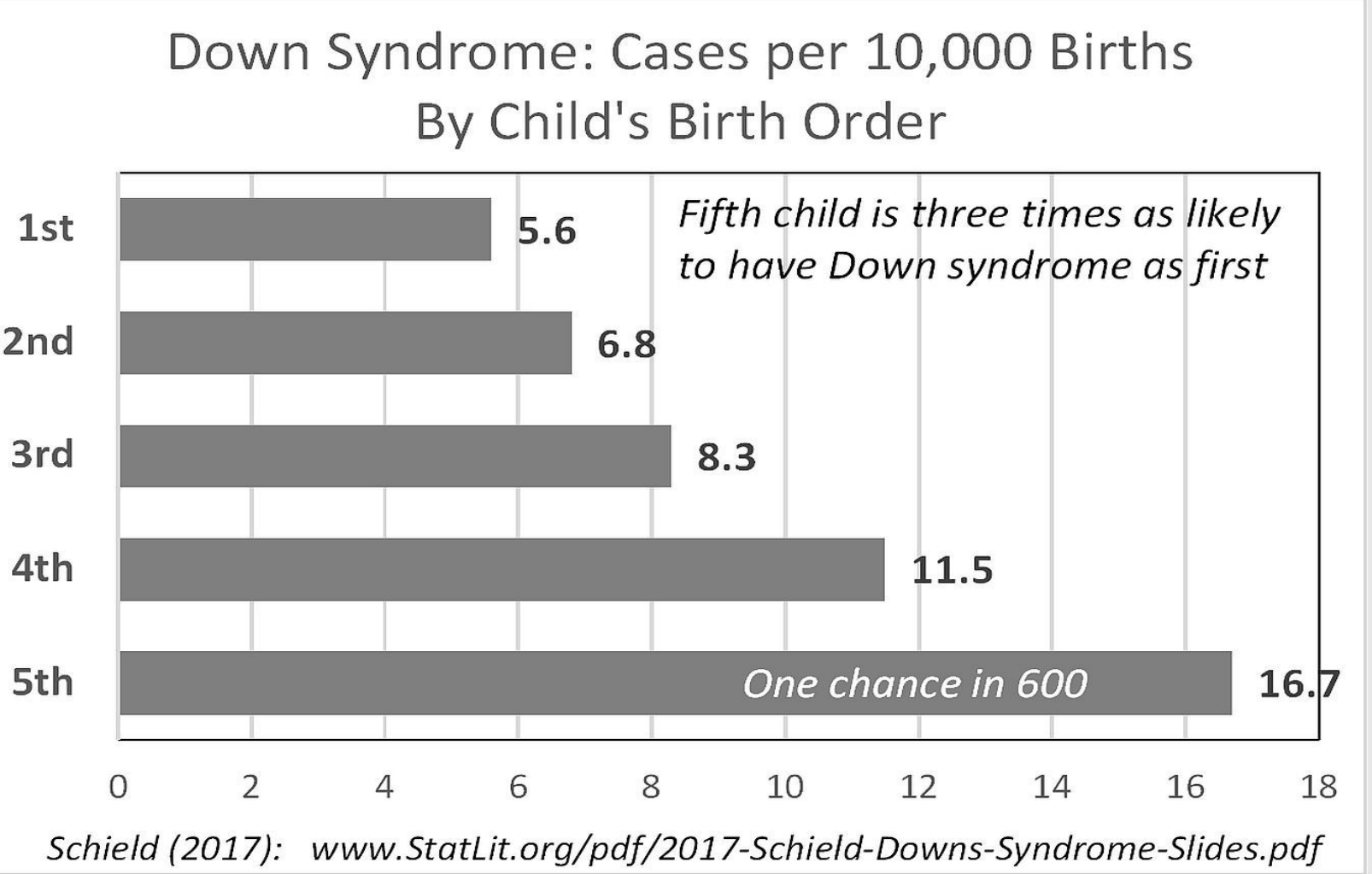
The number of children killed by gunfire **each year** has doubled since 1950

30% of middle school students are **bullied**.
Define bullying. Increase or decrease the percentage

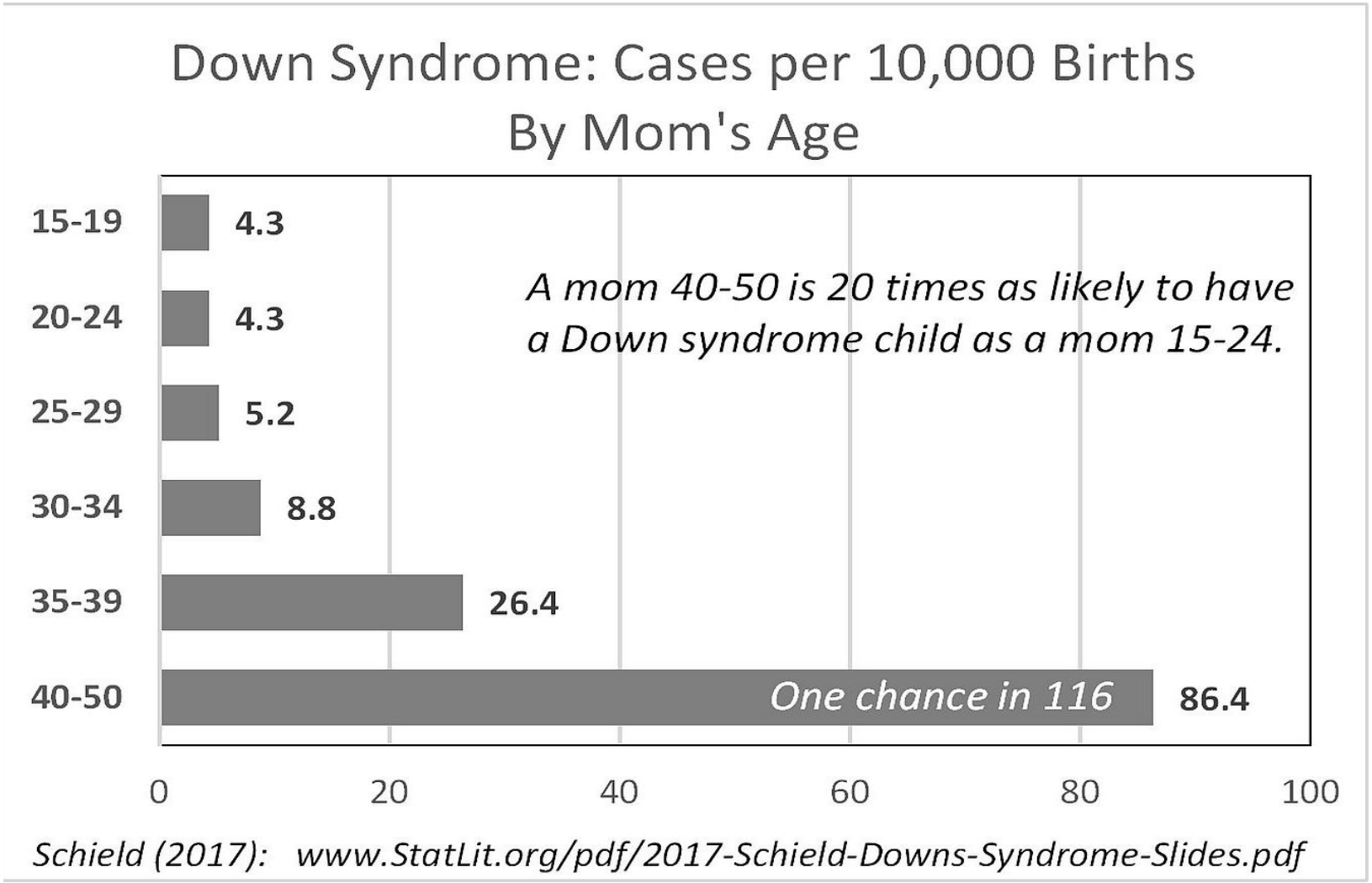
Seven Student Failures: Failure to recognize that…

1. Association is not causation
2. Statistics can be manipulated
3. **Statistics can be confounded: Simpson's paradox**
4. Bigger data, the more likely an unlikely statistic
5. Ratio statistics are ordered: confusion of the inverse
6. Spotty statistics are modeled
7. Statistical significance can be influenced

Statistics can be Confounded
Downs Syndrome



Statistics can be Confounded Down Syndrome



Schield (2017): www.StatLit.org/pdf/2017-Schield-Downs-Syndrome-Slides.pdf

Statistics can be Confounded
Vaccinated: More Likely to Die

In the UK, vaccinated cases are 2.4 times as likely to die from Covid as [are] the unvaccinated cases.

National Health Service. 268,169 cases.

A crude comparison (mixed fruit comparison).

What could confound this association?

Age!

Statistics can be Confounded
Vaccinated: More Likely to Die

UK 260,000 cases May-Sept 2021

Covid Death Rates per Case			**Crude**	Fraction of Cases			**Adjusted**
	<50	50+	All	<50	50+	All	Standard
Un-vac	0.03%	5.96%	**0.17%**	0.977	0.023	1.000	**0.71%**
Vaccinated	0.02%	1.68%	**0.41%**	0.767	0.233	1.000	**0.21%**
0.17% = 0.977*0.03% + 0.023*5.96%				0.885	0.115	1.000	
0.41% = 0.767*0.02% + 0.233*1.68%				**0.21%**=.885*.02%+.115*1.68%			
		Ratio	2.4	Ratio	10.2	Ratio	3.4

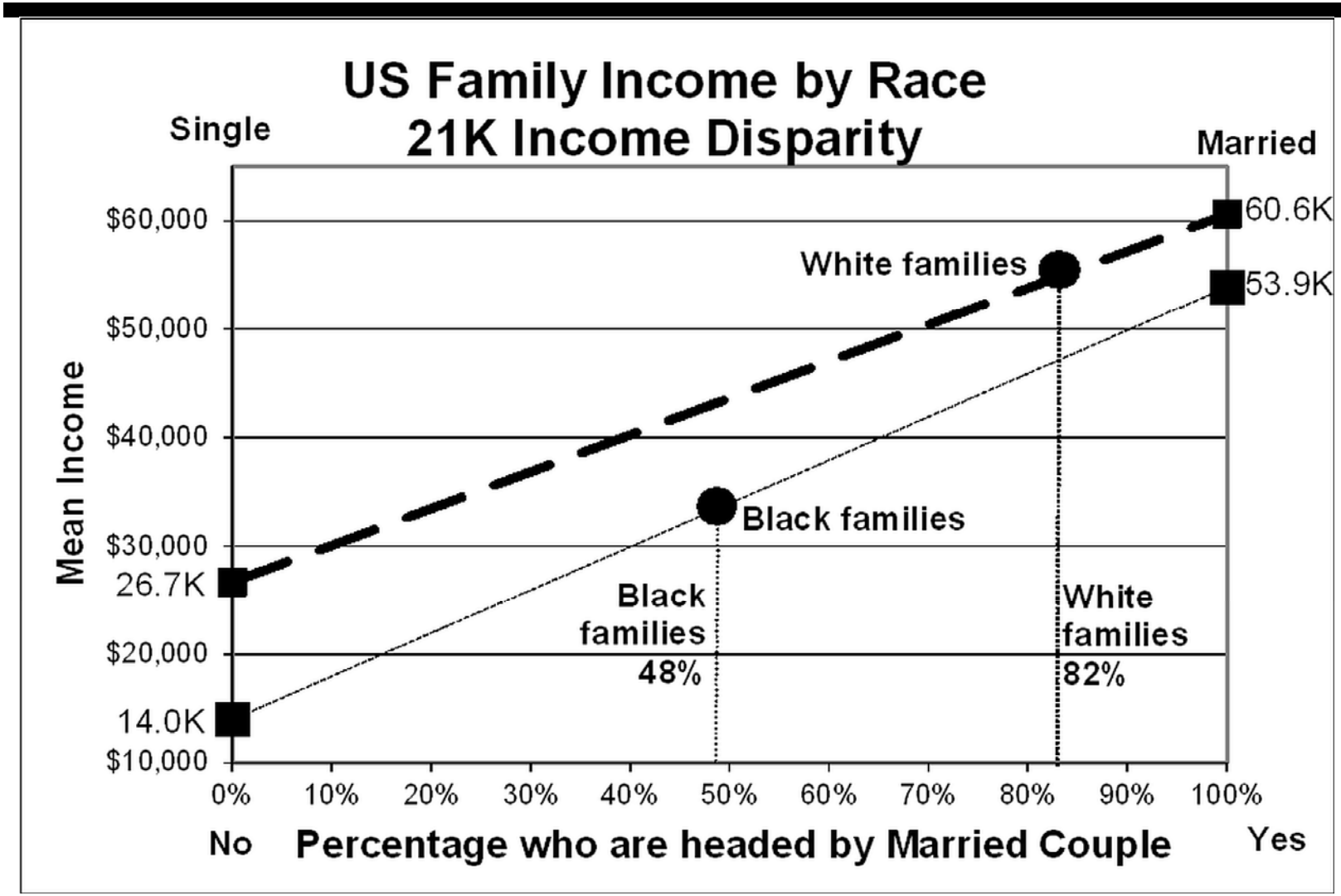
After standardizing, unvaccinated are more likely to die than [are] the vaccinated.

Black-White Income Gap
22K Income Disparity

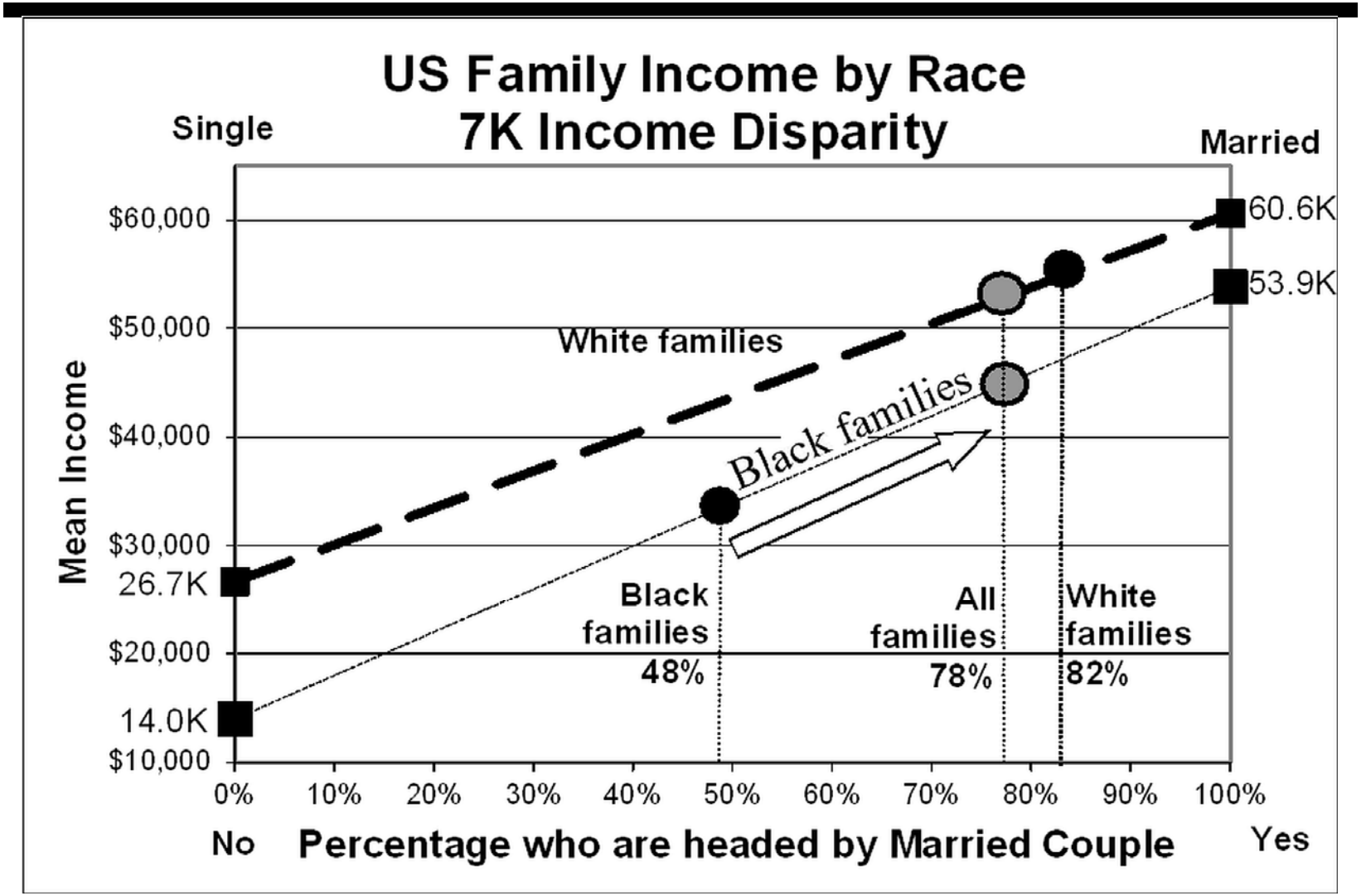
Family Income			
Total	Race	Single	Married
$55,000	Whites	$26,700	$60,600
$33,000	Blacks	$14,000	$53,900

Distribution	
Single	Married
18%	82%
52%	48%

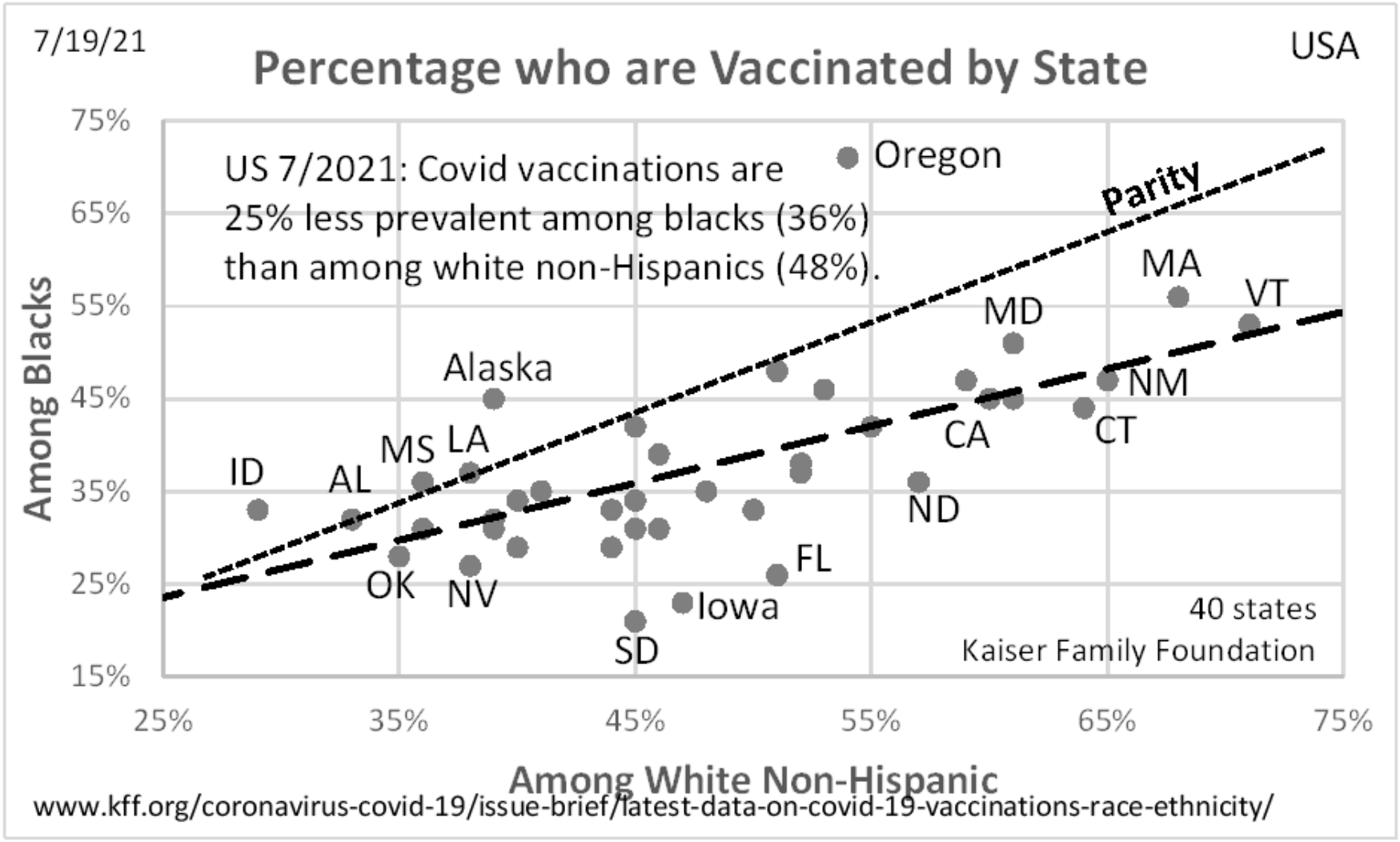
Standardization Graphically: 21K Income Disparity



Standardization Graphically: 7K Income Disparity



Blacks less likely to be Vaccinated than White NH



Students Need Practice with Confounding

Understand how a crude association can be a:

- Mixed fruit comparison
- An apples and oranges comparison.

Students need to really understand what it means:

- To take something into account
- To control for something.
- To standardize an association:
  To give both groups the same mix!

Seven Student Failures: Failure to recognize that…

1. Association is not causation
2. Statistics can be manipulated
3. Statistics can be confounded: Simpson's paradox
4. **Bigger data, the more likely an unlikely statistic**
5. Ratio statistics are ordered: confusion of the inverse
6. Spotty statistics are modeled
7. Statistical significance can be influenced

Randomness and Big Data

The bigger the data, the more likely a rare event.

The Law of Truly Large Numbers.

Suppose an unlikely event occurs one chance in N.

Given N tries,

- one such event is *expected*, and
- at least one such event is more likely than not.

Schield (2009).

The bigger the data, the more likely an unlikely outcome

Seven Student Failures: Failure to recognize that…

1. Association is not causation
2. Statistics can be manipulated
3. Statistics can be confounded: Simpson's paradox
4. Bigger data, the more likely an unlikely statistic
5. **Ratio statistics are ordered: confusion of the inverse**
6. Spotty statistics are modeled
7. Statistical significance can be influenced

Distinguish Numerator and Denominator

1. The percentage of smokers who are men.
2. The percentage of men who are smokers.
3. The percentage of men among smokers.

Confusion of the inverse

1. Guys are more likely to smoke than gals.
2. Guys are more likely among smokers than gals.

Distinguish Percent and Percentage Grammar

If "20% of guys are smokers",
then 20% is *the percentage of guys who smoke*.

So if 20% of guys who run are smokers,
then 20% is
the percentage of guys who run who are smokers.

This last phrase is ambiguous!

What is the status of *run*?

Seven Student Failures: Failure to recognize that…

1. Association is not causation
2. Statistics can be manipulated
3. Statistics can be confounded: Simpson's paradox
4. Bigger data, the more likely an unlikely statistic
5. Ratio statistics are ordered: confusion of the inverse
6. **Spotty statistics are modeled**
7. Statistical significance can be influenced

Distinguish Counted Counts from Modeled Counts

Failure #6

Responsible for …

Study blames diesel for deaths

By Jon Brodkin / Daily News Staff
Wednesday, February 23, 2005

Diesel pollution is responsible for more deaths than drunk drivers and homicides, according to a new study that estimates how many premature deaths, asthma attacks and heart attacks are caused by diesel pollution in every U.S. county.

Nationwide, diesel pollution causes 21,000 premature deaths each year, including 475 in Massachusetts and 81 in Middlesex County, robbing those who die of an average of 14 years of their lives.

2. Epidemiological statistics are common!
US Annual Deaths **Attributable To**:

Smoking: 467,000	Blood pressure: 395,000
Overweight: 216,000	Inactivity: 191,000
Blood sugar: 190,000	LDL cholesterol: 113,000
Dietary salt: 102,000	Low omega-3 : 84,000

High dietary trans fatty acids: 82,000

Alcohol use: 64,000 (90,000 less 26,000 averted)

Low intake of fruits and vegetables: 58,000

Low poly-unsaturated fatty acids: 15,000

www.emaxhealth.com/2/24/30740/
smoking-high-blood-pressure-obesity-top-preventable-death-causes.html

Distinguish Counted Counts from Modeled Counts

Counted deaths are coroner certified.
Some causal deaths are coroner certified: alcohol.

Modeled deaths are statistical deaths.
Smoking, second-hand smoke, obesity, etc.
These counts are speculative (spotty) statistics.

Schield (2009): Confound those Speculative Statistics
www.StatLit.org/pdf/2009SchieldASA.pdf
Schield (2011): Epidemiological Models and Spotty Statistics
www.StatLit.org/pdf/2011SchieldISI.pdf

Epidemiological statistics encourage seductive grammar

Consider these titles of news stories:

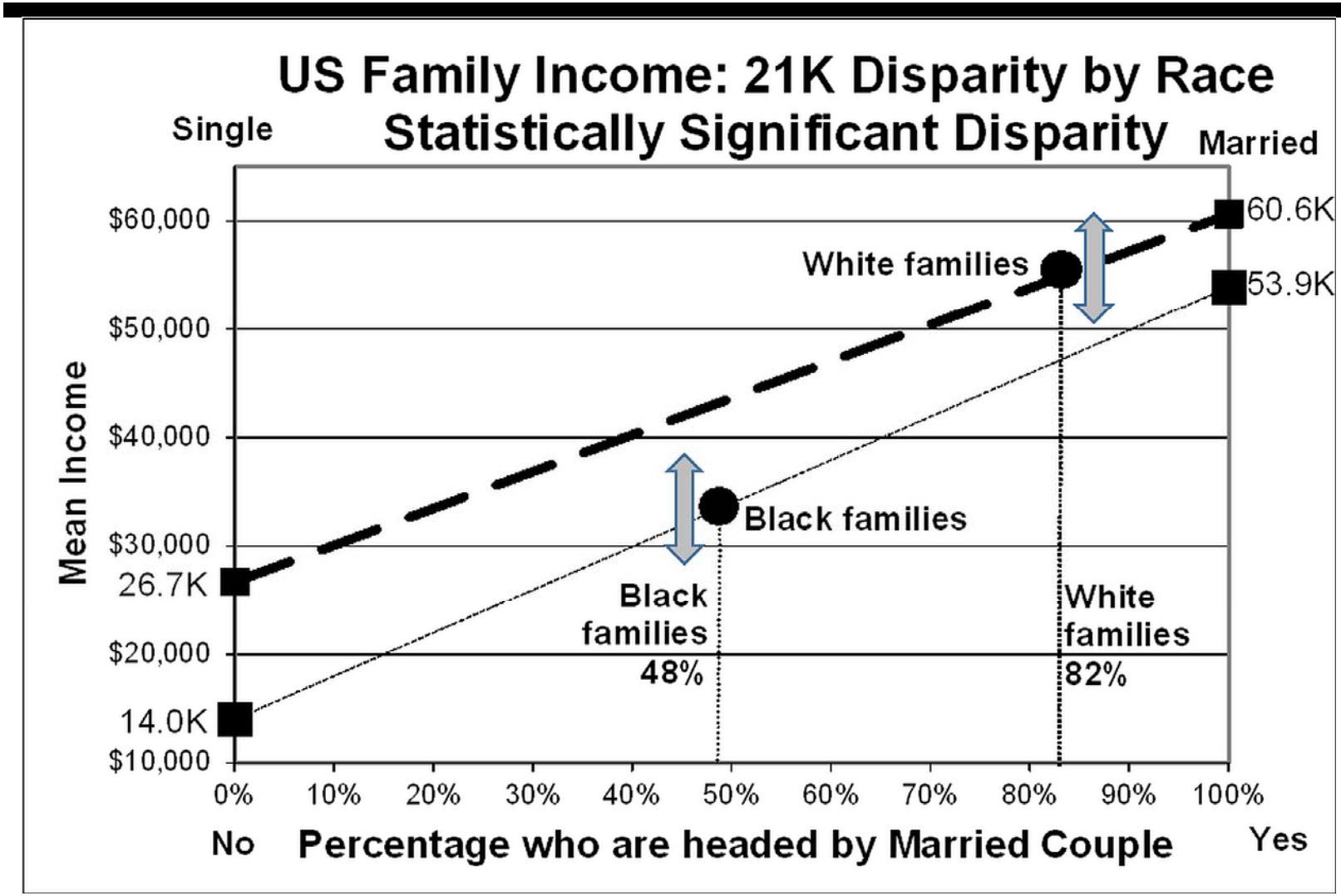
- 45,000 deaths *attributable to* uninsurance
- 45,000 deaths *associated with* lack of insurance
- Lack of insurance *linked to* 45,000 deaths
- 45,000 die … *because of* lack of health insurance
- Lack of Health Insurance *Kills* 45,000 a Year
- Lack of Health Insurance *cause* 44789 deaths
- Lack of insurance *to blame* for almost 45,000 deaths

Source: www.StatLit.org/pdf/2010SchieldICOTS.pdf

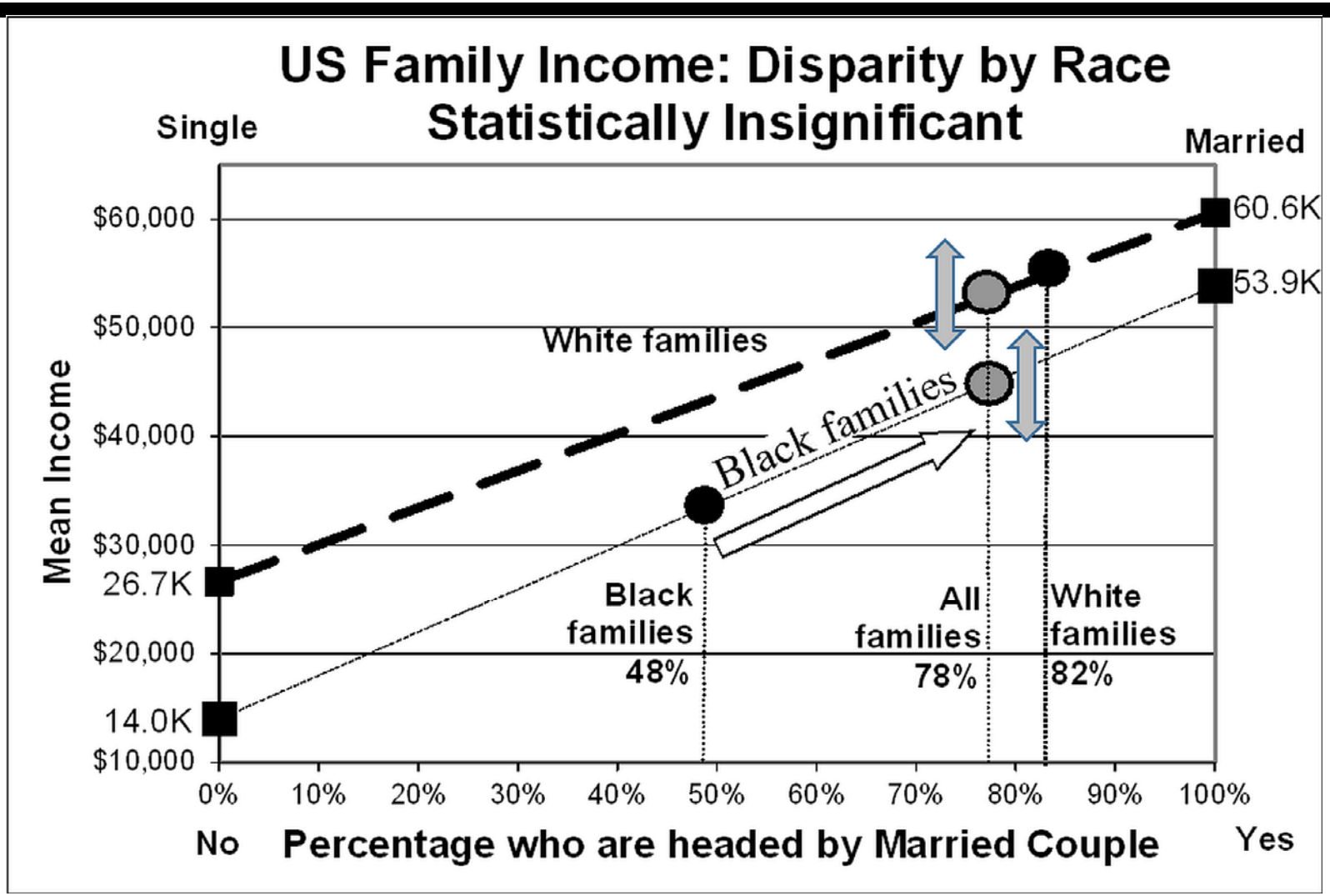
Seven Student Failures: Failure to recognize that…

1. Association is not causation
2. Statistics can be manipulated
3. Statistics can be confounded: Simpson's paradox
4. Bigger data, the more likely an unlikely statistic
5. Ratio statistics are ordered: confusion of the inverse
6. Spotty statistics are modeled
7. **Statistical significance can be influenced**

Statistical significance: Influenced by Confounding



Statistical Insignificance: Influenced by Confounding



Result:
Students should:

Know that statistics can be influenced.

Understand “control for” and “take into account.”

Know that standardizing converts a mixed fruit comparison into an apples and apples comparison.

Recognize the possibility of a Simpson’s paradox.

Math1300 Highlights

Asserts that *Association is Not Causation*
Asserts that *Disparity is Not Discrimination*

Focus on *The Story Behind the Statistics*

Shows how a *crude association* (mixed fruit comparison) may conceal the real story!

Shows students how to *control for* confounders

Shows students these things *without computers*

Anonymous Student Survey

Before finals, students are asked three questions.

1. Did this course improve your critical thinking?
2. Would you recommend this course to a friend?
3. Should all students be required to take this course?

When I teach traditional statistics I get 'Yes' (Agree or Strongly agree) from 15% to 25% of the students.

When I teach Statistical Literacy to art, music and management majors, I get 'Yes' from at least 50%.

Anonymous Student Comments (Fall 2021)

I like the content and critical thinking aspect of the class. As someone who had to drop the regular stats class, I was very happy to have this class as an option.

This course is an answer to my prayers, I am a music major and horrible at math so fulfilling my math requirement has been hard. This is the first math class I actually liked. … the material is about things I can apply to everyday life. …I would recommend this class for anyone.

Study Confounder-Based Statistical Literacy

Statistical Literacy: What Students Like about the Course
www.statlit.org/pdf/2021-Fall-UNM-MATH1300-S1.pdf

Statistical Literacy: The Diabolical Denominator
www.StatLit.org/pdf/2021-Schield-MathFest.pdf

Statistical Literacy: Teaching Confounding
www.StatLit.org/pdf/2021-Schield-USCOTS.pdf

University of New Mexico Offers Math 1300
www.StatLit.org/pdf/2021-Schield-ASA.pdf

Schield's papers: www.StatLit.org/Schield-Pubs.htm