Statistics Literacy For Decision Makers

Chapter 3: Measurements

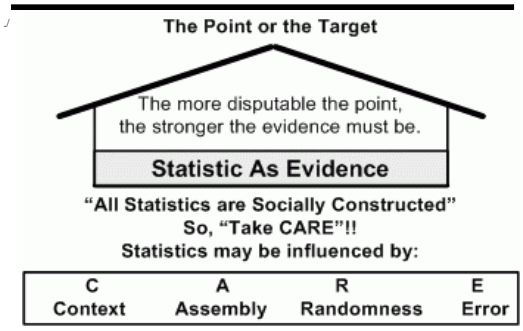
by
Milo Schield

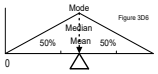
Half-Day Workshop
USCOTS May 16, 2019
www.StatLit.org/pdf/2019-Schield-USCOTS-Slides3.pdf

Measurements: Chapter 3 Outline

Distributions
Measures of center
Two-group comparisons of Means & Medians
Two-variable co-variation
Spread
Slope and simple regression

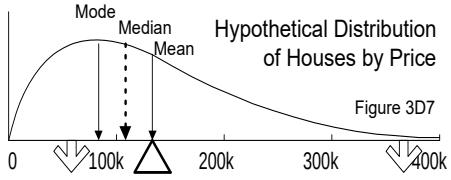
Stat Literacy: Study Statistics as Evidence in Arguments





Measures of Center

In an asymmetric distribution, mean, median and mode typically align alphabetically with mean most sensitive to extremes. Why?



Mean, median, mode: Alphabetically. Why?

Suppose that house prices in your town have a positive near-symmetric distribution

Suppose Bill and Melinda Gates move to your town. They built two Mac-Mansions.

How does that change the mode, median and mean of the original distribution?

Mode? Median? Mean?

Most relevant in the short run? In the long-run?

Issues:

1. Mean is more sensitive to outliers.
   Yet statisticians prefer the mean. Why?
2. Omit measure: *City1 income more than City2.*
3. Omit characteristic: *Midtown is a median city.*
4. Assume the mean exists. *1.8 kids per family.*
5. Ambiguity in specifying the group

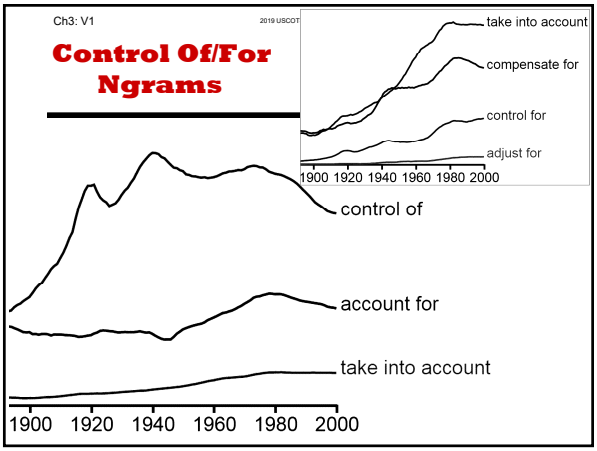
Controlling Confounding: Control Of

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

Controlling Confounding: Control For

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

Control Of/For Ngrams



Crude Associations

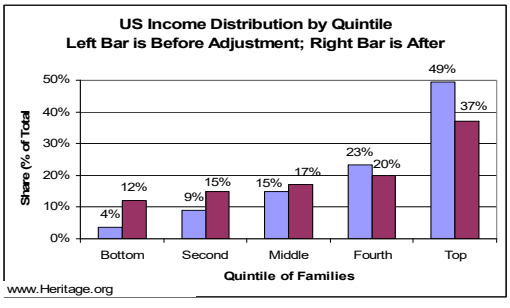
A **crude association** is an association in which nothing else has been taken into account.

Less likely to get pregnant:

- Short young adults than tall.
- Adults that shave daily than those that don't
- Adults with long hair than those with short.

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

Crude Association versus an Adjusted Association



Prison Expense: Crude vs Adjusted Associations

State	Total	# Inmates	Per Inmate
CA	$2.9B	136K	$21,385
NY	$1.9B	69K	$28,426

Total ↑	Per Inmate ↓
50% more	25% less

State	Total	# Inmates	Per Inmate
MD	$481M	21,623	$22,245
KS	$159M	7,148	$22,245

Total ↑	Per Inmate ↔
3 times..	Same

State	Total	# Inmates	Per Inmate
MN	$184M	4,865	$37,825
ME	$48M	1,424	$33,711

Total ↑	Per Inmate ↑
260% more	12% more

Crude Ratio Associations
It's the Mix!!!

Ratio associations can be still be confounded. Averages are ratios.

NAEP Math 8	Internet Access at Home		
State	All	Yes	No
Virginia (VA)	▲ 275	▼ 282	▼ 258
Texas (TX)	▲ 273	▼ 285	▼ 260

NAEP Math 8	Internet Access at home		
State	All	Yes	No
Virginia (VA)	▲ 275 (100%)	▲ 282 (69%)	▼ 258 (31%)
Texas (TX)	▲ 273 (100%)	▲ 285 (53%)	▼ 260 (47%)

Simpson's Paradox: Time
It's the Mix!!

SAT Verbal flat, but every group improved.

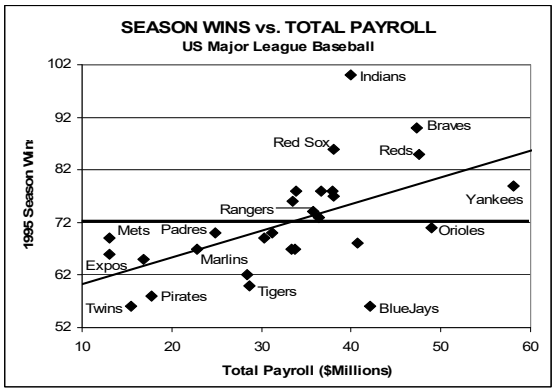
SAT-Verbal	---- Scores ----			---- Distribution ----		
Group	1981	2002*	Chg	1981	2002*	Points
White	519	527	+8	85%	65%	-20
Black	412	431	+19	9%	11%	+2
Asian	474	501	+27	3%	10%	+7
Mexican	438	446	+8	2%	4%	+2
Puerto Rican	437	455	+18	1%	3%	+2
Amer. Indian	471	479	+8	0%	1%	+1
ALL	504	504	0			

Will an Association Reverse?
The Cornfield Conditions

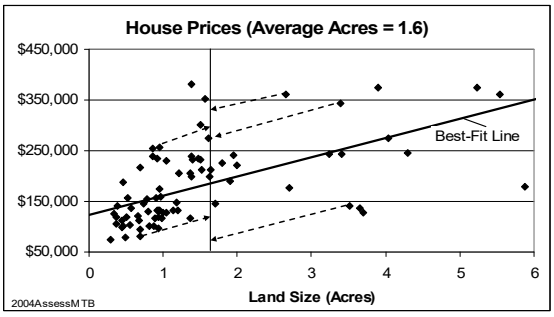
After learning about Simpson's Paradox, one student said, "I'll never trust another statistic." This is cynicism: not a good outcome.

Not all confounders can reverse an association. Jerome Cornfield proved that a confounder association must be "bigger" than the observed.

Cornfield's conditions are one of the three biggest contributions of statistics to human knowledge.



Regression Standardizes



Regression Standardizes
An Example:

The data shows that house prices increase by $39,000 per bedroom. This is a crude association.

$16,000 per bedroom if land is *controlled for*,

$9,000 per bedroom after *accounting for* land and house size,

$5,000 after *adjusting for* land, house size, and number of bathrooms.

TV for toddlers interferes with brain growth, says study:

Children under two should not be allowed to watch television because it increases their chances of suffering attention problems later in life, says an American study.

A study of 1,345 children found that each hour spent in front of the set every day increased the risks of attention deficit disorders by 10%.

U.S. journal, *Pediatrics*

Time to Double given Growth Rate

If a child's risk of Attention Deficit Disorder increases by 10% for every extra hour of watching TV, how many hours do they have to watch to double their risk?

Rule of 72*: Time to double = 72 / Rate

72 divided by 10% per hour = 7.2 hours

* Assuming compounding

How to Relate this to Math Colleagues

Don't talk about confounding or effect size.

Talk about assumptions.

- What one controls for is an assumption.
- What one fails to control for is an assumption.

AAU&C Quantitative Literacy VALUE rubric:

Assumptions: Ability to make and evaluate important assumptions in estimation, modeling, and data analysis.

AAC&U Quantitative Literacy VALUE Rubric

Interpretation, Representation, Calculation, Application, **Assumptions**, and Communication

Assumptions: Ability to **make and evaluate** important assumptions in estimation, modeling, and data analysis.

www.statlit.org/pdf/2009QuantitativeLiteracyRubricAACU.pdf
www.aacu.org/peerreivew/2014/summer/RealityCheck

Statistics Literacy For Decision Makers

Chapter 3: Measurements

by
Milo Schield

Half-Day Workshop
USCOTS May 16, 2019
www.StatLit.org/pdf/2019-Schield-USCOTS-Slides3.pdf

Measurements: Chapter 3 Outline

Distributions

Measures of center

Two-group comparisons of Means & Medians

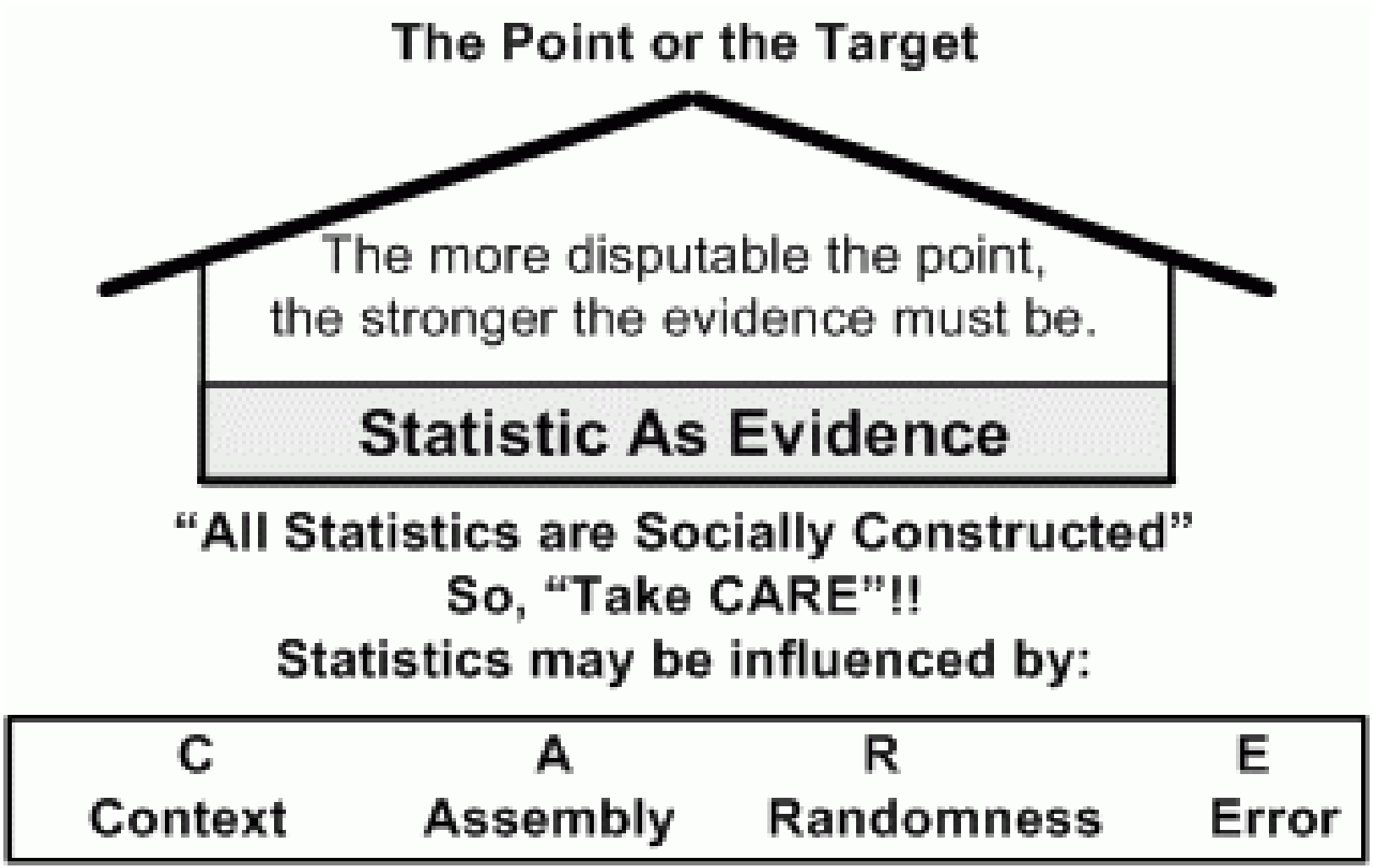
Two-variable co-variation

Spread

Slope and simple regression

Stat Literacy: Study Statistics as Evidence in Arguments

./



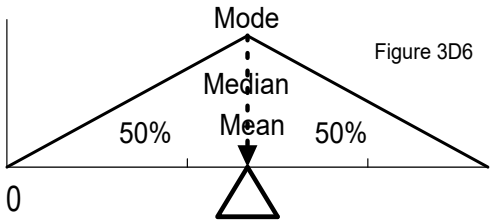


Figure 3D6

Measures of Center

In an asymmetric distribution, mean, median and mode typically align alphabetically with mean most sensitive to extremes. Why?

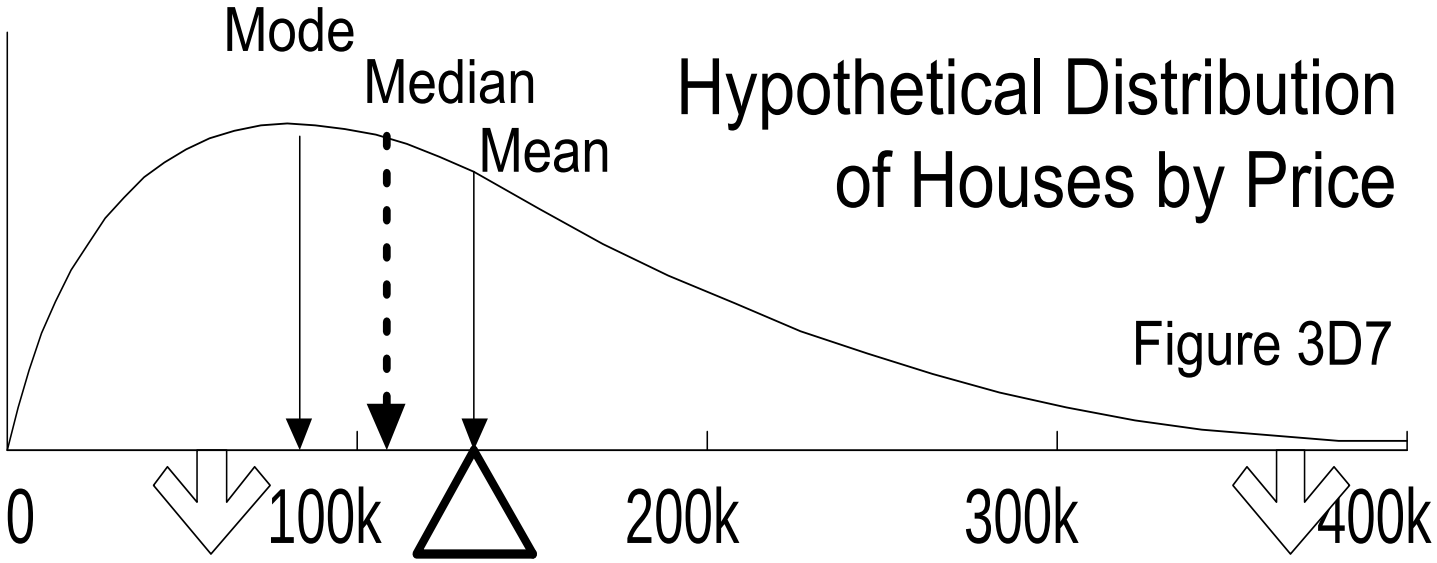


Figure 3D7

Mean, median, mode: Alphabetically. Why?

Suppose that house prices in your town have a positive near-symmetric distribution

Suppose Bill and Melinda Gates move to your town. They built two Mac-Mansions.

How does that change the mode, median and mean of the original distribution?

Mode? Median? Mean?

Most relevant in the short run? In the long-run?

Issues:

1. Mean is more sensitive to outliers.
   Yet statisticians prefer the mean. Why?
2. Omit measure: *City1 income more than City2.*
3. Omit characteristic: *Midtown is a median city.*
4. Assume the mean exists. *1.8 kids per family.*
5. Ambiguity in specifying the group

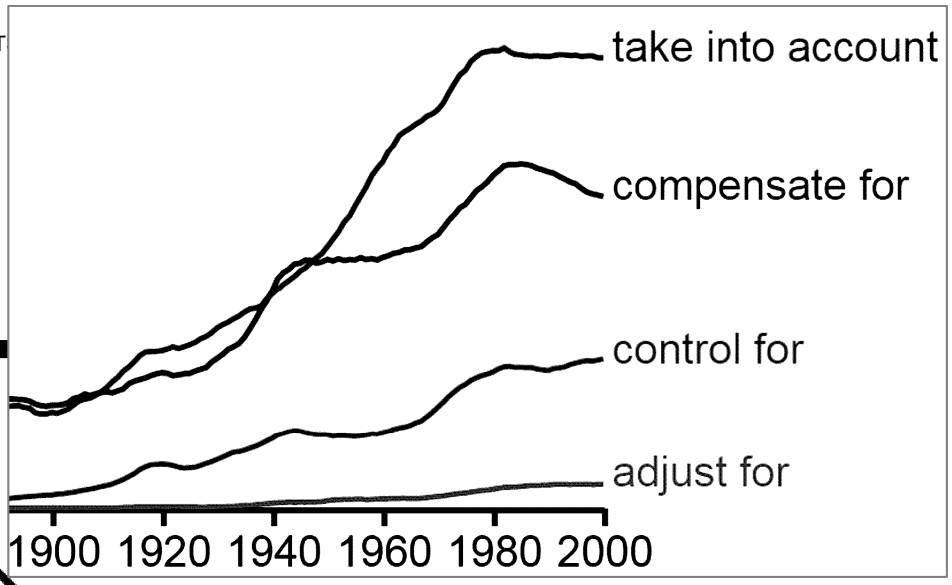
Controlling Confounding: Control Of

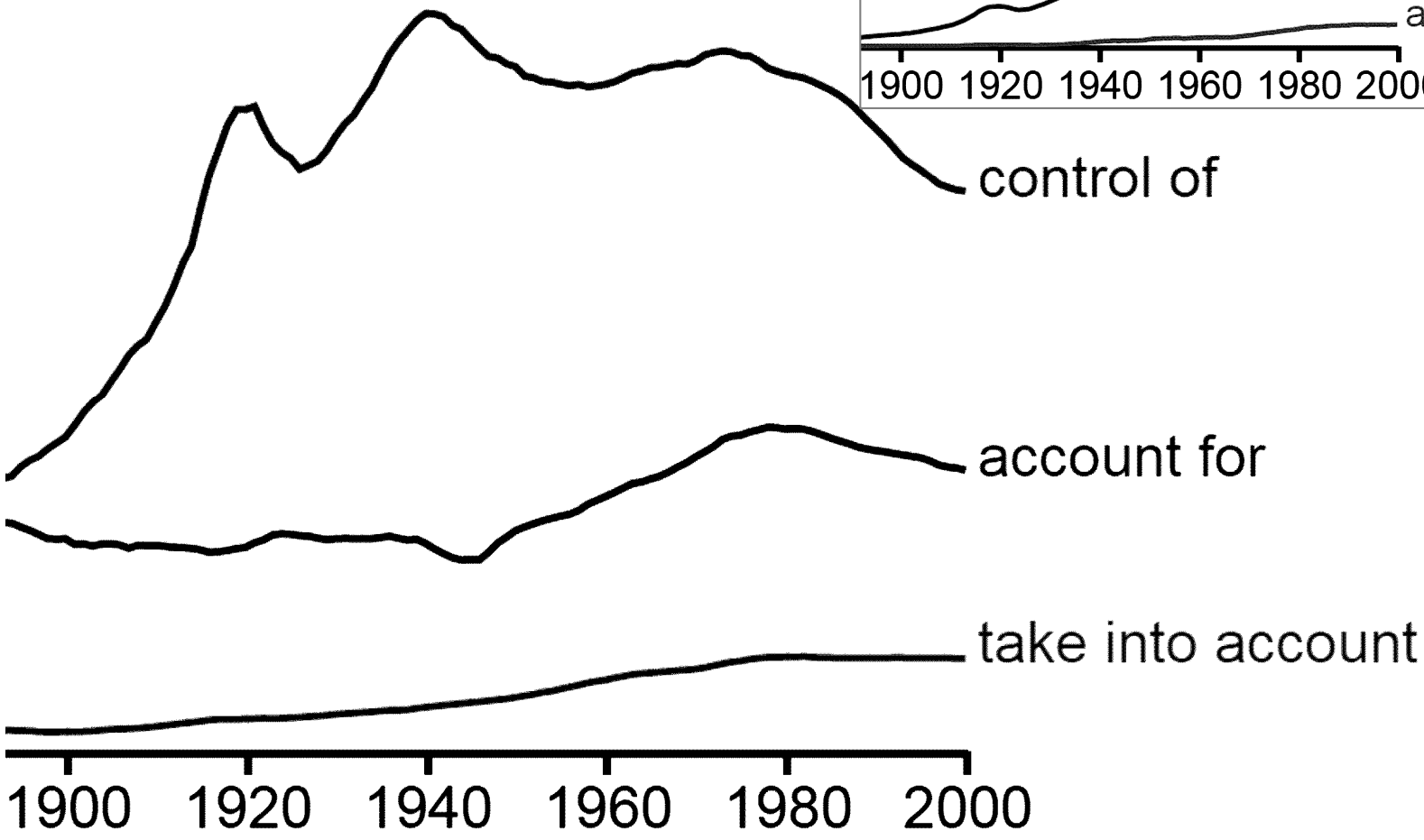
CONTROL OF CONFOUNDERS			
Physical Control (Grade = Quality)			
Experiment			Observational Study
A+	Scientific	C	Longitudinal
A-	Random Assign	D	Cross-sectional
B	Quasi-Exper	F	Anecdotal story

Controlling Confounding: Control For

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

Control Of/For Ngrams





Crude Associations

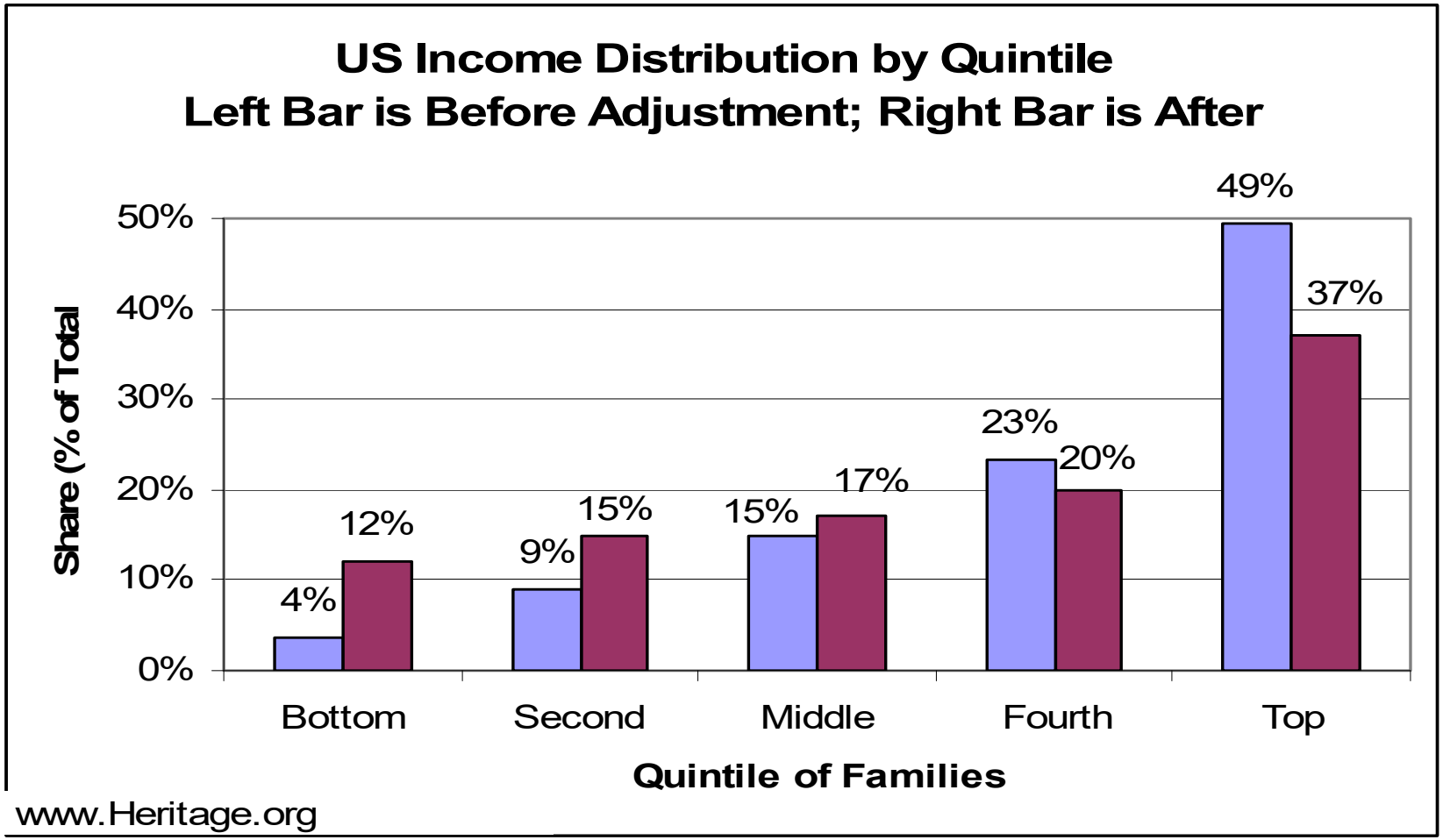
A **crude association** is an association in which nothing else has been taken into account.

Less likely to get pregnant:

- Short young adults than tall.
- Adults that shave daily than those that don't
- Adults with long hair than those with short.

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

Crude Association versus an Adjusted Association



Prison Expense: Crude vs Adjusted Associations

State	Total	# Inmates	Per Inmate
CA	$2.9B	136K	$21,385
NY	$1.9B	69K	$28,426

Total	Per Inmate
50% more	25% less

State	Total	# Inmates	Per Inmate
MD	$481M	21,623	$22,245
KS	$159M	7,148	$22,245

Total	Per Inmate
3 times..	Same

State	Total	# Inmates	Per Inmate
MN	$184M	4,865	$37,825
ME	$48M	1,424	$33,711

Total	Per Inmate
260% more	12% more

Crude Ratio Associations
It's the Mix!!!

Ratio associations can be still be confounded.
Averages are ratios.

NAEP Math 8	Internet Access at Home		
State	All	Yes	No
Virginia (VA)	275	282	258
Texas (TX)	273	285	260

NAEP Math 8	Internet Access at home		
State	All	Yes	No
Virginia (VA)	275 (100%)	282 (69%)	258 (31%)
Texas (TX)	273 (100%)	285 (53%)	260 (47%)

Simpson's Paradox: Time
It's the Mix!!

SAT Verbal flat, but every group improved.

/

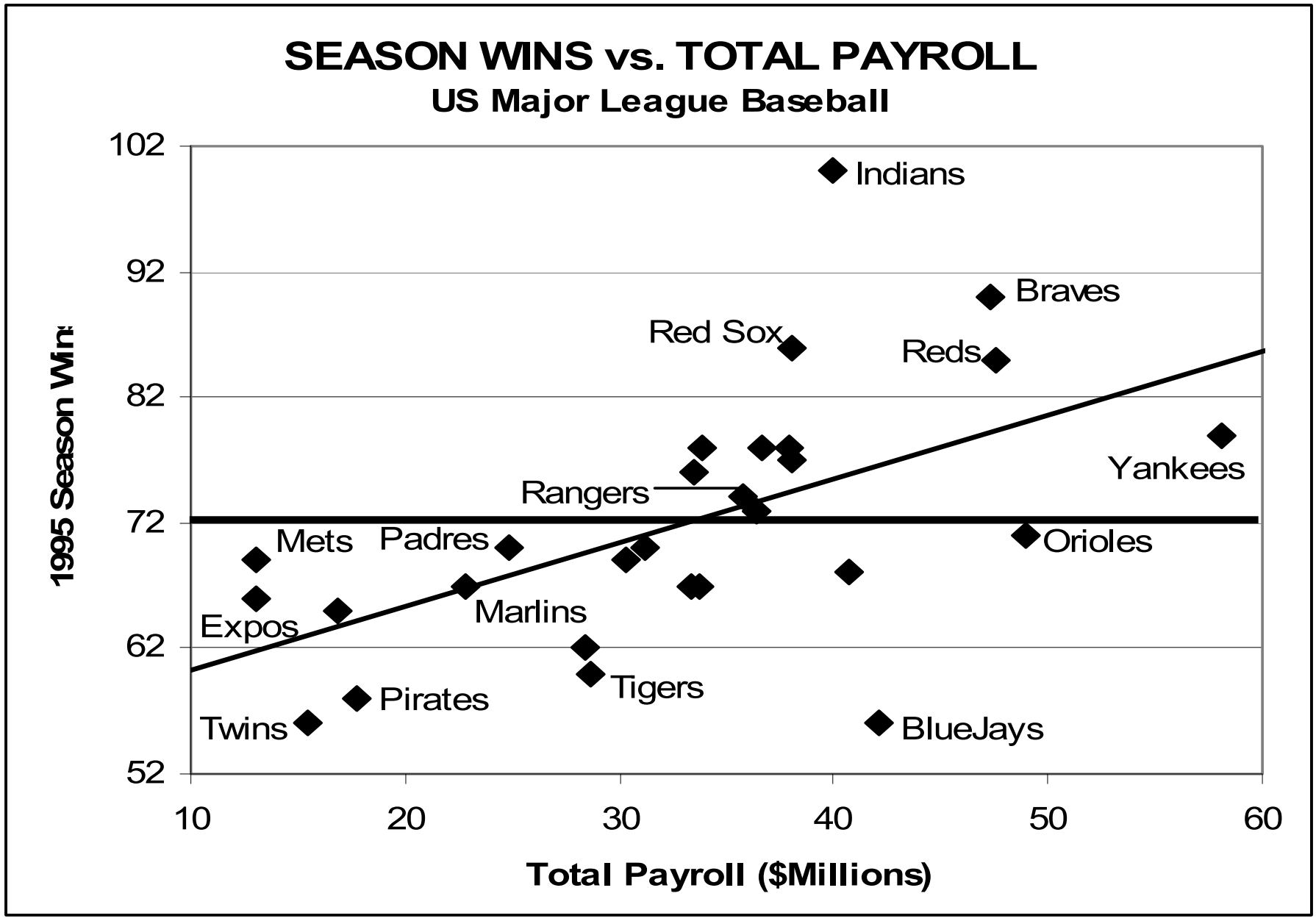
SAT-Verbal	---- Scores ----			---- Distribution ----		
Group	1981	2002*	Chg	1981	2002*	Points
White	519	527	+8	85%	65%	-20
Black	412	431	+19	9%	11%	+2
Asian	474	501	+27	3%	10%	+7
Mexican	438	446	+8	2%	4%	+2
Puerto Rican	437	455	+18	1%	3%	+2
Amer. Indian	471	479	+8	0%	1%	+1
ALL	504	504	0			

Will an Association Reverse? The Cornfield Conditions

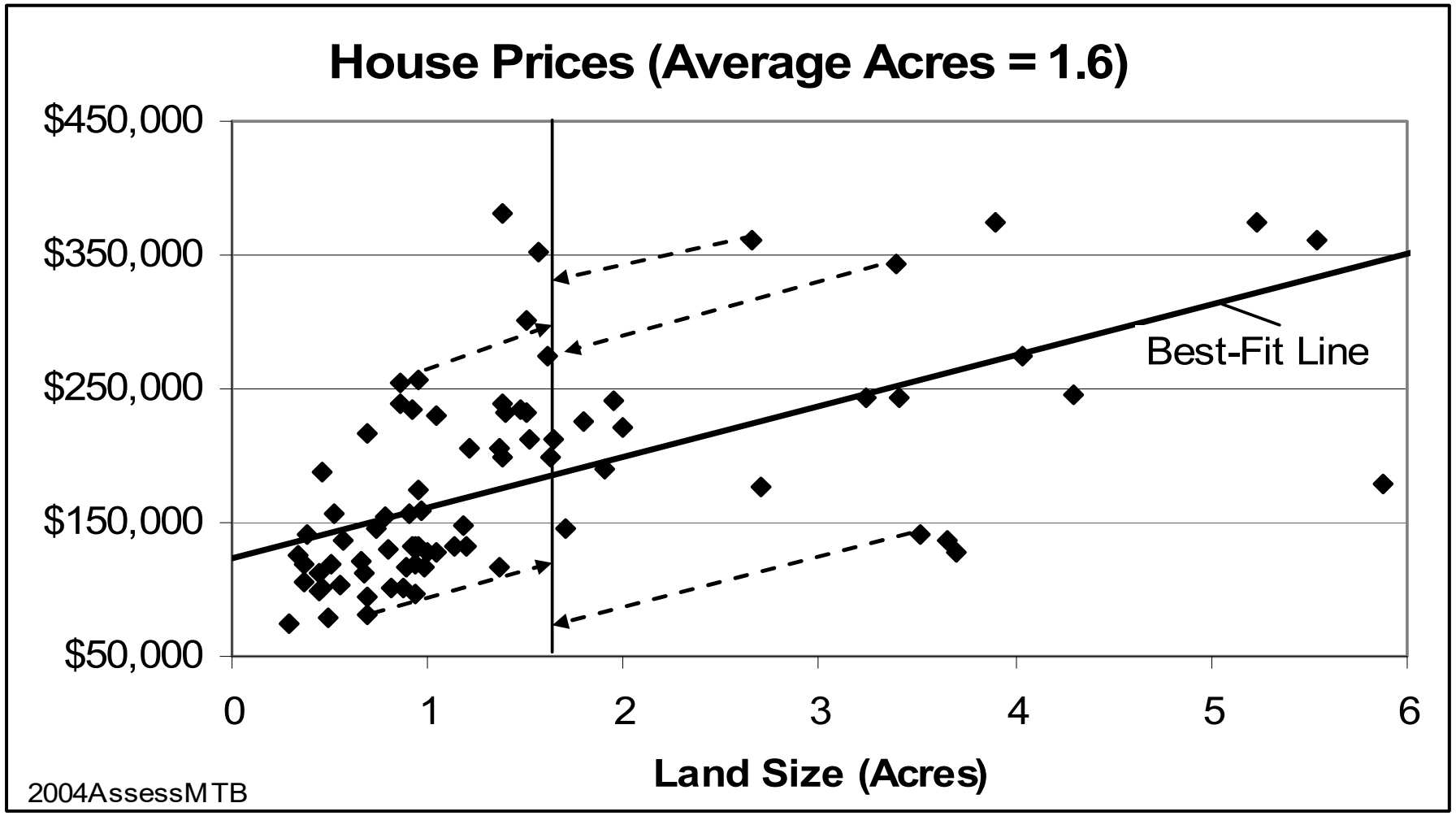
After learning about Simpson's Paradox, one student said, "I'll never trust another statistic." This is cynicism: not a good outcome.

Not all confounders can reverse an association. Jerome Cornfield proved that a confounder association must be "bigger" than the observed.

Cornfield's conditions are one of the three biggest contributions of statistics to human knowledge.


SEASON WINS vs. TOTAL PAYROLL
US Major League Baseball
1995 Season Wins
102
92
82
72
62
52
10
20
30
40
50
60
Total Payroll ($Millions)
Indians
Braves
Red Sox
Reds
Yankees
Rangers
Mets
Padres
Orioles
Expos
Marlins
Tigers
Pirates
Twins
BlueJays

Regression Standardizes



Regression Standardizes An Example:

The data shows that house prices increase by $39,000 per bedroom. This is a crude association.

$16,000 per bedroom if land is *controlled for*,

$9,000 per bedroom after *accounting for* land and house size,

$5,000 after *adjusting for* land, house size, and number of bathrooms.

TV for toddlers interferes with brain growth, says study:

Children under two should not be allowed to watch television because it increases their chances of suffering attention problems later in life, says an American study.

A study of 1,345 children found that each hour spent in front of the set every day increased the risks of attention deficit disorders by 10%.

U.S. journal, *Pediatrics*

Time to Double given Growth Rate

If a child's risk of Attention Deficit Disorder increases by 10% for every extra hour of watching TV, how many hours do they have to watch to double their risk?

Rule of 72*: Time to double = 72 / Rate

72 divided by 10% per hour = 7.2 hours

* Assuming compounding

How to Relate this to Math Colleagues

Don't talk about confounding or effect size.

Talk about assumptions.

- What one controls for is an assumption.
- What one fails to control for is an assumption.

AAU&C Quantitative Literacy VALUE rubric:

Assumptions: Ability to make and evaluate important assumptions in estimation, modeling, and data analysis.

AAC&U Quantitative Literacy VALUE Rubric

Interpretation, Representation, Calculation, Application, **Assumptions**, and Communication

Assumptions: Ability to **make and evaluate** important assumptions in estimation, modeling, and data analysis.

www.statlit.org/pdf/2009QuantitativeLiteracyRubricAACU.pdf
www.aacu.org/peerreivew/2014/summer/RealityCheck