Statistical Literacy: IDSL Keynote

Milo Schield
ASA Fellow
Visiting Professor: New College of Florida
Adjunct Researcher: Univ of New Mexico

Nov 25, 2025

International Day of Statistical Literacy

International Statistical Literacy Project

www.StatLit.org/pdf/
2025-Schield-IDSL-StatLit-Keynote-Slides.pdf

Strong Support for Statistical Literacy

Walker (1951). *Statistical Literacy in the Social Sciences*

Haack (1979). *Teaching Statistical Literacy*

Moore (1998). statistical ideas will educated people require in the 21st century?

Garfield (1999) The understanding of statistical language: words, symbols, and terms.

Snell (1999). the ability to understand statistical concepts and reason at the most basic level"

Strong Support for Statistical Literacy

GAISE (2005). define statistical literacy as understanding the basic language of statistics…

Weak Support for Statistical Literacy

Rumsey (2002). I will omit the phrase "statistical literacy" from my discussion…

Borers (2006) argued that "their existence is not dictated by empirical observations" so they are "unsuitable as learning goals for statistics education."

Everson (2015). concern that not everyone may understand what is meant by 'statistical literacy' and 'statistical thinking'.

Strong 'Support' for Statistical Literacy

Schield (2017). "There is evidence – abundant evidence – in the same document that the GAISE 2016 revision has actually accepted – and encouraged – statistical literacy. Their official recommendation has done more to advance the cause of statistical literacy than any previous paper, book, group or committee. Statistical literacy may have lost a battle as a phrase, but it may be winning the war as an idea."

Strong 'Support' for Statistical Literacy

Schield (2017). "Those reading the new GAISE report (GAISE, 2016) may not realize how big of a change this revision is. Multivariate thinking and confounding were not even mentioned in the earlier report (GAISE, 2005), neither in McKenzie's (2004) list of core concepts, nor in articles on statistical literacy by Watson (2003) and Gal (2002)."

Three Problems Defining Statistical Literacy

1. Problem defining two-word phrases. (Schield, 2017)
2. Problem defining ‘statistics’ (this talk)
3. Problem on whether to include confounding. (this talk)

2. 'Statistics' is Ambiguous

Statistics-Education		Statistical Literacy
Survey data	**Data**	Everyday # in context
Sample survey data	**Statistics**	Everyday stats in context
Context: outlier & bias		Context: all influences
Random variation		Systematic variation
Confidence Intervals		Simpson's paradox
Statistical Significance		Cornfield's conditions

3. ‘Confounding’ is Absent or Minimal

2021 Gal Civic Statistics. Correlate (1)

2019 Gal Understanding Statistical Literacy.
Adjusted (8), multivariable (4), correlate (1)

2018 [3] Understanding Civic Statistics
Multivariate (7), aggregate (6), confound (3)

2002 Gal Adult's Statistical Literacy

2002 Rumsey Statistical Literacy as a Goal

3. 'Confounding' is Absent or Minimal

In examining 100 intro statistics textbooks,

- 80 do not include confounding in the index
- 15 show one example (burglaries & ice cream)
- 5 show multiple examples.

None of them show how to work a problem without a computer and without algebra.

None of them show how statistical significance can be influenced.

3. 'Confounding': Comparison is Simpler

Most statisticians are trained in mathematics, where continuous data is the norm.

Two-group discrete is more common in the humanities and in many of the social sciences.

Adjusting for confouding		
Two-factor co-variation		Two-group comparison
Multivariate analysis		Standardization
Requires software		No software or algebra rqrd.

How Statistical Significance can be Influenced

Most introductory statistics textbooks present statistical significance as a ‘fact’ that is etched in stone regardless of the kind of study.

As teachers, we know that statistical significance can be influenced in an observational study.

But we don’t show it.

Yet most of our students taking statistics will be dealing with observational studies.

Here is how it can be shown easily.

Statistical Significance: The Idea (very abstract)

Statistically significant: statistically unlikely. The less likely the chance, the more 'significant' the result.

Examples:

- Dog that didn't bark. Sherlock Holmes, Silver Blaze
- Goldfinger meets Bond 3 times, 3 different places
- Five heads in five flips of a fair coin.

Statistically significant:

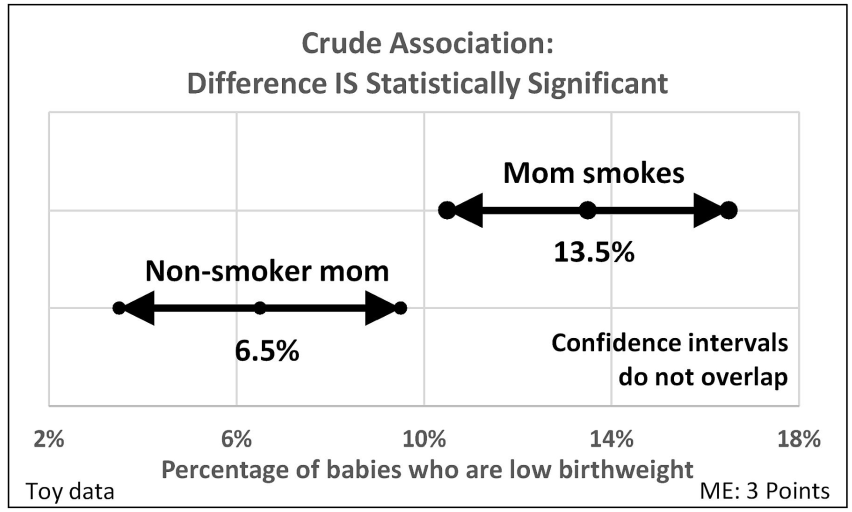
- Not necessarily important
- is arbitrary (< 1 chance in 20) but not unreasonable
- **can be influenced**

Statistical Significance: Influenced by Confounding

Crude Association:
Difference is stat. signif.

Young: 50% of smokers
10% of non-smokers

Adjust for mom's age:

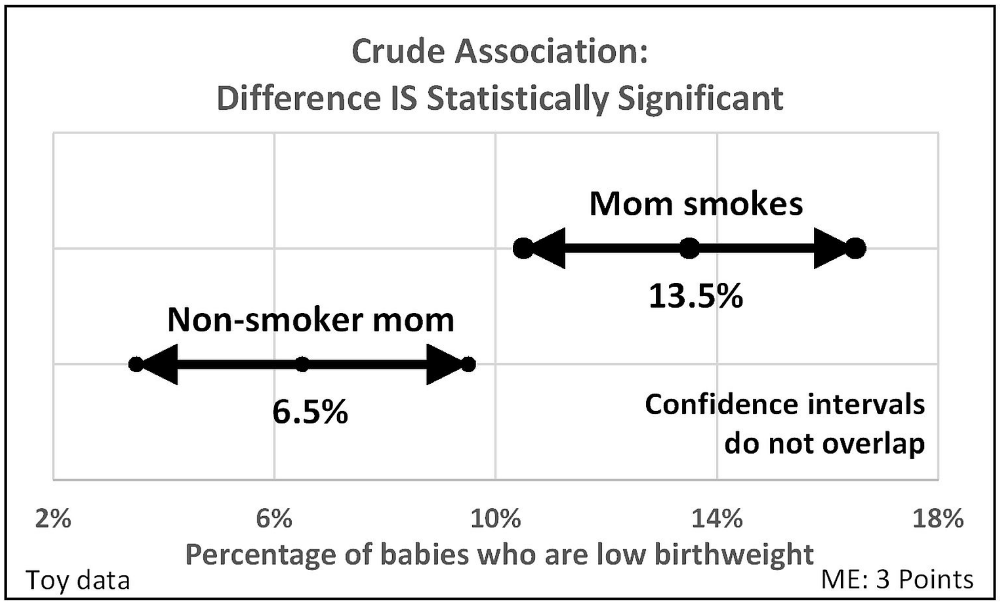


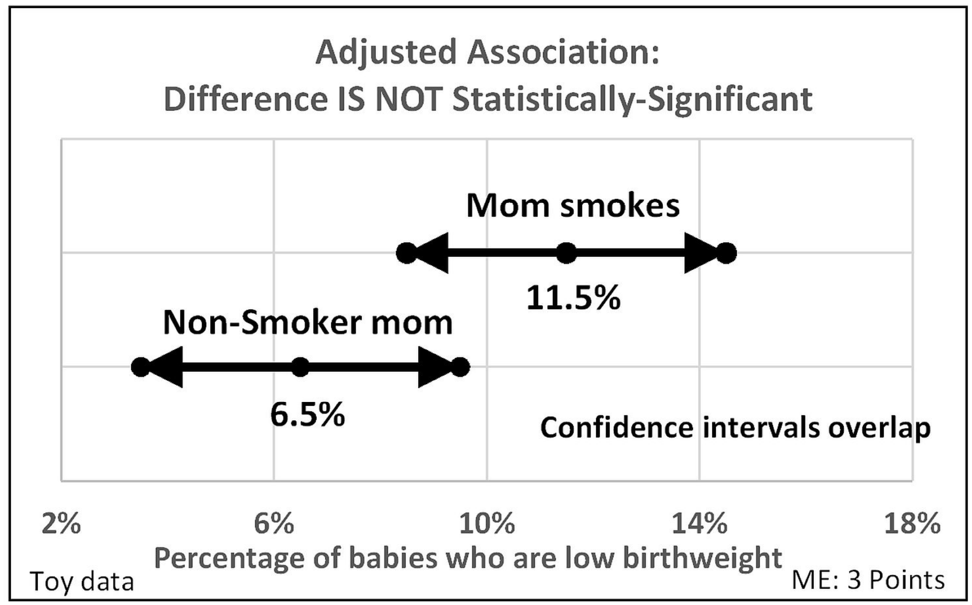
A	B	C	D	1	F	G	H	I
Low-weight birth rates: crude				2	**Low-weight birth rates: mix-match**			
Moms	Teens	Older	ALL	3	Moms	Teens	Older	ALL
Non-Smokers	11%	6%	**6.5%**	4	Non-Smokers	11%	6%	6.5%
Smokers	16%	11%	**13.5%**	5	Smokers	16%	11%	**11.5%**
DISTRIBUTION (weights):				6	DISTRIBUTION (weights):			
Non-Smokers	10%	90%	100%	7	Non-Smokers	10%	90%	100%
Smokers	50%	50%	100%	8	Smokers	**10%**	**90%**	100%
D4=B4*B7 + C4*C7	=0.11*0.10 + 0.06*0.90				I4	Same as D4		
D5=B5*B8 + C5*C8	=0.16*0.50 + 0.11*0.50				**I5=G5*G8 + H5*H8**	**=0.16*0.10 + 0.11*0.90**		

Statistical Significance: Influenced by Confounding

Crude Association: Difference statistically significant

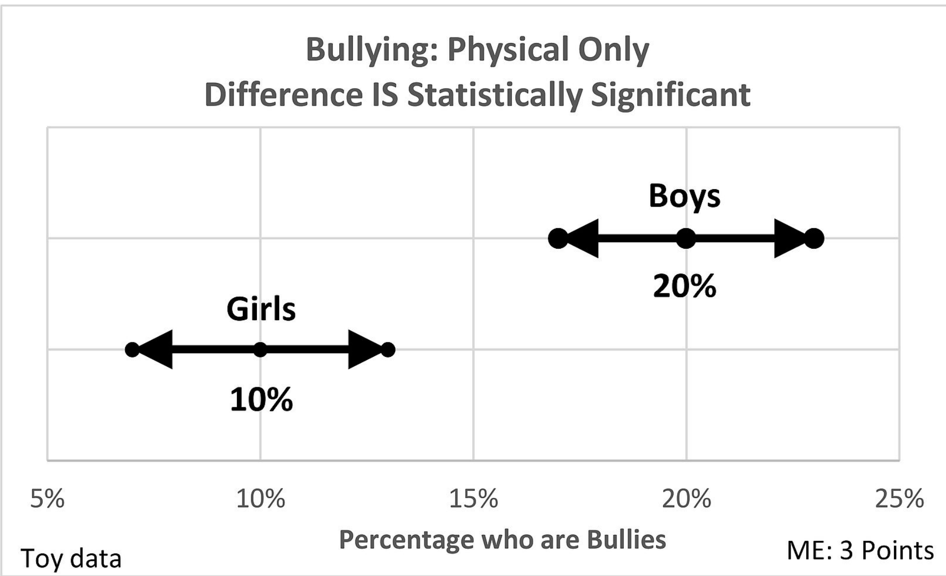
Adjusted Comparison (same mix of ages):
Difference is not statistically significant.

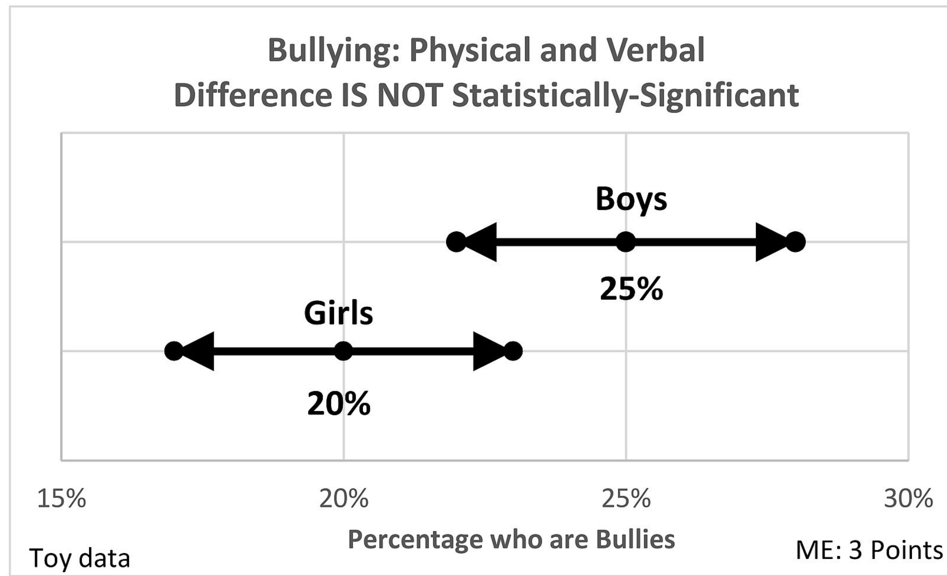




Statistical Significance: Influenced by Assembly

Bullying #1: physical only. 10% of girls. 20% of boys.





Bullying #2: physical & verbal. Girls 20%; boys 25%.
Change in definition (assembly) changes
statistically-significant into statistically insignificant

Statistical Significance: Influenced by Randomness

Clinical trial: difference is statistically significant.

If the observed difference is totally spurious (bogus), how could statistical significance be expected?

1. One trial repeated 20 times.
2. 20 outcomes in single trial.
3. 20 researchers each: one trial and one outcome.

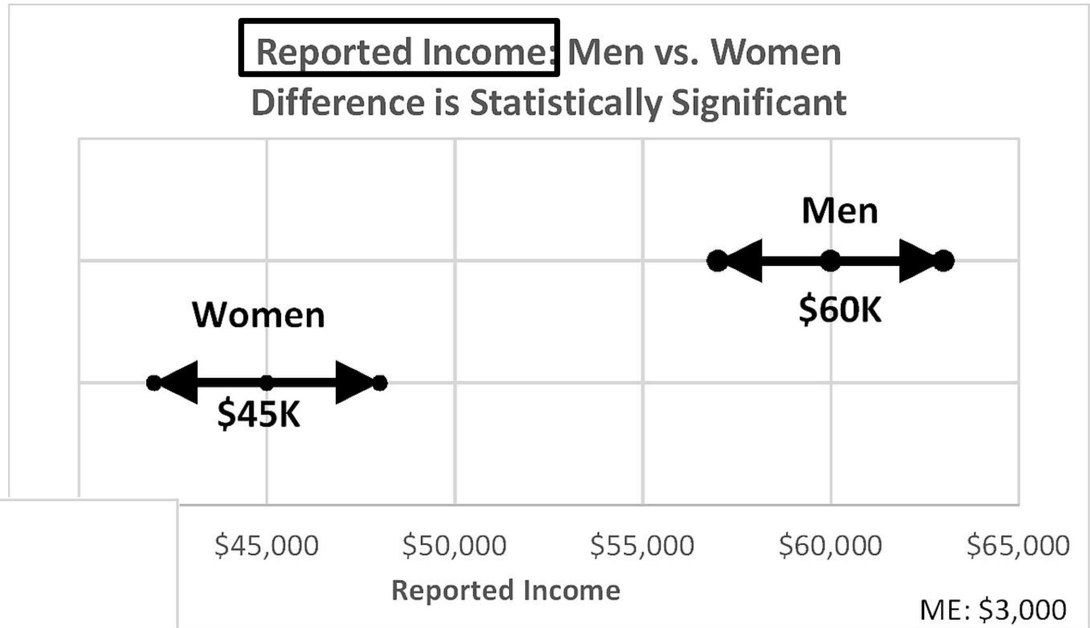
Also, one researcher does 20 different trials each with one outcome.

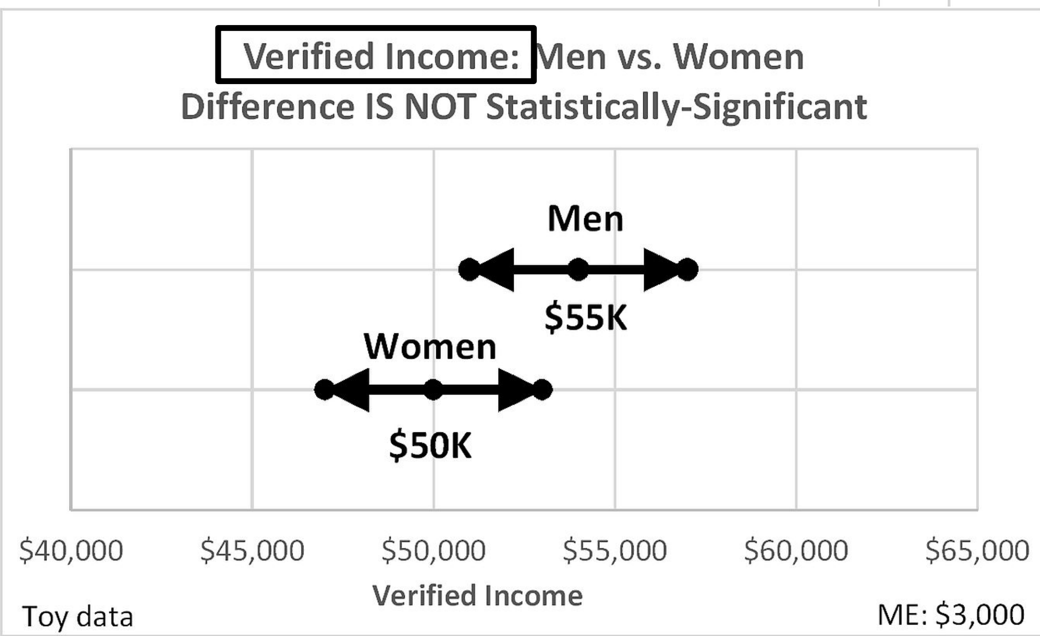
Statistical Significance: Influenced by Error/bias

Income by Sex:

Reported vs. verified.

Subject bias!





Chance: Common Misunderstandings

1. If two outcomes, then 50% chance of either. Maybe.
2. If 90% chance of rain and no rain, forecast was wrong. Maybe. Can't tell from a single event.
3. If excess from expected, offset is predicted. No. Dilution is slower but maintains independence.
4. If medical test is 99% accurate and result is positive, then 99% chance that patient has disease. Maybe.
5. The more unlikely the outcome, the more likely it is causal. Maybe. Unlikely sometime, somewhere or unlikely then, there?

Hypothesis Tests and Criminal Trials

Criminal trial: Innocent until proved guilty
Statistical trial: No difference until proved different

Type 1: false alarm. Type 2: silent alarm

--------------- Criminal Trial --------------		
	Defendant (Reality)	
Jury verdict	**INNOCENT**	Guilty
Not-guilty	OK	Type 2
Guilty BRD	**Type 1**	OK
Type 1 error: Innocent person found guilty		
Type 2 error: Guilty person found not-guilty		
Minimize Type 1 error. Presume innocent.		

-------------- Hypothesis Test -------------		
	Claim (Reality)	
Statistical result	**FALSE**	True
Stat. Insignificant	OK	Type 2
Stat. Significant	**Type 1**	OK
Type 1: Statistically significant result is false		
Type 2: Statistically-insignificant result is true		
Minimize Type 1 error. Presume false		

Interpreting Statistical Significance Bayesian Medical Tests (>50%)

Shift from confirmation (column %) to prediction (row %).

--------------- Medical Test --------------			
People	*Has the diesease (Reality)*		
Test result	**No**	Yes	TOTAL
Negative	# OK	# Type 2	Sum(Row1)
Positive	**# Type 1**	# OK	Sum(Row2)
TOTAL	Sum(No)	Sum(Yes)	SUM
Type 1 error: Disease-free person tests positive			
Type 2 error: Diseased person tests negative			

-------------- Hypothesis Test --------------			
Samples	*Real difference (Population)*		
Test result	**FALSE**	True	TOTAL
Stat. Insignificant	# OK	# Type 2	Sum(Row1)
Stat. Significant	**# Type 1**	# OK	Sum(Row2)
TOTAL	Sum(False)	Sum(True)	SUM
Type 1 error: Statistically-significant result is false			
Type 2 error: Statistically-insignificant result is true			

Medical test	DISEASED		
Result	No	Yes	Total
Negative	95	5	100
Positive	5	95	100
Total	100	100	200

50% are diseased; 95% confirmation accuracy

Hypothesis test	Research Hypothesis (RH)		
Result	FALSE	TRUE	Total
Not Stat Sig	95	5	100
Stat. Signif.	5	95	100
Total	100	100	200

50% chance RH is true; 95% confirmation accuracy

Interpreting Statistical Significance Bayesian Medical Tests (5%)

Unlikely to be true in reality:

- Group: prone to heart attacks on 4th day of the month
- Women who eat cereal are more likely to have boys.
- Hurricanes with female names are deadlier.

Extraordinary claims require extraordinary evidence (Ecree)

Medical test	DISEASED		
Result	No	Yes	Total
Negative	1,805	5	1,810
Positive	95	95	190
Total	1,900	100	2,000

5% are diseased; 95% confirmation accuracy

Hypothesis test	Research Hypothesis (RH)		
Result	FALSE	TRUE	Total
Not Stat. Signif.	1,805	5	1,810
Stat. Signif.	95	95	190
Total	1,900	100	2,000

5% chance RH is True; 95% confirmation accuracy

Interpreting Statistical Significance
Bayesian Medical Tests (5%)

If research hypothesis (RH) is more unlikely than not:

1. Confirmation error must be equally unlikely so that a statistically-significant result give 50% chance that RH is true.
2. If 50% chance that RH is true, then confirmation error must be less than prior chance that RH is true.

Summary

1. Statistical significance can be influenced
2. When research hypothesis is more likely to be true, and sample statistic is statistically significant (1 chance in 20), then the result is strong evidence that research hypothesis is true in population.
3. When research hypothesis is less likely to be true, then criteria for statistical significance must be equally small.

Statistical Literacy: IDSL Keynote

Milo Schield
ASA Fellow
Visiting Professor: New College of Florida
Adjunct Researcher: Univ of New Mexico

Nov 25, 2025

International Day of Statistical Literacy
International Statistical Literacy Project

www.StatLit.org/pdf/ 2025-Schield-IDSL-StatLit-Keynote-Slides.pdf

1

Strong Support for Statistical Literacy

Walker (1951). *Statistical Literacy in the Social Sciences*
Haack (1979). *Teaching Statistical Literacy*
Moore (1998). statistical ideas will educated people require in the 21st century?
Garfield (1999) The understanding of statistical language: words, symbols, and terms.
Snell (1999). the ability to understand statistical concepts and reason at the most basic level"

2

Strong Support for Statistical Literacy

GAISE (2005). define statistical literacy as understanding the basic language of statistics…

3

Weak Support for Statistical Literacy

Rumsey (2002). I will omit the phrase "statistical literacy" from my discussion…
Borers (2006) argued that "their existence is not dictated by empirical observations" so they are "unsuitable as learning goals for statistics education."
Everson (2015). concern that not everyone may understand what is meant by 'statistical literacy' and 'statistical thinking'.

4

Strong 'Support' for Statistical Literacy

Schield (2017). "There is evidence – abundant evidence – in the same document that the GAISE 2016 revision has actually accepted – and encouraged – statistical literacy. Their official recommendation has done more to advance the cause of statistical literacy than any previous paper, book, group or committee. Statistical literacy may have lost a battle as a phrase, but it may be winning the war as an idea."

5

Strong 'Support' for Statistical Literacy

Schield (2017). "Those reading the new GAISE report (GAISE, 2016) may not realize how big of a change this revision is. Multivariate thinking and confounding were not even mentioned in the earlier report (GAISE, 2005), neither in McKenzie's (2004) list of core concepts, nor in articles on statistical literacy by Watson (2003) and Gal (2002)."

6

Three Problems Defining Statistical Literacy

1. Problem defining two-word phrases. (Schield, 2017)
2. Problem defining 'statistics' (this talk)
3. Problem on whether to include confounding. (this talk)

7

2. 'Statistics' is Ambiguous

Statistics-Education		Statistical Literacy
Survey data	**Data**	Everyday # in context
Sample survey data	**Statistics**	Everyday stats in context
Context: outlier & bias		Context: all influences
Random variation		Systematic variation
Confidence Intervals		Simpson's paradox
Statistical Significance		Cornfield's conditions

8

3. 'Confounding' is Absent or Minimal

2021 Gal Civic Statistics. Correlate (1)
2019 Gal Understanding Statistical Literacy.
Adjusted (8), multivariable (4), correlate (1)
2018 [3] Understanding Civic Statistics
Multivariate (7), aggregate (6), confound (3)

2002 Gal Adult's Statistical Literacy
2002 Rumsey Statistical Literacy as a Goal

9

3. 'Confounding' is Absent or Minimal

In examining 100 intro statistics textbooks,
- 80 do not include confounding in the index
- 15 show one example (burglaries & ice cream)
- 5 show multiple examples.

None of them show how to work a problem without a computer and without algebra.

None of them show how statistical significance can be influenced.

10

3. 'Confounding': Comparison is Simpler

Most statisticians are trained in mathematics, where continuous data is the norm.
Two-group discrete is more common in the humanities and in many of the social sciences.

Adjusting for confouding	
Two-factor co-variation	Two-group comparison
Multivariate analysis	Standardization
Requires software	No software or algebra rqrd.

11

How Statistical Significance can be Influenced

Most introductory statistics textbooks present statistical significance as a 'fact' that is etched in stone regardless of the kind of study.
As teachers, we know that statistical significance can be influenced in an observational study.
But we don't show it.
Yet most of our students taking statistics will be dealing with observational studies.
Here is how it can be shown easily.

12

8.3, P 322, 324 2025 Schield IDSL StatLit Keynote Slides V1 13

Statistical Significance: The Idea (very abstract)

Statistically significant: statistically unlikely. The less likely the chance, the more 'significant' the result.

Examples:

- Dog that didn't bark. Sherlock Holmes, Silver Blaze
- Goldfinger meets Bond 3 times, 3 different places
- Five heads in five flips of a fair coin.

Statistically significant:

- Not necessarily important
- is arbitrary (< 1 chance in 20) but not unreasonable
- **can be influenced**

13

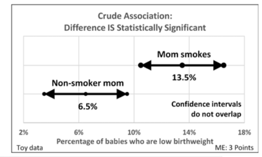
8.4.1: P 324, 325 2025 Schield IDSL StatLit Keynote Slides V1 14

Statistical Significance: Influenced by Confounding

Crude Association:
Difference is stat. signif.

Young: 50% of smokers
10% of non-smokers

Adjust for mom's age:



A	B	C	D	1	F	G	H	I
Low-weight birth rates: crude				2	**Low-weight birth rates: mix-match**			
Moms	Teens	Older	ALL	3	Moms	Teens	Older	ALL
Non-Smokers	11%	6%	**6.5%**	4	Non-Smokers	11%	6%	6.5%
Smokers	16%	11%	**13.5%**	5	Smokers	16%	11%	**11.5%**
DISTRIBUTION (weights):				6	DISTRIBUTION (weights):			
Non-Smokers	10%	90%	100%	7	Non-Smokers	10%	90%	100%
Smokers	50%	50%	100%	8	Smokers	**10%**	**90%**	100%
D4=B4*B7 + C4*C7	=0.11*0.10 + 0.06*0.90				I4	Same as D4		
D5=B5*B8 + C5*C8	=0.16*0.50 + 0.11*0.50				**I5=G5*G8 + H5*H8**	**=0.16*0.10 + 0.11*0.90**		

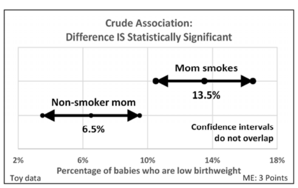
14

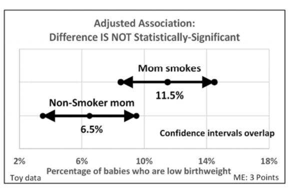
8.4.1: P 326 2025 Schield IDSL StatLit Keynote Slides V1 15

Statistical Significance: Influenced by Confounding

Crude Association: Difference statistically significant

Adjusted Comparison (same mix of ages):
Difference is not statistically significant.



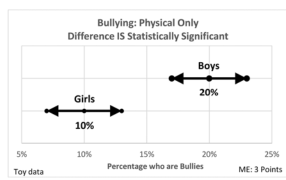


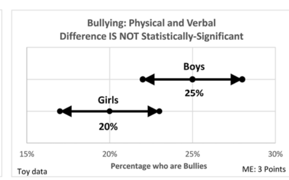
15

8.4.2; P 326 2025 Schield IDSL StatLit Keynote Slides V1 16

Statistical Significance: Influenced by Assembly

Bullying #1: physical only. 10% of girls. 20% of boys.





Bullying #2: physical & verbal. Girls 20%; boys 25%.
Change in definition (assembly) changes
statistically-significant into statistically insignificant

16

8.4.3, P 327 2025 Schield IDSL StatLit Keynote Slides V1 17

Statistical Significance: Influenced by Randomness

Clinical trial: difference is statistically significant.

If the observed difference is totally spurious (bogus), how could statistical significance be expected?

1. One trial repeated 20 times.
2. 20 outcomes in single trial.
3. 20 researchers each: one trial and one outcome.

Also, one researcher does 20 different trials each with one outcome.

17

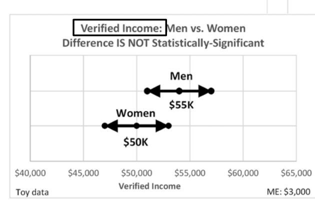
8.4.4, P 328 2025 Schield IDSL StatLit Keynote Slides V1 18

Statistical Significance: Influenced by Error/bias

Income by Sex:

Reported vs. verified.

Subject bias!



18

8.10, P 341 — 2025 Schield IDSL StatLit Keynote Slides — V1 19

Chance: Common Misunderstandings

1. If two outcomes, then 50% chance of either. Maybe.
2. If 90% chance of rain and no rain, forecast was wrong. Maybe. Can't tell from a single event.
3. If excess from expected, offset is predicted. No. Dilution is slower but maintains independence.
4. If medical test is 99% accurate and result is positive, then 99% chance that patient has disease. Maybe.
5. The more unlikely the outcome, the more likely it is causal. Maybe. Unlikely sometime, somewhere or unlikely then, there?

19

8.5, P 329 — 2025 Schield IDSL StatLit Keynote Slides — V1 20

Hypothesis Tests and Criminal Trials

Criminal trial: Innocent until proved guilty
Statistical trial: No difference until proved different

Type 1: false alarm. Type 2: silent alarm

---------------- Criminal Trial --------------		
	Defendant (Reality)	
Jury verdict	**INNOCENT**	Guilty
Not-guilty	OK	Type 2
Guilty BRD	**Type 1**	OK
Type 1 error: Innocent person found guilty		
Type 2 error: Guilty person found not-guilty		
Minimize Type 1 error. Presume innocent.		

--------------- Hypothesis Test --------------		
	Claim (Reality)	
Statistical result	**FALSE**	True
Stat. Insignificant	OK	Type 2
Stat. Significant	**Type 1**	OK
Type 1: Statistically significant result is false		
Type 2: Statistically-insignificant result is true		
Minimize Type 1 error. Presume false		

20

8.12.1, P 343 — 2025 Schield IDSL StatLit Keynote Slides — V1 21

Interpreting Statistical Significance Bayesian Medical Tests (>50%)

Shift from confirmation (column %) to prediction (row %).

---------------- Medical Test --------------			
People	*Has the diesease (Reality)*		
Test result	**No**	Yes	TOTAL
Negative	# OK	# Type 2	Sum(Row1)
Positive	**# Type 1**	# OK	Sum(Row2)
TOTAL	Sum(No)	Sum(Yes)	SUM
Type 1 error: Disease-free person tests positive			
Type 2 error: Diseased person tests negative			

--------------- Hypothesis Test --------------			
Samples	*Real difference (Population)*		
Test result	**FALSE**	True	TOTAL
Stat. Insignificant	# OK	# Type 2	Sum(Row1)
Stat. Significant	**# Type 1**	# OK	Sum(Row2)
TOTAL	Sum(False)	Sum(True)	SUM
Type 1 error: Statistically-significant result is false			
Type 2 error: Statistically-insignificant result is true			

Medical test	DISEASED		
Result	No	Yes	Total
Negative	95	5	100
Positive	5	95	100
Total	100	100	200

50% are diseased; 95% confirmation accuracy

Hypothesis test	Research Hypothesis (RH)		
Result	FALSE	TRUE	Total
Not Stat Sig	95	5	100
Stat. Signif.	5	95	100
Total	100	100	200

50% chance RH is true; 95% confirmation accuracy

21

8.12.1, P 345, 346 — 2025 Schield IDSL StatLit Keynote Slides — V1 22

Interpreting Statistical Significance Bayesian Medical Tests (5%)

Unlikely to be true in reality:

- Group: prone to heart attacks on 4th day of the month
- Women who eat cereal are more likely to have boys.
- Hurricanes with female names are deadlier.

Extraordinary claims require extraordinary evidence (Ecree)

Medical test	DISEASED		
Result	No	Yes	Total
Negative	1,805	5	1,810
Positive	95	95	190
Total	1,900	100	2,000

5% are diseased; 95% confirmation accuracy

Hypothesis test	Research Hypothesis (RH)		
Result	FALSE	TRUE	Total
Not Stat. Signif.	1,805	5	1,810
Stat. Signif.	95	95	190
Total	1,900	100	2,000

5% chance RH is True; 95% confirmation accuracy

22

8.12.1, P 346 — 2025 Schield IDSL StatLit Keynote Slides — V1 23

Interpreting Statistical Significance Bayesian Medical Tests (5%)

If research hypothesis (RH) is more unlikely than not:

1. Confirmation error must be equally unlikely so that a statistically-significant result give 50% chance that RH is true.
2. If 50% chance that RH is true, then confirmation error must be less than prior chance that RH is true.

23

8.11, P 342 — 2025 Schield IDSL StatLit Keynote Slides — V1 24

Summary

1. Statistical significance can be influenced
2. When research hypothesis is more likely to be true, and sample statistic is statistically significant (1 chance in 20), then the result is strong evidence that research hypothesis is true in population.
3. When research hypothesis is less likely to be true, then criteria for statistical significance must be equally small.

24