Expanding Opportunities by Reducing Algebraic Barriers

Patti Frazer Lock, St. Lawrence University

Robin Lock, St. Lawrence University

Kari Lock Morgan, Penn State University

USCOTS 2021

Better Title

Algebra is the Problem!

Maybe We Can Be (Part of) the Solution

Zoom Poll #1

Is there any Algebra prerequisite or co-requisite for your Intro Stat course?

A. Yes
B. No
C. It depends
D. I don’t know

Do City Dwellers React Differently to Stress than Country Dwellers?

A study used brain scans to measure reactions to stress of city dwellers and country dwellers.

In order to induce stress in the participants, they were given a difficult math test.

In increase the stress of the participants, those conducting the study tried to humiliate the participants by telling them how poorly they were doing on the test.

Great question for students:
Is this an experiment or an observational study?

Conclusion:
Results were significant; city dwellers showed more stress.

Abbott A, "City Living Marks the Brain," *Nature*, **474**, 429 (2011).
Lock5: 1.117

Let's Do a Math Test...

With a bit of role-playing....

1. Solve for x: $x^2 - x = 6$

Factor x out on the left????

NO! WE SET IT EQUAL TO ZERO, OF COURSE.

Let's Do a Math Test...

With a bit of role-playing....

2. Solve for x: $2x = 6$

Set it equal to zero????

NO! WE DIVIDE BOTH SIDES BY 2, OF COURSE.

Let's Do a Math Test...

With a bit of role-playing....

3. Solve for x: $e^x = 6$

Set it equal to zero???? Divide both sides by 6????

NO! WE USE LOGARITHMS, OF COURSE.

Let's Do a Math Test...

With a bit of role-playing....

4. Solve for x: $e^x = 2x$

Set it equal to zero???? Divide both sides by 2???? Use logarithms???

(Um, no. This equation – and most equations – can't be solved by algebra. But let's not talk about that.)

Let's Do a Math Test...

With a bit of role-playing....

5. Simplify: $\frac{x}{3} + \frac{5}{3} = \frac{8}{3}$

I have no idea. I'm completely stressed out and my brain has shut down.

THAT'S OKAY. JUST PAY ATTENTION.
WE CAN MAKE THIS EASIER BY MULTIPLYING THROUGH BY 3.

Let's Do a Math Test...

With a bit of role-playing....

6. Simplify: $\frac{x^2}{3} + \frac{5x}{3} - \frac{8}{3}$

Okay, I'm trying. Multiply through by 3???

NO!! YOU CAN'T JUST MULTIPLY AN EXPRESSION BY 3!

Conclusions

1. Algebra is hard!

2. Most citizens of the world (even in STEM fields!) will do very little algebra by hand, ever.

3. And expecting students to do lots of algebra has real consequences ...

Expanding Opportunities

According to the *Common Vision* report of the Mathematical Association of America:

Algebra is "**the most significant barrier**" to students finishing a degree in both STEM and non-STEM fields.

A Common Vision for Undergraduate Mathematical Sciences Programs in 2025, Mathematical Association of America, 2015

Expanding Opportunities

These barriers are most acute for low income students, first generation college students, and underrepresented groups.

Bob Moses, an icon of the civil rights movement, argues in The Algebra Project that:

“math is the civil rights issue of the twenty-first century.”

The Algebra Project: Bob Moses on math literacy as a civil right

What Can We Do?

What can <u>we</u> do, in our roles as stat or DS instructors, to increase student access and student opportunities?

This whole conference is about "Expanding Opportunities"

We've already heard many great ideas at this conference about ways to expand opportunities. Hooray!

But we're told by the MAA that algebra is the most significant barrier to student success (and lots of education research backs this up).

Maybe the answer isn't to find more ways to teach the algebraic parts (prerequisite courses, co-requisite courses, stretch courses,...).

This whole conference is about "Expanding Opportunities"

Maybe instead we can try to:

Carefully construct our courses to focus more on understanding data and less (maybe much less) on algebraic methods.

Research on barriers to student success tell us that this might be the <u>most</u> effective thing we can do.

A (partial) Solution?

Can we find ways to teach a rigorous Intro Stat or Intro Data Science course that meets our students' needs, while also removing barriers and being more inclusive?

What Intro Stat?

Calculus-Based I Why??

Algebra-Based Why??

~~Intro Stat as a low-level math course?~~ No!!

(But having it replace College Algebra as the dominant required quantitative course would be great...)

Intro Stat as a course focused on applied data analysis, serving as an introduction to the dynamic field of statistics (for STEM majors or non-STEM majors,...)

A (partial) Solution?

Can we find ways to teach a rigorous Intro Stat or Intro Data Science course that meets our students' needs, while also removing barriers and being more inclusive?

Zoom Poll #2

Do your students need to know the formula for hyperbolic asymptotes?

(The asymptotes for the hyperbola $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ are $y = \pm\frac{b}{a}x$.)

A. Yes, this is essential
B. Nice but not essential
C. Not necessary at all

Zoom Poll #3

Do your students need to know how to use a number line? (e.g. to understand an interval or to know which numbers are larger than other numbers)

A. Yes, this is essential
B. Nice but not essential
C. Not necessary at all

How Much is Necessary?

So (perhaps?) we are agreed that our students need SOME math knowledge.

But how much? What is the minimum?

Key Question: How much can we reduce the amount without hurting the integrity of the course?

First: Descriptive Statistics

Let’s focus first only on descriptive statistics.

(We’ll move on to inference later.)

Zoom Poll #4

Which of the following summary statistics, if any, should students compute <u>by hand from a formula</u>:
(check all that apply)

- ❑ mean
- ❑ median
- ❑ standard deviation
- ❑ quartiles
- ❑ correlation
- ❑ slope of a regression line
- ❑ none of the above

Zoom Poll #5:

Which of the following graphs, if any, should students draw by hand: (check all that apply)

- ❑ histogram
- ❑ dotplot
- ❑ stem-and-leaf plot
- ❑ boxplot
- ❑ plotting points on a scatterplot
- ❑ drawing the regression line on a scatterplot
- ❑ none of the above

Breakout Rooms

Discuss the results of the two polls. In addition:

There is more to descriptive statistics than just those basic summary statistics and graphs. Which of the following, if any, should students compute by hand:

- Z-scores?
- Residuals?
- Proportions?
- Proportions from a two-way table?
- Other?

Discuss: Should they do these things by hand? If so, what are the minimal mathematical requirements?

Reducing Algebraic Barriers

1. **<u>All</u>** summary statistics and **<u>all</u>** graphs are found using technology.

    Far less algebra;

    Questions focused on the data are definitely NOT less rigorous

2. Move toward Data Science!

    Multiple variables! Data Visualization!

    Far less algebra;

    Questions focused on the data are definitely NOT less rigorous

Example

The five-number-summary for number of vaccinated residents of states (in 100,000's as of 6/26/21 according to the CDC) is

`min=2.0   Q1=7.5   m=19.0   Q3=40.9   max=193.6.`

(a) What would expect for the **shape** of the distribution?

- symmetric
- skewed to the left
- skewed to the right

(b) Which of the following is most likely to be the **mean**?

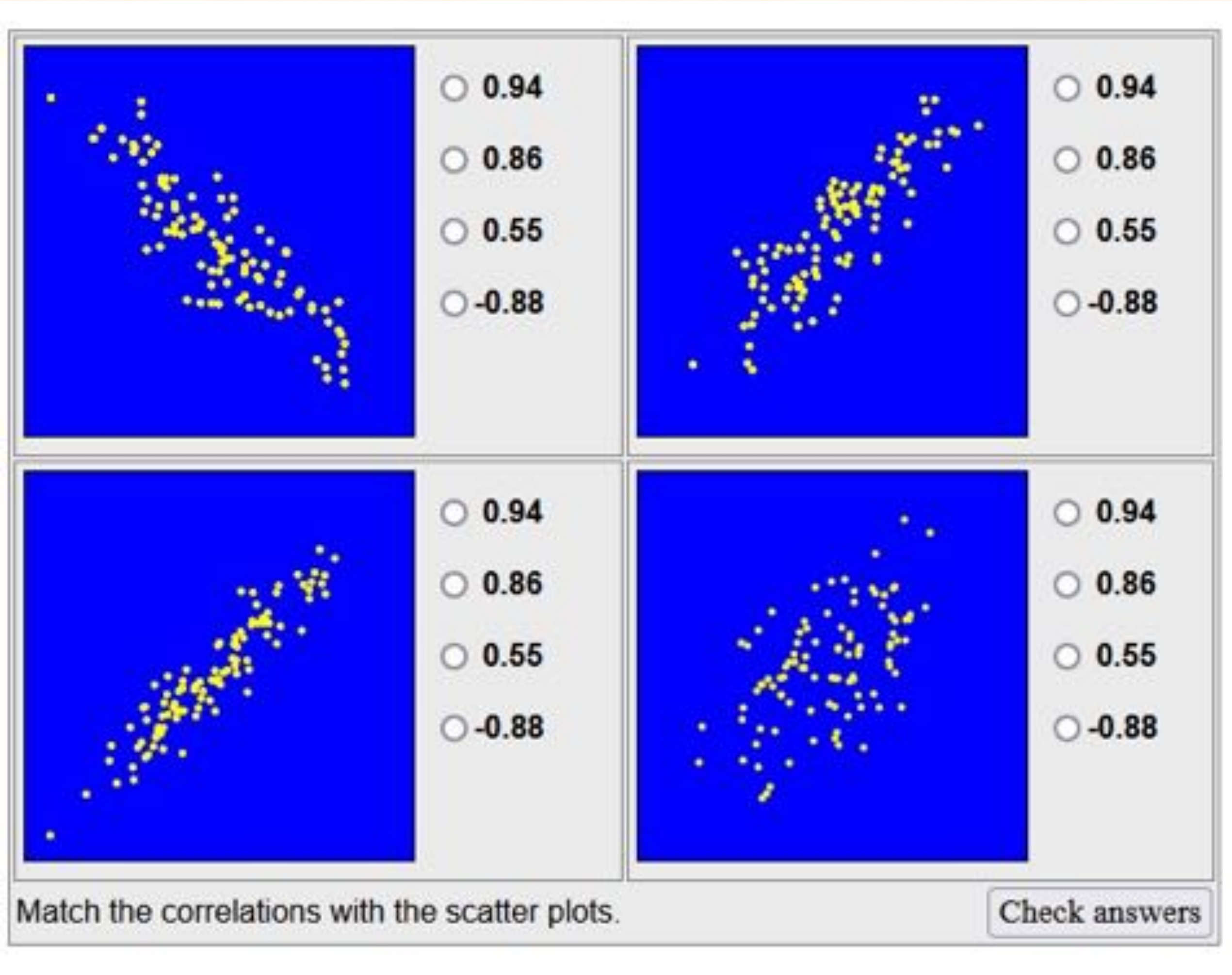
`19        30        53        98`

What about the percentage of residents vaccinated in each state?

`min=37.8   Q1=48.5   m=55.1   Q3=62.7   max=75.0`

Example: Correlation Guessing Game

istics.net/Correlations

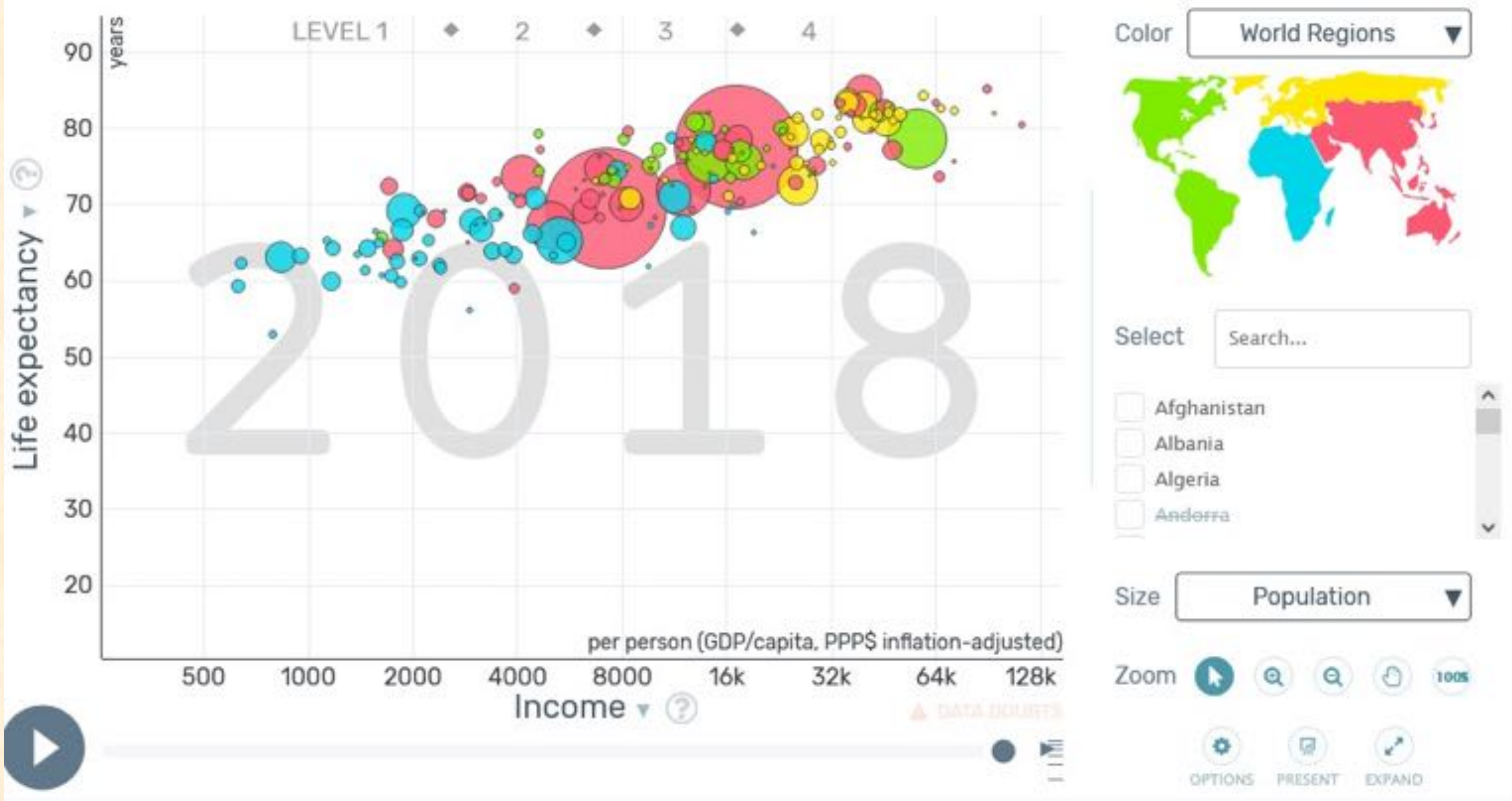


Match correlations with scatterplots.

Have a class competition!

Example: Gapminder

gapminder.org/tools



Cases?

Variables?

Association?

Outliers?

Trend over time?

Example: Association

Side-by-side boxplots are shown for two different datasets, A and B, each involving one categorical variable (with group 1 and group 2) and one quantitative variable. (All sample sizes are large.)
Which dataset (A or B) appears to show the strongest association between the two variables?

Dataset A

Non-athlete

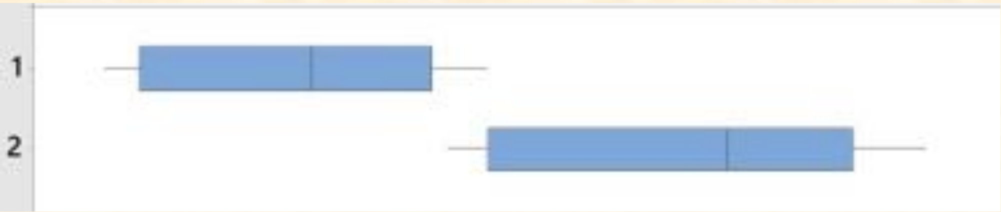
Athlete

Exercise

Dataset B

Non-athlete

Athlete



Exercise

Lock5: 2.167

Next: Inference

Let’s turn now to inference.

Lily Lock Morgan!

Zoom Poll #6

Which of the following computations about ***estimation***, if any, should students compute by hand from a formula:
(check all that apply)

- ❑ Standard error
- ❑ Normal or t-distribution values from a table or calculator
- ❑ Confidence interval
- ❑ Determining sample size
- ❑ None of the above

Zoom Poll #7

Which of the following computations about ***testing***, if any, should students compute <u>by hand from a formula</u>:
(check all that apply)

- ❑ Standard error
- ❑ Standardized test statistic
- ❑ Normal or t-distribution values from a table or calculator
- ❑ None of the above

Zoom Poll #8

In order to calculate and/or understand confidence intervals or hypothesis tests, which of the following algebraic techniques, if any, are necessary: (check all that apply)

- ❑ Solving equations
- ❑ Simplifying expressions
- ❑ Evaluating a formula at specific values
- ❑ Order of operations
- ❑ Understanding intervals on a number line
- ❑ Understanding relative magnitude of numbers
- ❑ Other (put in chat)
- ❑ None of the above

Breakout Rooms

Discuss the results of the three polls. In addition:

Discuss: How necessary are formulas for student understanding? Do they help? Do they hurt? What is the minimal amount of algebra (or mathematical knowledge) needed?

Reducing Algebraic Barriers

1. **<u>All</u>** inference results are found using technology!

   Students need to be able to interpret, explain, make conclusions

   Far less algebra;

   Questions focused on the results are definitely NOT less rigorous

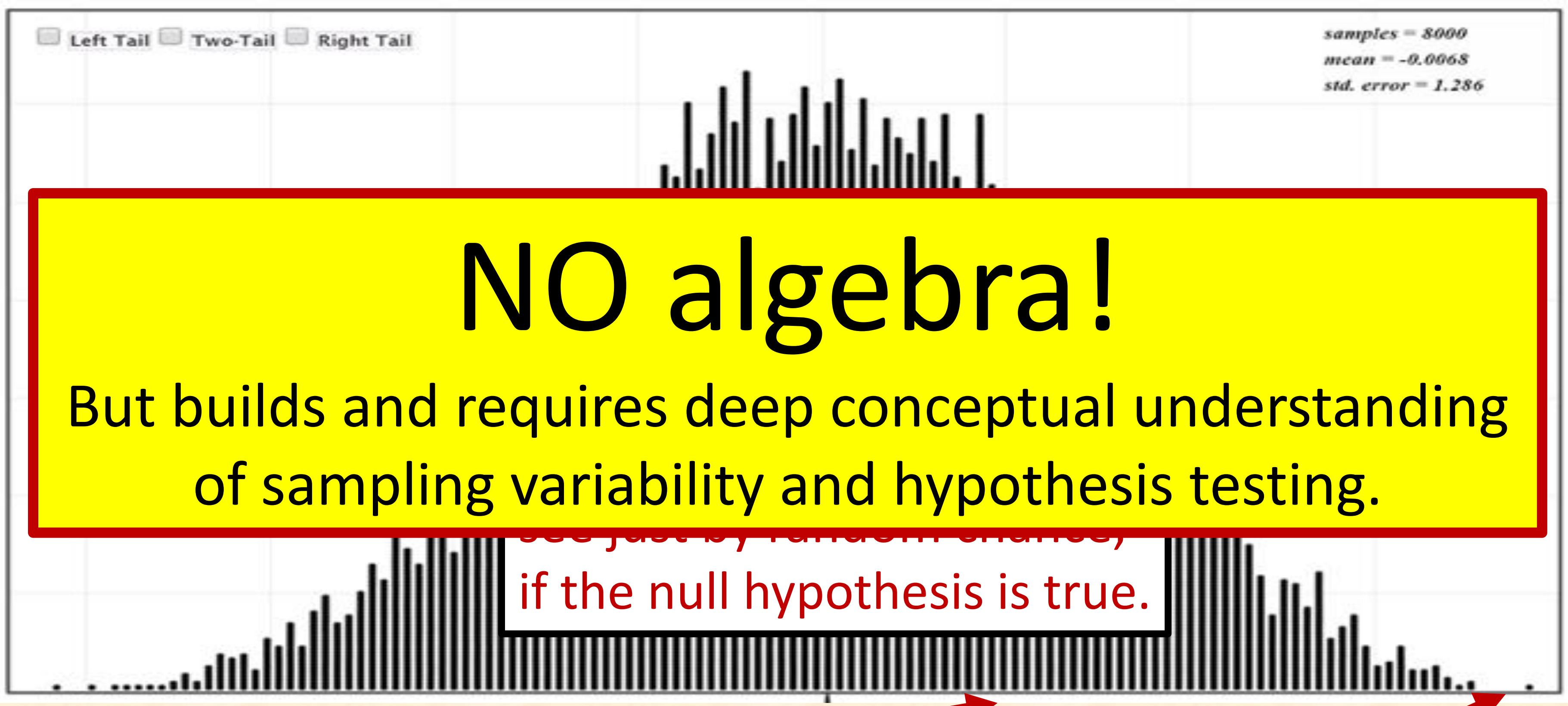
2. Simulation-based inference!

   Same intervals and tests; More understanding

   Far less algebra;

   Questions focused on the results are definitely NOT less rigorous

Simulation-Based Hypothesis Test



NO algebra!

But builds and requires deep conceptual understanding of sampling variability and hypothesis testing.

Sample statistic here:
Could be just random chance.
Not enough evidence to make a clear conclusion.
(NOT a small p-value)

Sample statistic way out here:
Unlikely to be just random chance!
There is evidence for the effect!
(small p-value)

Traditional Hypothesis Test

1. Check conditions

2. Which formula?

$$t = \frac{\bar{x}_L - \bar{x}_D}{\dots}$$

$$\sqrt{\frac{2.97^2}{10} + \frac{1.56^2}{8}}$$

4. Chug with calculator

$$t = 2.41$$

0.02< p-value < 0.025

5. Which theoretical distribution?

6. df?

TABLE B: t-DISTRIBUTION CRITICAL VALUES

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

LOTS of algebra!

Process requires algebraic skill but does not build or require conceptual understanding of sampling variability or hypothesis testing.

Not Less Rigorous!

Inference is conceptually hard.

The concepts we want our students to understand are not intrinsically connected to the algebra!

Maybe separating these key statistical ideas from algebraic manipulation will help our students understand.

Not Less Rigorous!

- Students still need to be able to:
  - *Understand sampling variability: Hard!!!*
  - *Interpret (and understand) the meaning of a confidence interval: Hard!!*
  - *State the hypotheses.*
  - *WAIT! First: What are the relevant parameters? Apparently, Hard!*
  - *Understand what a p-value actually means: Hard!!*
  - *Give a conclusion in context based on a p-value: Hard!!*

  - *Calculate intervals and tests based on real data:*
    (hard with algebra, easier with technology!)

Example: Intervals

Researchers had participants take a computer motion test, and response time and accuracy were recorded. Half of the participants were regular video-game players and half were non-players.

A 95% confidence interval for difference in mean *response time* (players minus non-players) is -1.8 to -1.2 seconds.

A 95% confidence interval for difference in mean *accuracy score* (players minus non-players) is -4.2 to +5.8 seconds.

The point is *understanding* not the algebraic calculation!

Green, et al., "Improved probabilistic inference as a general learning mechanism with action video games," *Current Biology*, 20, 2010.
(Lock5: 3.73)

Example: Tests

A well-designed study randomly assigned students in a college physics course to either an active learning class or a passive learning (lecture) class. Students were given a test to measure how much they *actually* learned and also rated how much they *thought* they had learned.

Two tests were conducted:

- One tested whether students with active learning learned more, on average.
- Another tested whether students with active learning thought they learned less, on average.

The p-values in both cases were very small. What can we conclude?

The point is *understanding* not the algebraic calculation!

Deslaurier L, et al., "Measuring actual learning vs feelings of learning in response to being actively engaged in the classroom," *PNAS*, 116(39), 2019.
(Lock5: 1.15 and 4.17)

Next: More Advanced Methods

These could all be done using technology:

- Chi-Square Tests
- Analysis of Variance for Difference in Means
- Inference for Regression (Simple)
- Multiple Regression with Inference

Next: Probability

Just kidding! We think this should be omitted from a one-semester Intro Stats or Intro Data Science course.

If formal probability (probability rules, Bayes' Rule, etc) is included, the amount of algebra required goes up significantly.

But this is not necessary for a course focused on applied data analysis.

What Else?

If we minimize algebra, how might we otherwise spend that time???

- More focus on understanding?
- More on real data context and conclusions?
- More on using technology?
- More active learning?
- More on multiple variables?
- More on data visualization?
- More on confounding / causal conclusions?
- More on cautions?
- More on ethics?
- More methods / procedures ?
- More data science?
- … [Type your “If I only had time” wish in chat… but don’t press enter yet!]

Takeaways

Algebra is hard!

Algebra is an (unnecessary!) barrier to success for too many students.

Typical intro topics can be rigorously taught with minimal algebra.

There are <u>many</u> better ways to spend valuable class time!

By reducing algebra requirements, we can ***<u>expand opportunities</u>***!

Thank you!

QUESTIONS?

Patti Frazer Lock: plock@stlawu.edu
Robin Lock: rlock@stlawu.edu
Kari Lock Morgan: klm47@psu.edu

Slides posted at USCOTS and at
www.lock5stat.com