Informal statistical inference: Using simulation to build on students' intuition about probability

Todd Swanson, Associate Professor of Mathematics, Hope College

Nathan Tintle, Professor of Statistics, Dordt College

Today

- Motivation for alternative approach to build student intuition for inferential thinking
- What is simulation?
- Three examples of simulations for informal inference
  - Hands-on
  - Free computer applets
- Assessment results
- Bigger picture

Today

- Emphasis
  - First exposures to statistical inference
  - Bridging life experience/intuition to informal inference to formal inference
  - MS, HS or post-secondary

Reflecting on how we teach statistics

- Descriptive statistics?
  - Can be fun, easy. Why?
  - Can be challenging and mundane. Why?
- Probability and sampling distributions?
  - Often challenging for students. Why?
- Inference?
  - Often challenging for students. Why?

The consequence

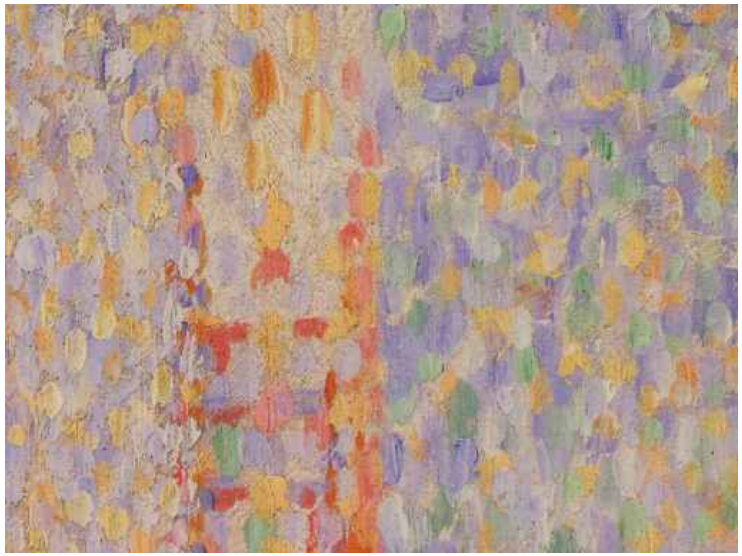
- Few students ever leave our course seeing statistics as this

The consequence

- The better students may get a fuzzy impression

The consequence

- All too many noses stay too close to the canvas, and see disconnected details



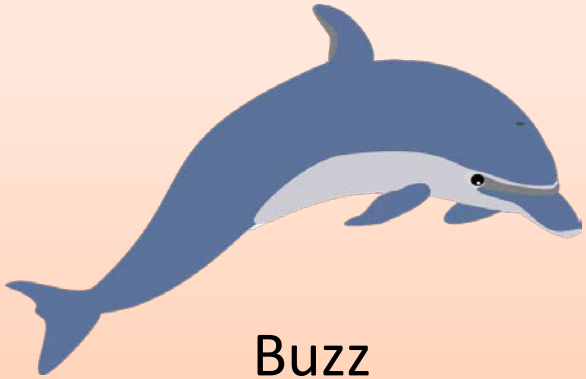
Is there a better way?

- Simulation-based inference (SBI)
  - Directly explore “what could happen by chance alone”?
  - Active, in-class experiences
  - Emphasizes statistical variability vs. deterministic thinking (GAISE K-12)
  - “Simulation-Based Inference” allows student exploration of statistical inference with minimal prior knowledge

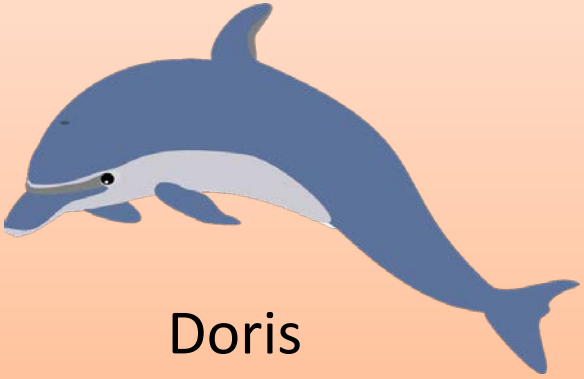
Is there a better way?

- Overarching statistical investigation process
  - 6-steps
    - Ask question, design study, explore data, inferences from data, draw conclusions, look back and ahead
  - PPDAC – Problem, Plan, Data, Analysis, Conclusion

Can dolphins communicate abstract ideas?

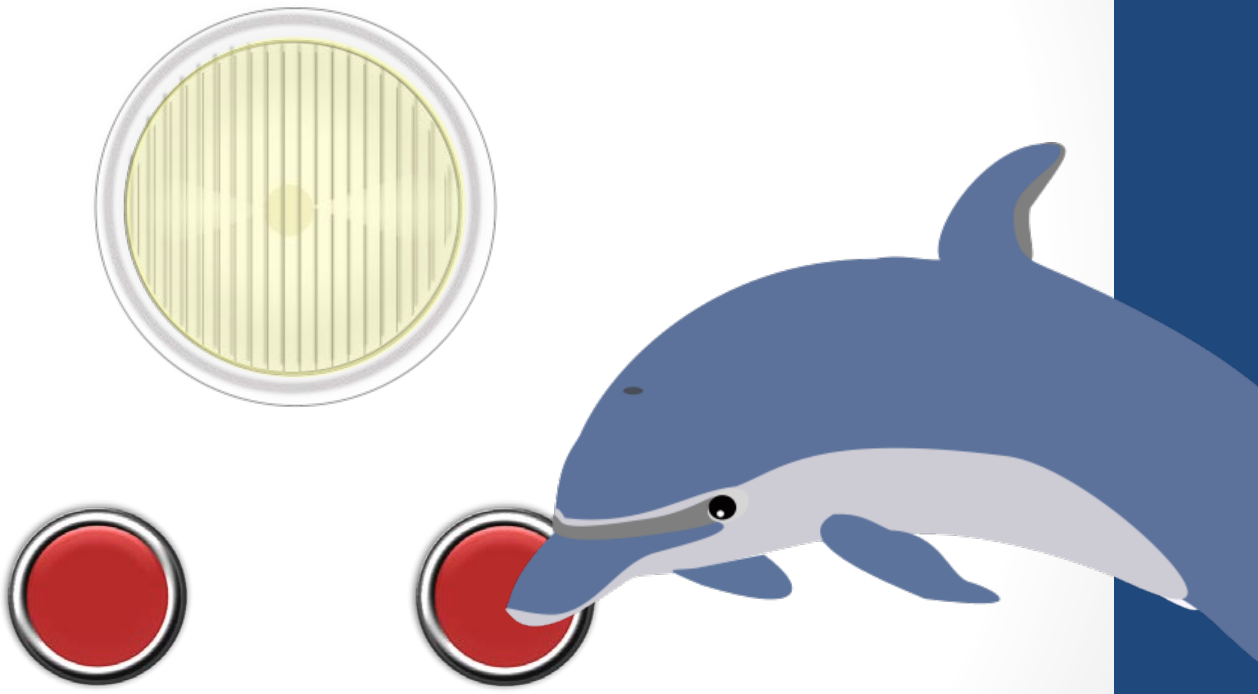


Buzz

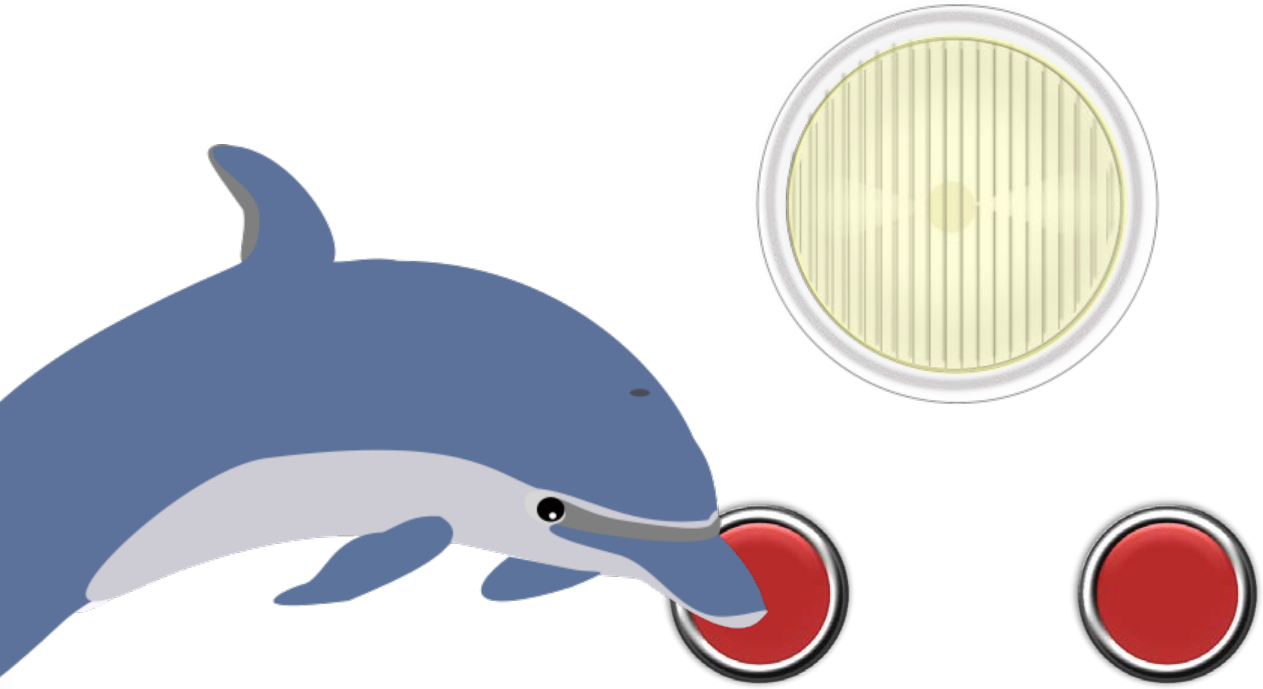


Doris

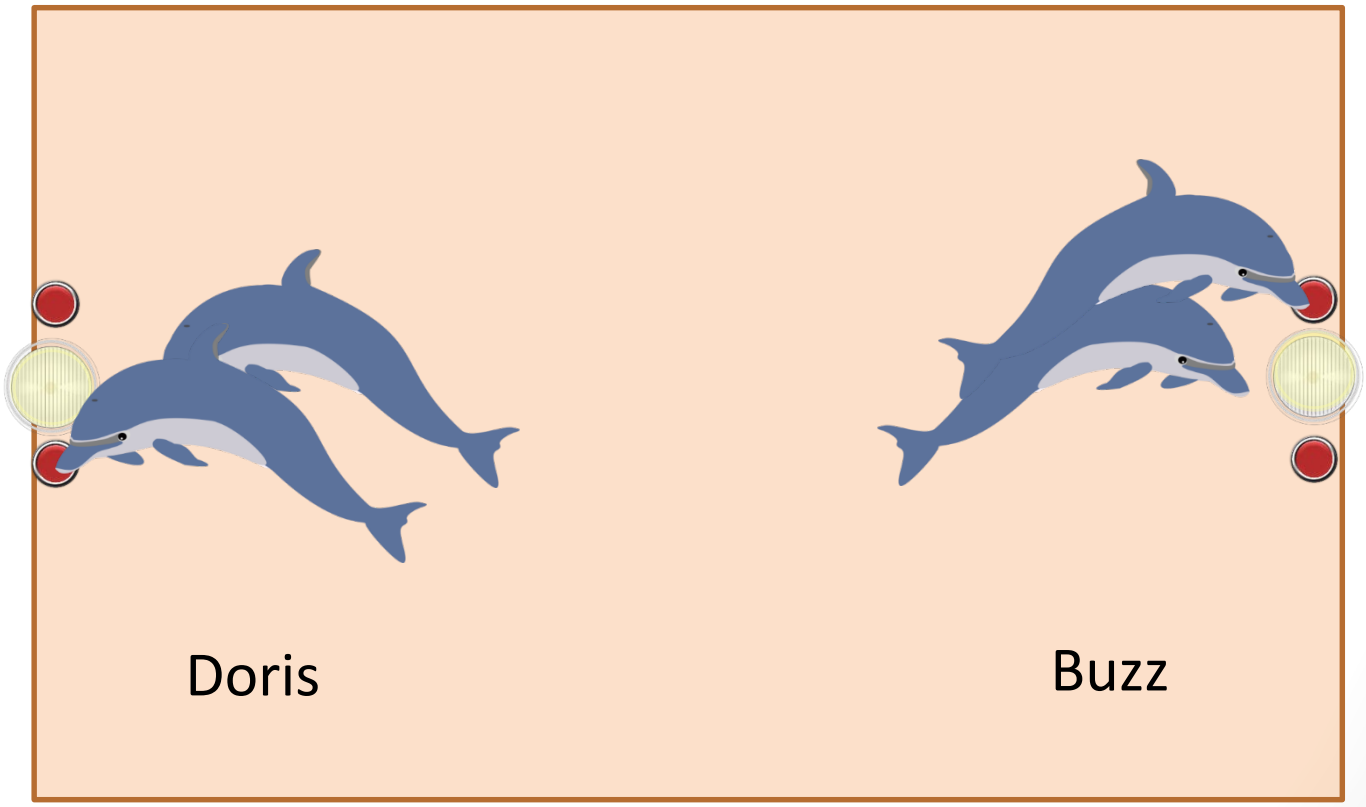
Step 1: Learn the Signals



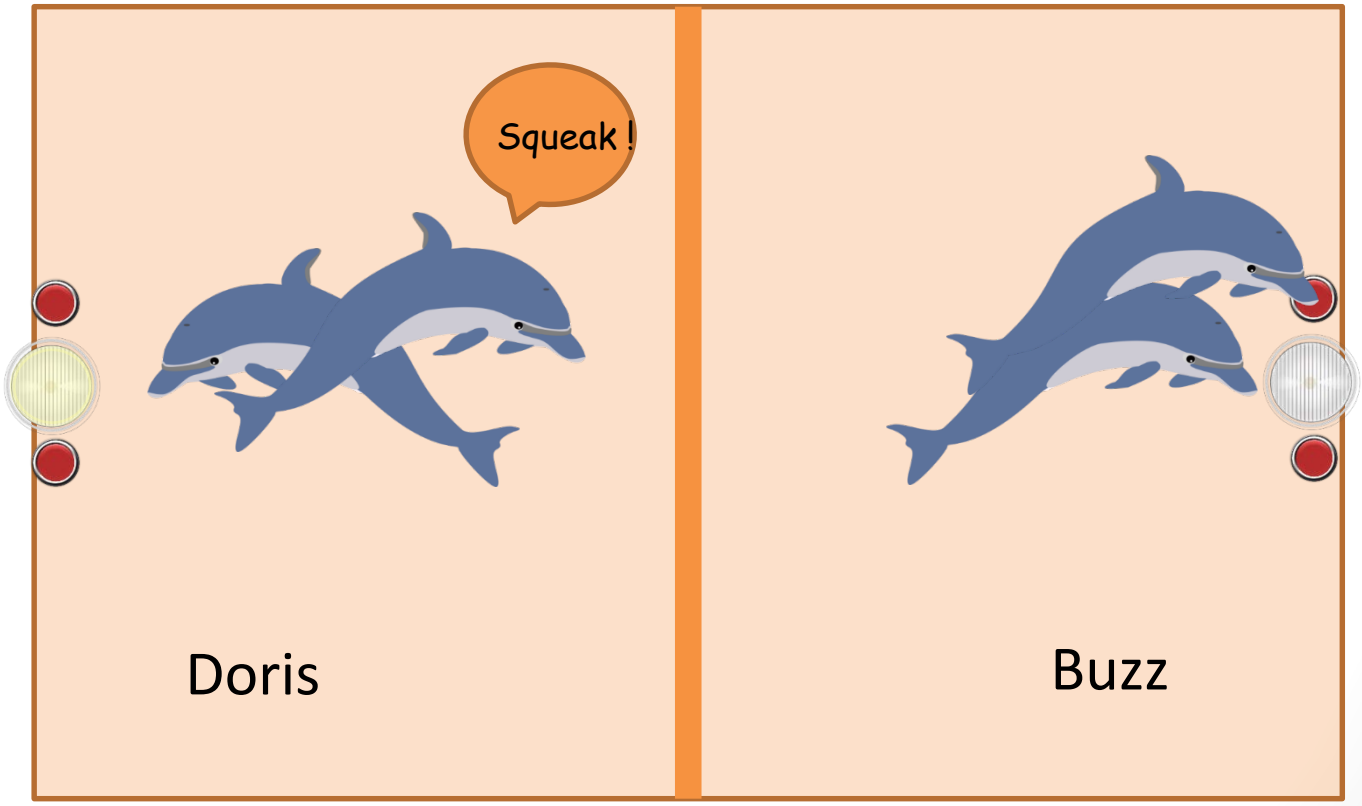
Step 1: Learn the Signals



Step 2: Learn the Response



Step 3: Communicate!



Can dolphins communicate abstract ideas?

- In one set of trials, Buzz chose the correct button 15 out of 16 times.

- Do you think that dolphins can communicate abstract ideas?

What do students' say?

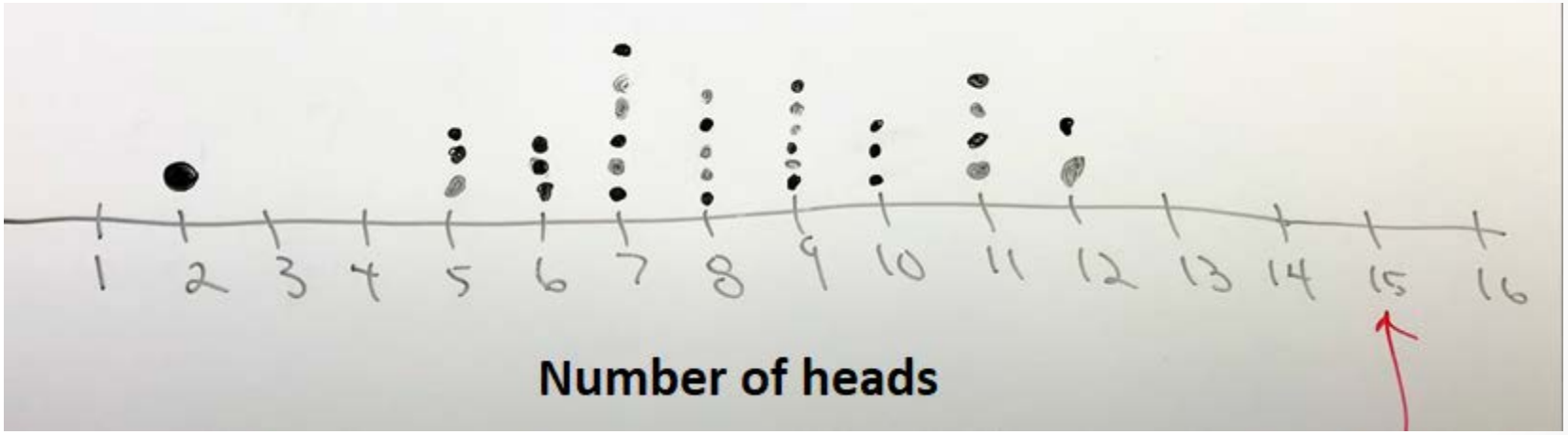
- https://www.youtube.com/watch?v=yKWtJVbStoc&feature=youtu.be (2:58 – 3:50)

Let’s try it

Flipping coins and sharing our results

Students plotting points for Buzz and Doris example

What could happen in your class



Why flip coins?

Each student is one dot on the graph

Witness random variability first hand

Students can often tell you why might want more dots, then turn to technology

Demo applet

www.isi-stats.com

Key components of technology

Same approach but much faster

Ability to make predictions/ask “what if” questions and immediately test intuition

- Ask new questions

Combine visual interpretation

- Ultimately: What does that p-value measure?

“Simulation-Based Inference” allows student exploration of statistical inference with minimal prior knowledge

- Distribution
- Variability
- Randomness

Review: Is there a better way?

- Simulation-based inference (SBI)
  - Tactile simulations
  - Bridge to intuitive technology that builds on active, in-class experiences
- Overarching process of statistics inference: 6-steps
- Advantages:
  - Keep students close to data and study design
  - Active learning
  - Conceptual – overarching process of statistical inference
  - Natural bridge from student experience with probability to informal inference (before terms) to formal inference
  - Emphasizes statistical variability vs. deterministic thinking (GAISE K-12)

It's not just us saying this!

- Core standards: Grade 7
  - Investigate chance processes and develop, use, and evaluate probability models: Find probabilities of compound events using organized lists, tables, tree diagrams, and ***simulation***.
  - Use random sampling to ***draw inferences*** about a population.

It's not just us saying this!

- High School: Statistics & Probability » Making Inferences & Justifying Conclusions
  - Understand and evaluate random processes underlying statistical experiments: Decide if a specified model is consistent with results from a given data-generating process, e.g., using ***simulation.***
  - Understand statistics as a ***process*** for making inferences about population parameters based on a random sample from that population.
  - Use data from a randomized experiment to compare two treatments; use ***simulations*** to [draw inferences].

3S Strategy

- **Statistic:** Compute the statistic from the observed data.
- **Simulate**: Identify a model that represents a chance explanation. Repeatedly simulate values of the statistic that could have happened when the chance model is true and form a distribution.
- **Strength of evidence**: Consider whether the value of the observed statistic is unlikely to occur when the chance model is true. Where does the observed **statistic** lie in the **simulated** null distribution?

Which is Bob and which is Tim?

Research Question: Do people have a tendency to associate certain facial features with a name?

Bob and Tim

More specifically, our research question is: Will a majority of people correctly identify which face belongs to Bob and which to Tim?

- What is our **statistic**?

The last time I used this in class I had 25/30 ≈ 83% correctly make the identification. This is quite typical. Statistics range from about 70% to 90%.

Bob and Tim

- If we were to use coins to construct a null distribution, how many times would we have to flip the coin?
- What would we mark on the board?

Bob and Tim

- When we go to the One Proportion Applet, what numbers go in the boxes?

Probability of heads: [ ]

Number of tosses: [ ]

Number of repetitions: [ ]

(•) Number of heads
( ) Proportion of heads

As extreme as [≥] [ ] [Count]

Bob and Tim

Let's go to the **One Proportion applet** and make a null distribution for the **proportion** of heads.

www.isi-stats.com/isi

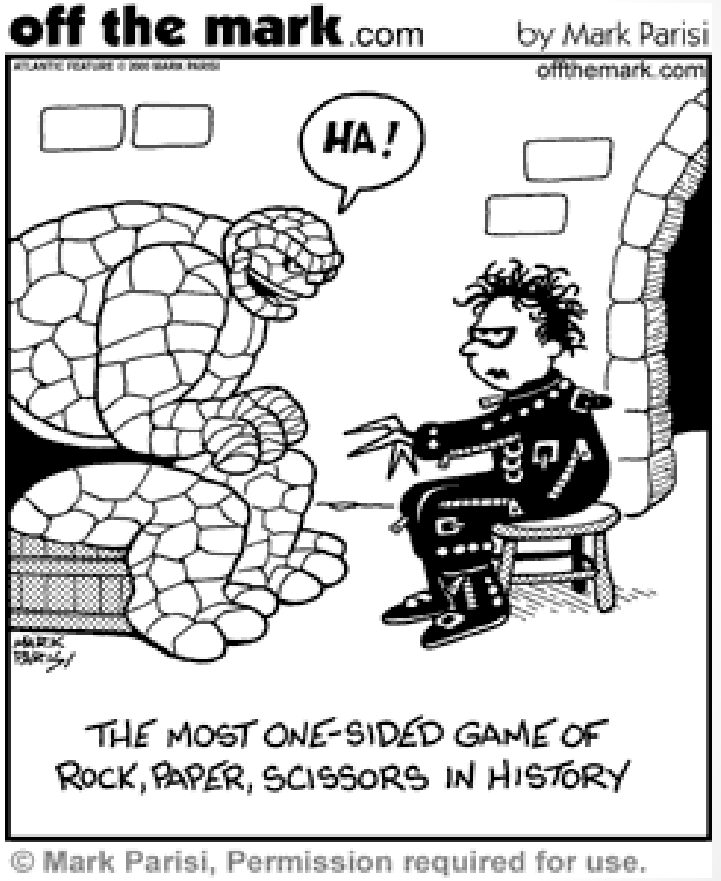
- What is your approximate p-value?
- Do we have strong evidence that the probability people will correctly identify Bob and Tim is more than 0.50?

We could take a look at Rock-Paper-Scissors-Lizard-Spock

- **Scissors cut paper**
- **Paper covers rock**
- **Rock crushes lizard**
- Lizard poisons Spock
- Spock smashes scissors
- Scissors decapitate lizard
- Lizard eats paper
- Paper disproves Spock
- Spock vaporizes rock
- (and as it always has) Rock crushes scissors

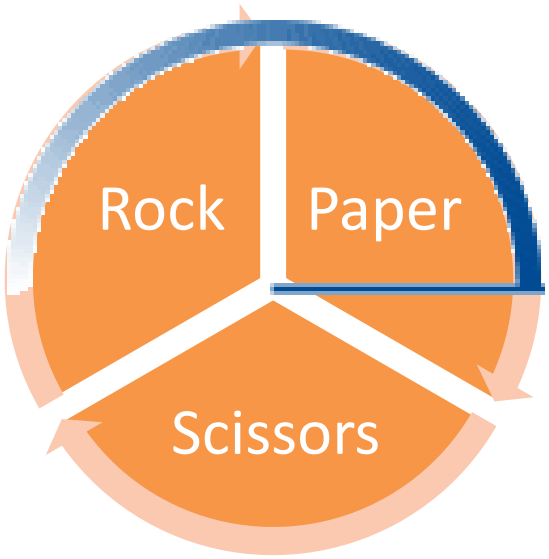
Rock-Paper-Scissors

- Rock smashes scissors
- Paper covers rock
- Scissors cut paper
- Are these choices used in equal proportions (1/3 each)?
- One study suggests that scissors are chosen less than 1/3 of the time. (29.6% of the time)



Setting up a Chance Model

- Because the Buzz and Doris example had a 50% chance outcome, we could use a coin to model the outcome from one trial. What could we do in the case of Rock-Paper-Scissors?



Three S Strategy

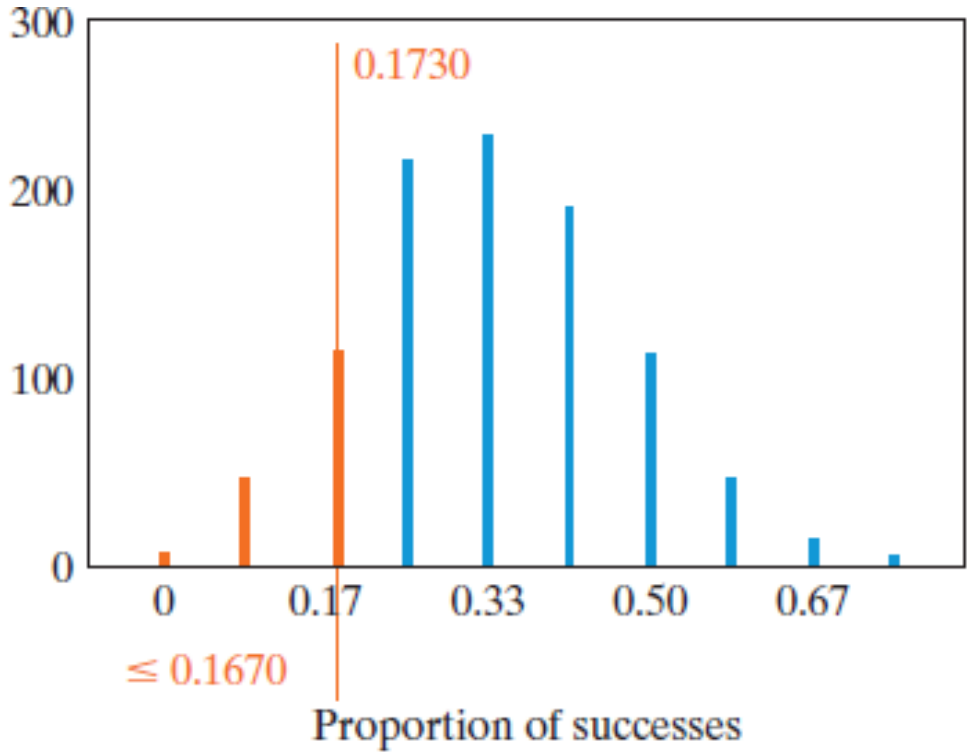
- **Statistic:** Compute the statistic from the observed data. [In a class of 12 students, 2 picked scissors.]
- **Simulate**: Identify a model that represents a chance explanation. Repeatedly simulate values of the statistic that could have happened when the chance model is true and form a distribution.
- **Strength of evidence**: Consider whether the value of the observed statistic is unlikely to occur when the chance model is true.

Applet

- We will use the One Proportion Applet for our test.
- This is the same applet we used last time except now we will change the proportion under the null hypothesis.
- Let's go to the applet and run the test. (Notice the use of symbols in the applet.)

Null Distribution

- The **null distribution** is the distribution of simulated statistics that could have happened in the study assuming the null hypothesis was true.



P-value

- The ***p-value*** is the proportion of the simulated statistics in the null distribution that are *at least as extreme* (in the direction of the alternative hypothesis) as the value of the statistic actually observed in the research study.
- We should have seen something similar to this in the applet:
  - Proportion of samples: 173/1000 = 0.173

What can we conclude?

- Do we have strong evidence that less than 1/3 of the time scissors gets thrown?
- How small of a p-value would you say gives strong evidence?

- **Remember the smaller the p-value, the stronger the evidence against the null.**

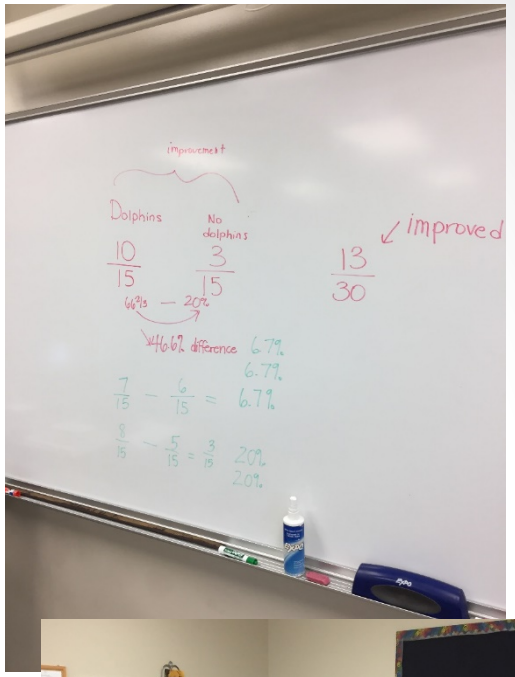
What can we conclude?

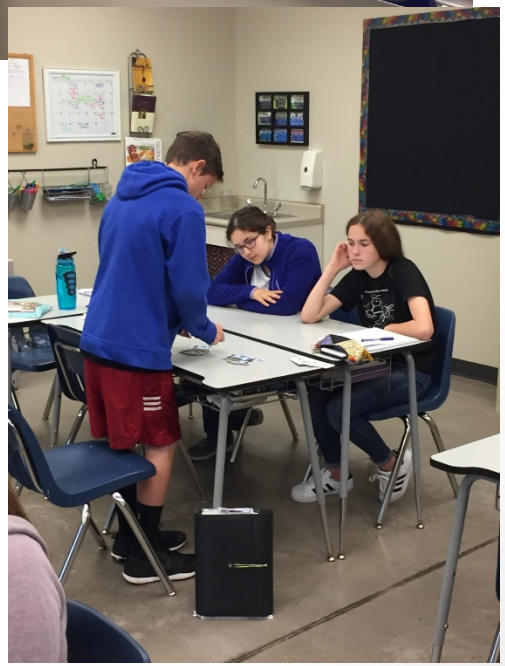
- So we do not have strong evidence that fewer than 1/3 of the time scissors is thrown.
- Does this mean we can conclude that 1/3 of the time scissors is thrown?
- Is it plausible that 1/3 of the time scissors is thrown?
- Are other values plausible? Which ones?
- Suppose 1/12 of our sample chose scissors instead of 2/12. How would the p-value change?
- What could we do in our study design to have a better chance of getting strong evidence that scissors is thrown less that 1/3 of the time?

Middle-school and High School

- Can this approach be adapted for use in middle and high school classrooms?
  - Numerous examples at both levels

- **Re: Doris and Buzz. Seventh grade math teacher Tammy Veenstra**: “It was really fun today….The kids all thought 2 dolphins were not enough dolphins to make a claim and that they could have been related or prodigy dolphins. Any time in their class contributions that they would talk about Doris and Buzz, they referred to them as "the prodigy dolphins." Most of my class said 15/16 was not a strong enough statistic to say dolphins could communicate but it was strong enough to say Doris and Buzz could. :) The class was very engaged! :)”


improvement
Dolphins
No dolphins
10/15
3/15
13/30
improved
66.7% — 20%
46.6% difference
6.7%
6.7%
7/15 − 6/15 = 6.7%
8/15 − 5/15 = 3/15
20%
20%



Assessment results

- Numerous papers at the college level (visit www.isi-stats.com) for details

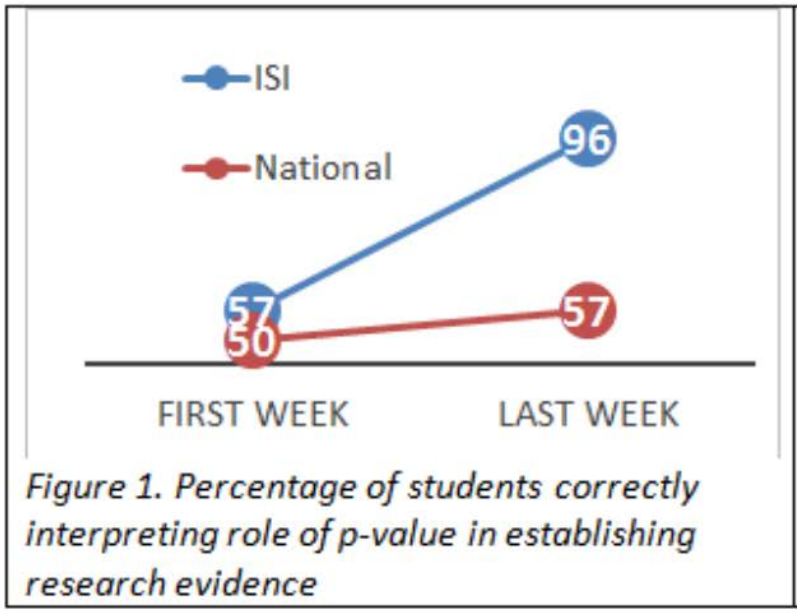


Figure 1. Percentage of students correctly interpreting role of p-value in establishing research evidence

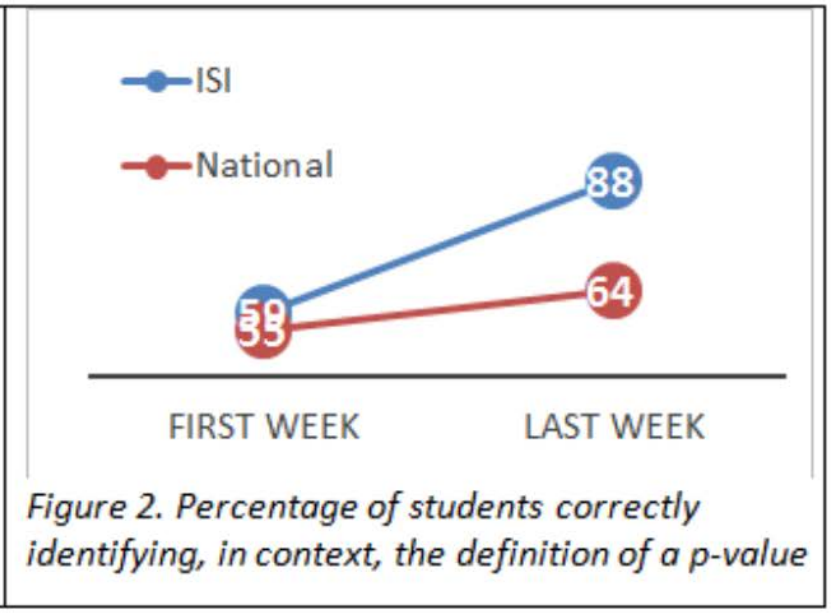


Figure 2. Percentage of students correctly identifying, in context, the definition of a p-value

Assessment results

- Emerging evidence at the HS level (e.g., Roy and McDonnel, 2018)
  - 28 high schools in 16 states; 2016-2017
    - 19 public, 6 private, 1 public charter, 1 private charter, 1 boarding school
    - 24 instructors
    - 630 students
    - Pre-test and post-test data on conceptual assessment

Assessment results

Table 1 – Descriptive statistics on the instructors

	ISI	NonISI	Overall
N	10	14	24
Female instructor (%)	50%	50%	41.7%
Mean number of years of teaching intro statistics (SD)	10.1 (6.2)	11.1 (4.8)	10.7 (5.3)
Whether class was full year	50.6%	95.2%	91.1%

Table 4 – Adjusted mean gain (± SD) separated by curriculum

	ISI	NonISI
n	196	435
Achievable gain	0.214 (0.051)	0.182 (0.058)
Total gain (out of 24 questions)	0.124 (0.025)	0.098 (0.025)
Score on subscale		
Data collection/Scope of conclusions (4 questions)	0.104 (0.041)	0.074 (0.048)
Descriptive statistics (5 questions)	0.077 (0.049)	0.080 (0.056)
Confidence intervals (5 questions)	0.152 (0.036)	0.199 (0.044)
Significance Tests (7 questions)	0.146 (0.037)	0.063 (0.043)*
Simulation/Sampling variability (3 questions)	0.143 (0.051)	0.098 (0.060)

Note: on the subscales, statistics are average adjusted differences in percentage correct (SD); * indicates that the difference is significant at 5% level of significance.

Assessment results

- More data needed
- Digging deeper into existing data needed – digging into learning trajectories
- Better understanding of what's working and what's not
- Informing further improvements to the approach

Summary

- Simulation-based inference increasing in popularity at all levels (MS, HS, post-secondary)
- Increasing evidence of efficacy (assessment data/publications)

Summary

- Some curricular materials available
  - *Statistical Reasoning in Sports* (Tabor and Franklin)
  - *Introduction to Statistical Investigations: AP Edition* (Tintle, Carver, Chance, Cobb, Rossman, Roy, Swanson and VanderStoep) or non-AP edition
    - Key elements
      - Overarching process of statistical reasoning
      - Embracing SBI as a bridge from intuition and to formal inference
      - Guided discovery/active learning pedagogy
      - Online technology – applets, traditional and auto-graded HW, videos, AP exams, etc.

Getting started

- Try one (or two!) days of simulation in place of a traditional approach to inference (or as a complement)
- Try a hands-on simulation in class
- Assess your students to get a baseline sense of where they are on statistical understanding before/after your class
- Jump in! Use one of the full-length curricula available

Acknowledgements

- **Curriculum development team:** Ruth Carver, Beth Chance, George Cobb, Allan Rossman, Soma Roy, and Jill VanderStoep
- **National Science Foundation funding:** This material is based upon work supported from the National Science Foundation under (Grant DUE-1140629) and (Grant DUE-1323210) and (Grant DUE-1612201)

- **Website:** http://www.isi-stats.com/
  - Applets, Free curricular materials, Links to applets and datasets, Links to published assessment findings/publications
  - Link to full curriculum published through Wiley (AP and non-AP course introductory statistics course versions)
- **Blog:** www.causeweb.org/sbi