Teaching Randomization-based Methods in an Introductory Statistics Course: The CATALST Curriculum

Robert delMas
University of Minnesota
NSF DUE-08144333

Workshop materials developed in collaboration with
Andrew Zieffler and Nicola Parker

Overview of Workshop

- Overview/background of CATALST course
- Example course activities (Unit 1)
  - *Matching Dogs to Owners*
- Example course activities (Unit 2)
  - *Sleep Deprivation*
- Example course activities (Unit 3)
  - *Kissing the "Right" Way*

Overview/Background of CATALST

Project Team

University of Minnesota

- Joan Garfield
- Andrew Zieffler
- Robert delMas

Graduate Students

- Rebekah Isaak
- Laura Le
- Laura Ziegler

Collaborators

- Allan Rossman

  California Polytechnic State University, San Luis Obispo
- Beth Chance

  California Polytechnic State University, San Luis Obispo
- John Holcomb

  Cleveland State University
- George Cobb

  Mt. Holyoke College (retired)
- Herle McGowan

  North Carolina State University

Course Pedagogy

- Course activities designed to develop students' statistical reasoning and promote transfer
  - Based on research in cognition and learning
  - Instructional design principles
- Student-centered approach
  - Cooperative groups
- Minimal lectures ("just-in-time" as needed)
  - Small-group and class discussion

Course Emphasis: Inference

The core logic of inference as the foundation (Cobb, 2007)

- **Specify a Model:** Specify a model to reasonably approximate the variation in outcomes attributable to the random process
- **Randomize & Repeat**: Use the model to generate simulated data and collect a summary measure
- **Evaluate**: Examine the distribution of the resulting summary measures

Course Content

- No *t*-tests; Use of probability for simulation and modeling (TinkerPlots™)
- Coherent curriculum that builds ideas of models, chance, simulated data & inference from first day
- Immersion in statistical thinking
- Activities and homework assignments are based on actual research studies using real data

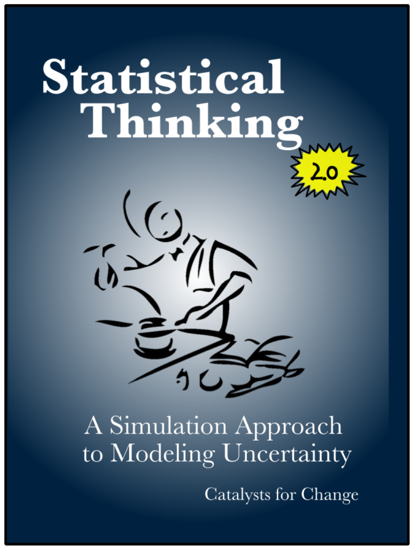
Three Course Units
15 Week Semester

- **Unit 1: Modeling and Simulation**
  - Building and using probability models
  - Learning to use the core logic of inference
- **Unit 2: Comparing Groups**
  - Randomization Tests
  - Design: Random Assignment and Random Sampling
- **Unit 3: Sampling and Estimation**
  - Standard Error of a Sample Statistic (Bootstrapping)
  - Confidence Intervals

TinkerPlots™ Software

- TinkerPlots™ was chosen because of its unique visual capabilities for simulation
  - Allows students to see the random devices they select (e.g., sampler, spinner)
  - Easily use models to simulate and collect data
  - Allows students to visually examine and evaluate distributions of statistics
- The author, Cliff Konold, is currently seeking a new publisher for TinkerPlots™

Resources

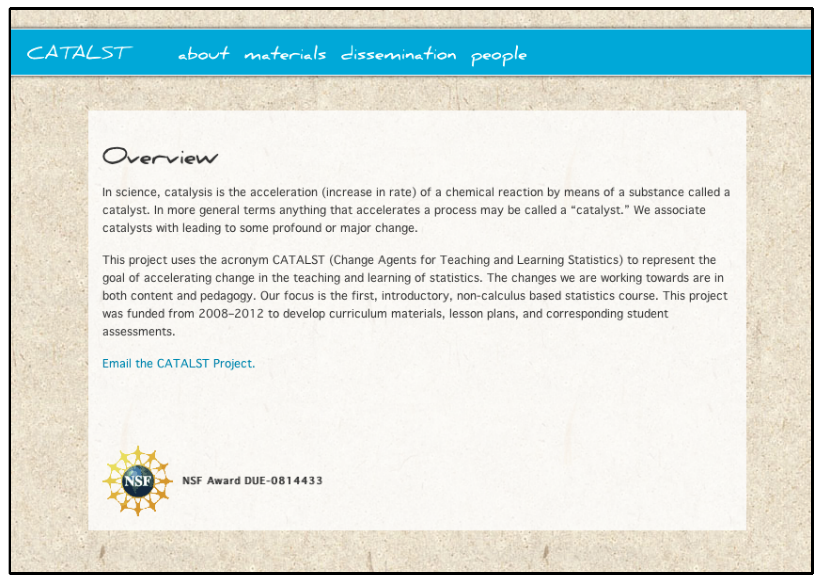


Textbook and data

https://github.com/zief0002/Statistical-Thinking

More information, lesson plans, etc.

http://www.tc.umn.edu/~catalst/



Example Course Activity

Unit 1

Modeling & Simulation

Matching Dogs to Owners

- **Course Location:** Unit 1: Modeling and Simulation
- **Learning Goals:** (1) Begin to develop understanding of the reasoning process of statistical inference; (2) Understand that if an observed result is very unlikely under a particular model, then the result provides strong evidence against that particular model
- **Preparation:**
  - *Introduction to Statistical Hypothesis Testing* [textbook]

Matching Dogs to Owners

- Now please put on your "student hat"!
- Afterward, you will get to put your "teacher hat" back on

Research Question

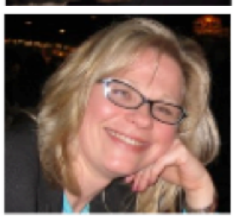
Are humans able to match dogs to their owners better than blind luck?

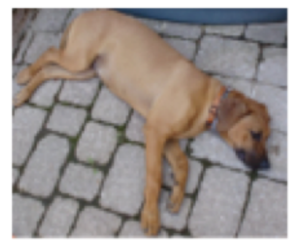
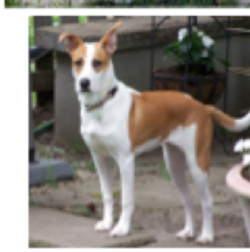
Matching Dogs to Owners

- On a piece of paper, create a two column grid.
- First column OWNER, second column DOG
- Write the numbers 1 through 6 under OWNER

OWNER	DOG
1	
2	
3	
4	
5	
6	

- Match the dogs to the owners as best as you can.

1. 
2. 
3. 
4. 
5. 
6. 

1. 
2. 
3. 
4. 
5. 
6. 

Matching Dogs to Owners

- READY ?.... here come the answers.....

OWNER

1

DOG

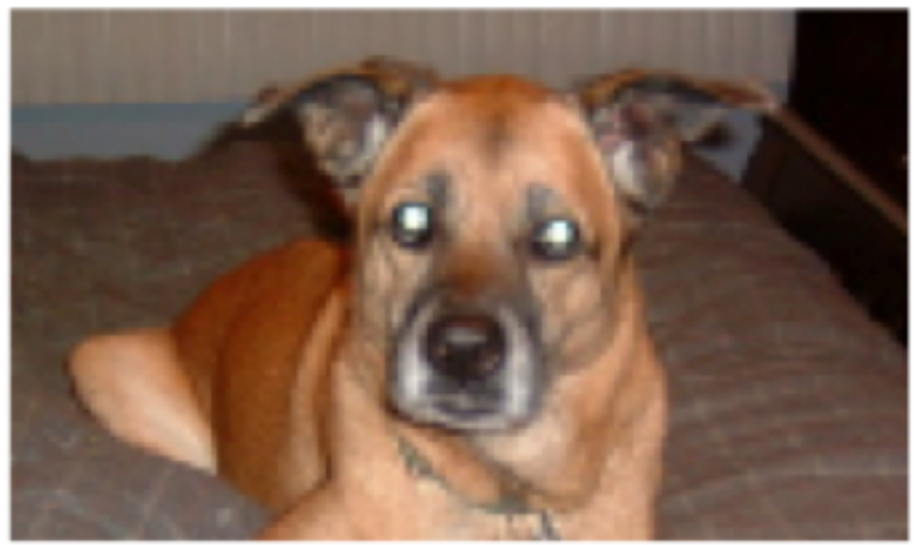
1

OWNER

2

DOG

3



OWNER

3

DOG

6

OWNER

4

DOG

5

OWNER	DOG
5	2

OWNER

6

DOG

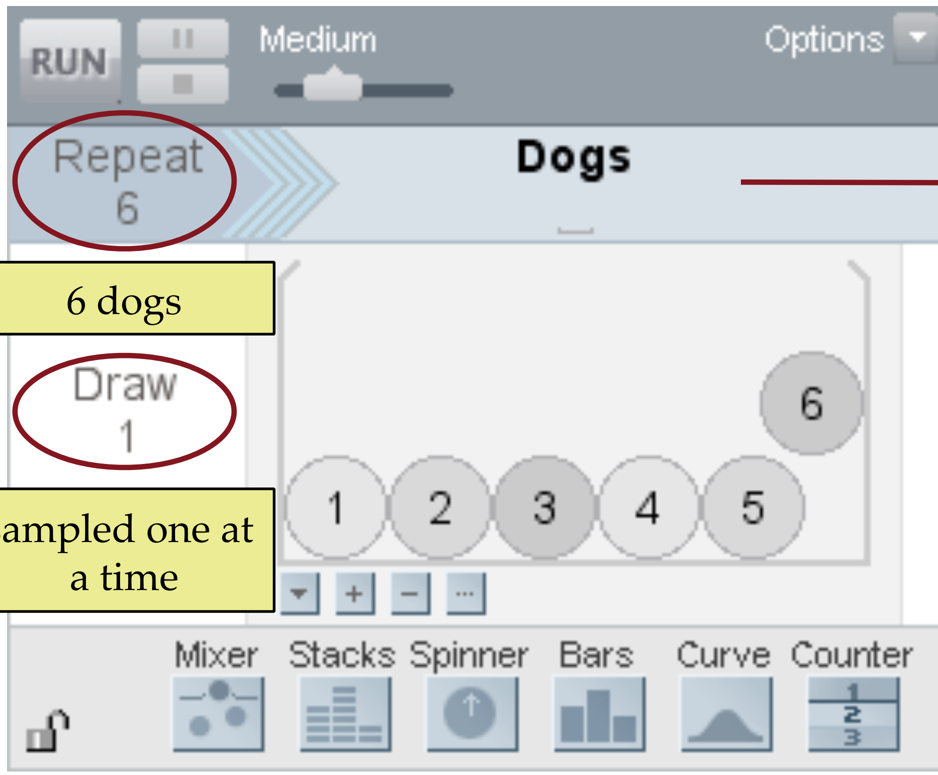
4

Matching Dogs to Owners

Now, we will look at how TinkerPlots™ can be used to shuffle the dogs and match them to the owners randomly.

Model

The dogs are selected randomly.
ASSUMPTION: humans have no ability to match dogs with owners (blind guessing).

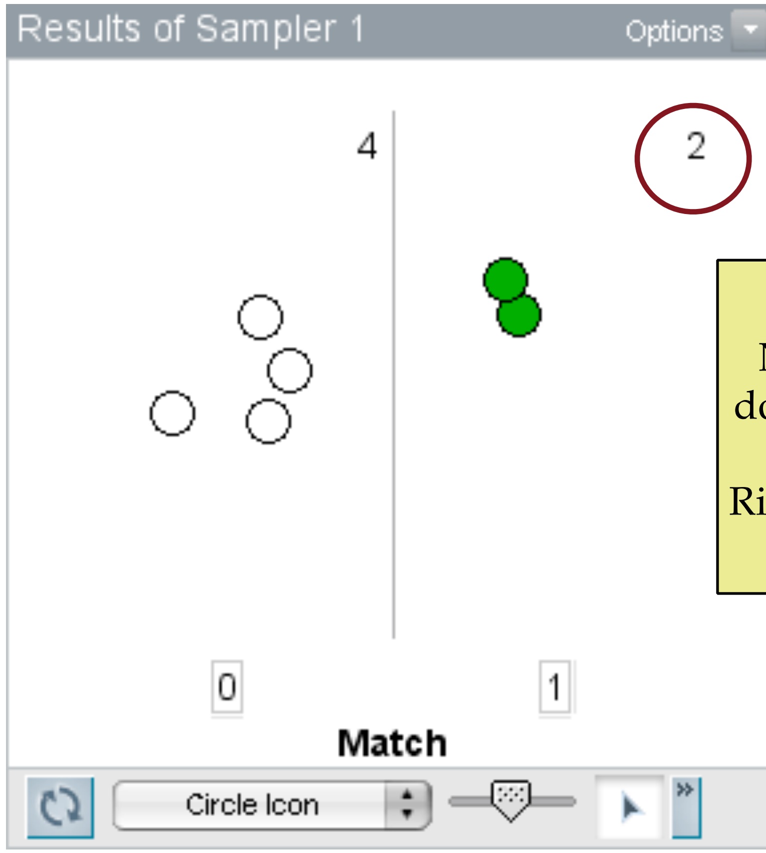


6 dogs

sampled one at a time

Sampling without replacement.

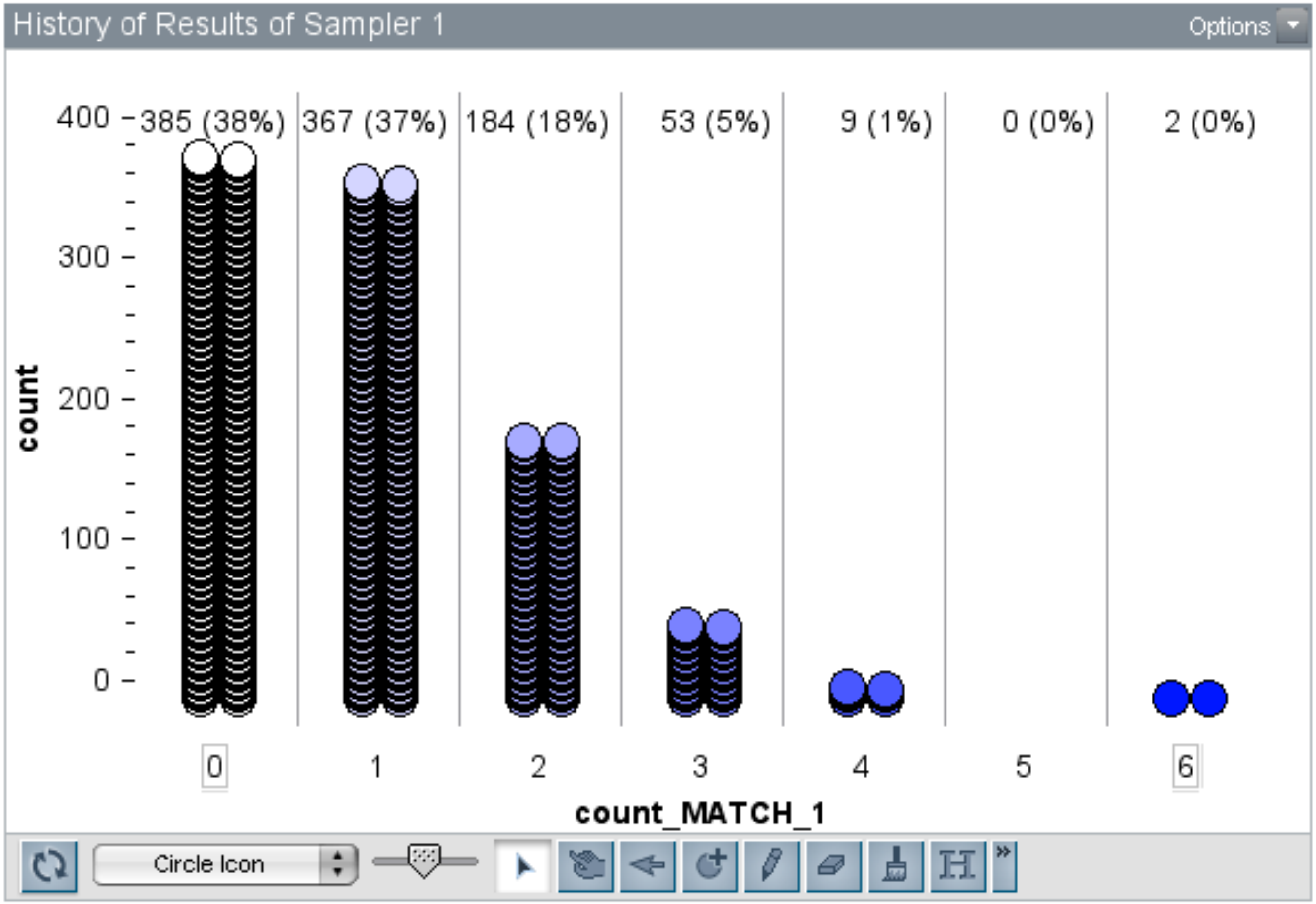
Simulate



Trial Result
Number of matches between dog number and owner number

Right-click on the Trial Result to "Collect" it

Evaluate



Which results are likely or unlikely under the null model?

Matching Dogs to Owners WRAP UP

- What does the “blind guessing” model mean?
- Why do we use the “blind guessing” model to simulate our data?
- What does it mean to have evidence that “supports” or “doesn’t support” the blind guessing model?

Matching Dogs to Owners

REFLECTION

Teacher hats back on!

- What about this activity (content, format…) do you think might help maximize student learning in your classroom?
- What are your hesitations / what do you think might hinder student learning in your classroom? What questions do you still have about the implementation of such a course?
- Presuming that you *wanted* to implement these activities in your courses, how comfortable would you feel doing so? Why or why not?

Example Course Activity
Unit 2
Comparing Groups

Sleep Deprivation

- **Course Location:** Unit 2: Comparing Groups
- **Learning Goals:**
  1) Compare two groups with quantitative outcome variable using TinkerPlots™ to carry out a randomization test;
  2) Find and use the *p*-value from the randomization test to draw a conclusion
- **Preparation:**
  - Stickgold, R., James, L., & Hobson, J. A. (2000). Visual discrimination learning requires sleep after training. *Nature Neuroscience, 3*(12), 1237–1238.
  - *Mission Improbable: A Concise and precise definition of p-value.* [web article]

Research Question

Does the effect of sleep deprivation last, or can a person "make up" for sleep deprivation by getting a full night's sleep on subsequent nights?

brief communications

Visual discrimination learning requires sleep after training

Robert Stickgold, LaTanya James and J. Allan Hobson

Laboratory of Neurophysiology, Department of Psychiatry, Harvard Medical School, 74 Fenwood Road, Boston, Massachusetts 02115, USA

Correspondence should be addressed to R.S. (rstickgold@hms.harvard.edu)

Performance on a visual discrimination task showed maximal improvement 48–96 hours after initial training, even without intervening practice. When subjects were deprived of sleep for 30 hours after training and then tested after two full nights of recovery sleep, they showed no significant improvement, despite normal levels of alertness. Together with previous findings[11] that subjects show no improvement when retested the same day as training, this demonstrates that sleep within 30 hours of training is absolutely required for improved performance.

Skill learning represents one of several classes of procedural memory, and is defined as experience-dependent improvement in

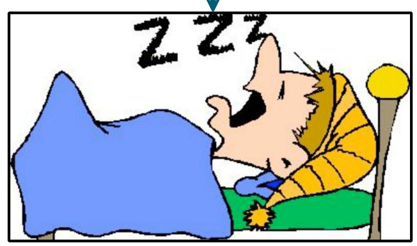
was deprived of sleep from the time of testing until 9:00 p.m. the following evening, and was then allowed unrestricted sleep on the second and third nights, before retesting after 72 hours (**Fig. 1**). Subjects were monitored during the period of sleep deprivation using the Nightcap ambulatory vigilance monitor, a wallet-sized recording device that accurately identifies periods of wake and sleep[13].

In the visual discrimination task[8,11], a target screen was displayed for 17 ms, followed by a blank screen for a variable interstimulus interval (ISI), and then a mask, also displayed for 17 ms. The target screen consisted of three diagonal bars in one quadrant of the screen, in either a vertical or horizontal array, displayed against a background of horizontal bars, with the letter 'T' or 'L' displayed at the fixation point (see web supplement, http://neurosci.nature.com/web_specials/). After presentation of the mask, subjects were asked to determine whether the fixation letter had been a 'T' or an 'L' and whether the array of diagonal bars had been vertically or horizontally oriented. Subjects were tested over a range of interstimulus intervals, and the minimum ISI required to reach a threshold accuracy on the horizontal–vertical discrimination task of 80% was determined. Improvement for a subject was defined as the decrease in threshold ISI at retest compared to training. Training and test sessions each contained 1250 trials in 25 blocks. To avoid floor effects, subjects with

Study Design

Volunteers between the ages of 18 and 25 trained and tested on a visual discrimination task

Night following training, subjects randomly assigned to have **unrestricted sleep** or to be **deprived of sleep**



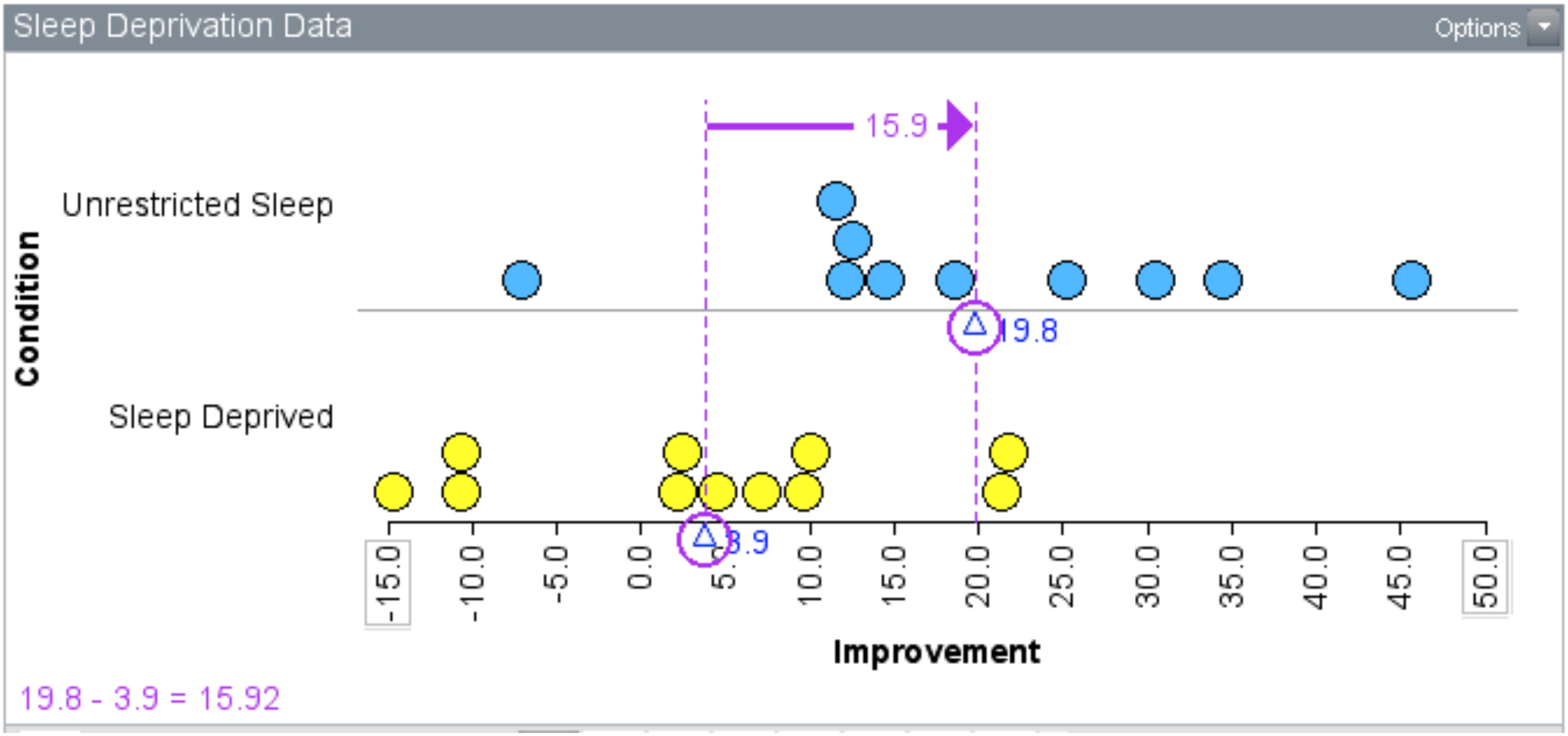
($n = 10$)

($n = 11$)

Next two nights subjects were allowed as much sleep as they wanted then they were re-tested

How were the groups determined: Random assignment!

Observed Difference

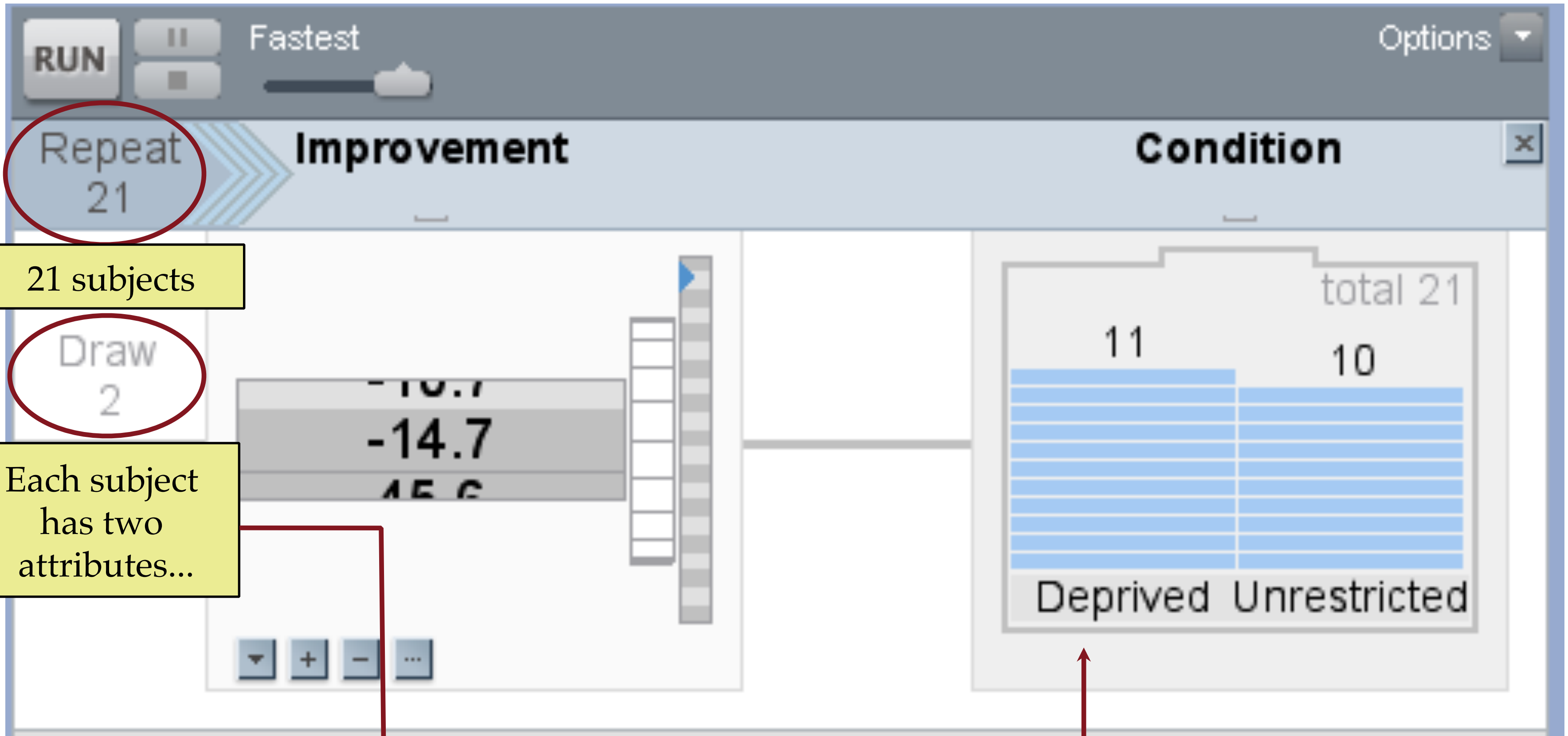


Could the observed sample mean difference of 15.92 have happened just by chance?

Sleep Deprivation

- Similar to activity that students complete early in the course
- Again, put on your "student hat"!
- Afterward, you will get to put your "teacher hat" back on

Model



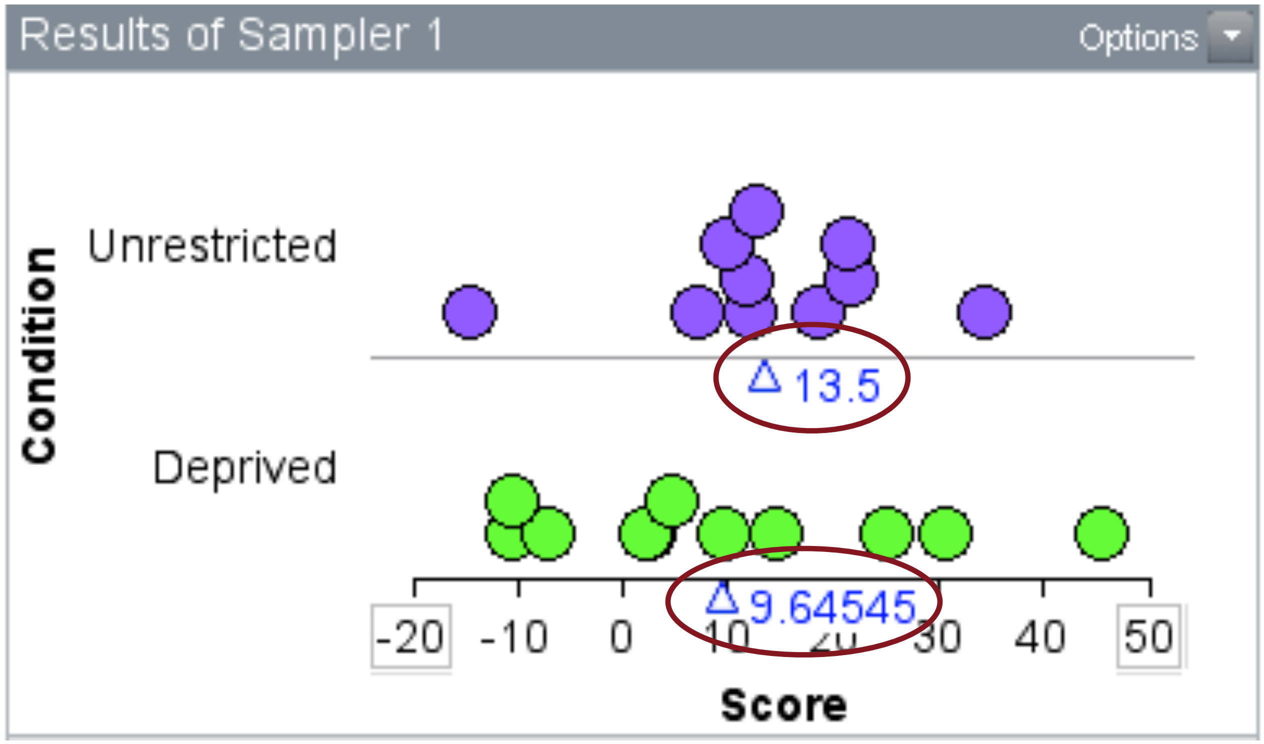
21 subjects

Each subject has two attributes...

Fixed Response

Condition
Randomly assigned.
Sampling without replacement.

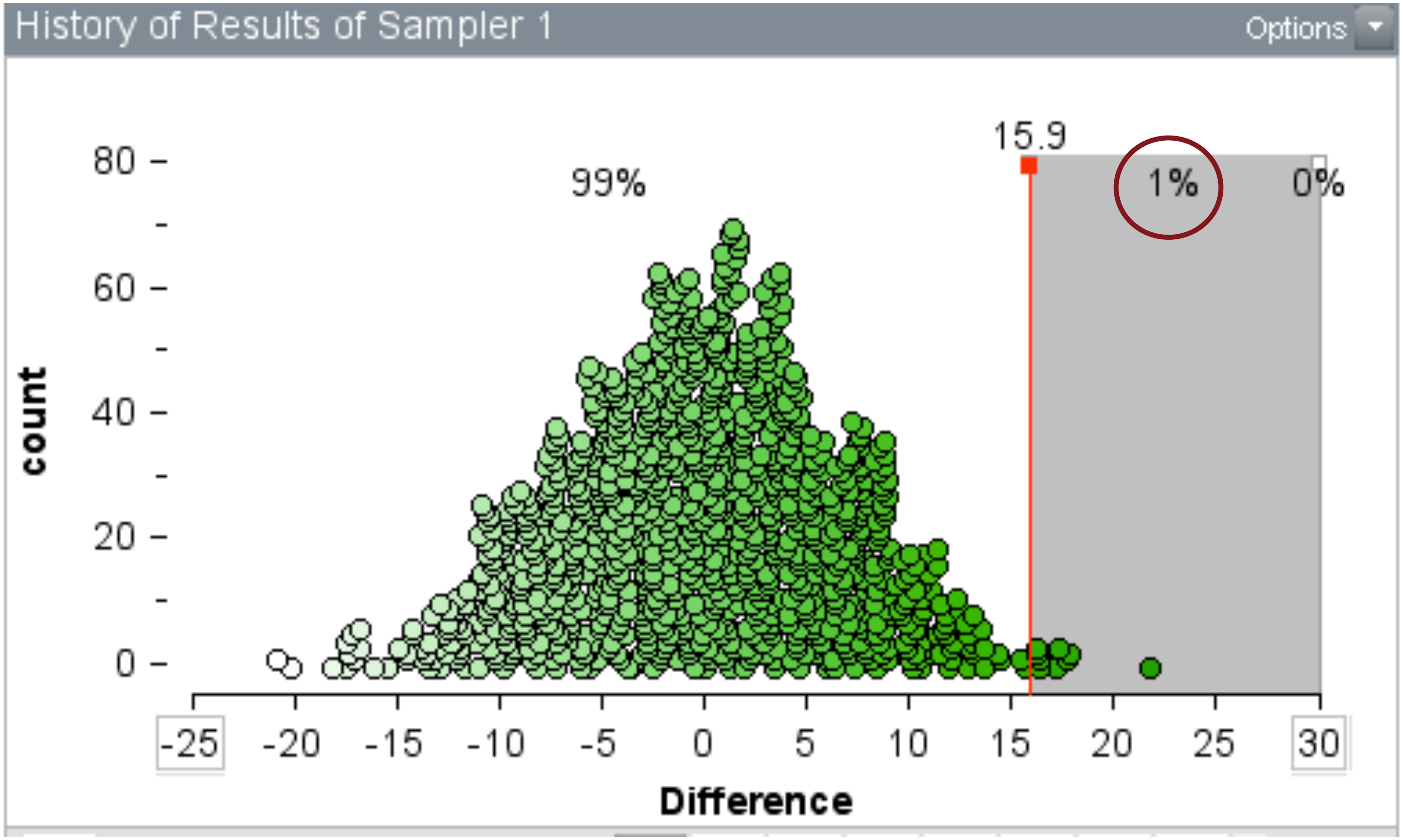
Simulate



Trial Result
Difference between treatment and control is 3.9.

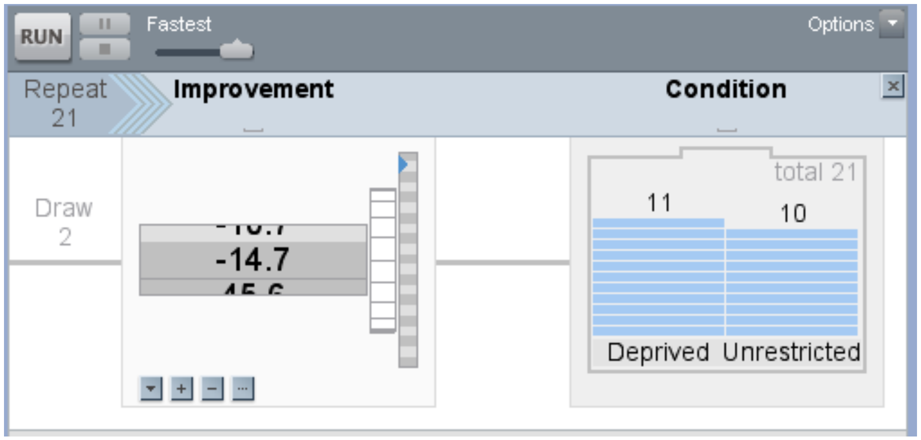
Collect the Statistics
The mean Improvement for both Control & Treatment

Evaluate

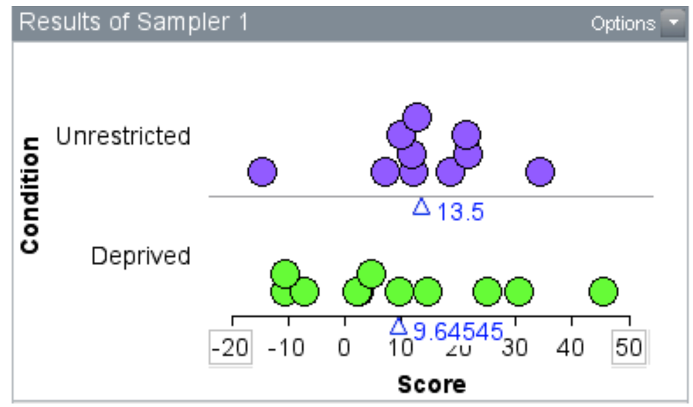


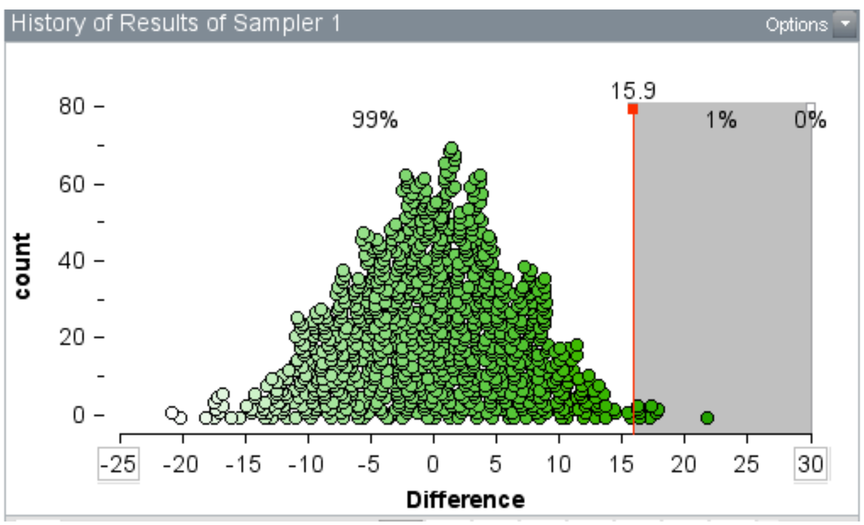
What proportion of results are *at least* as extreme as the observed result of 15.9?

Model: There is no effect of unrestricted sleep.



Simulate: Carry out the random assignment and compute the mean difference.





Evaluate: Under the model, the observed result of 15.9 (or a more extreme result) is very unlikely to have occurred by chance.

Sample Wrap-Up Questions

- What is the purpose of random assignment in this study?
- Why do we need to conduct a test? Why can't we just answer the research question using the differences found in the observed data?
- What was the model (null model) used to simulate the data?
- Where was the plot of the simulated differences centered (location)? Why does that location make sense?
- What was the *p*-value you obtained? What does this suggest about the answer to the research question?

Example Course Activity

Unit 3

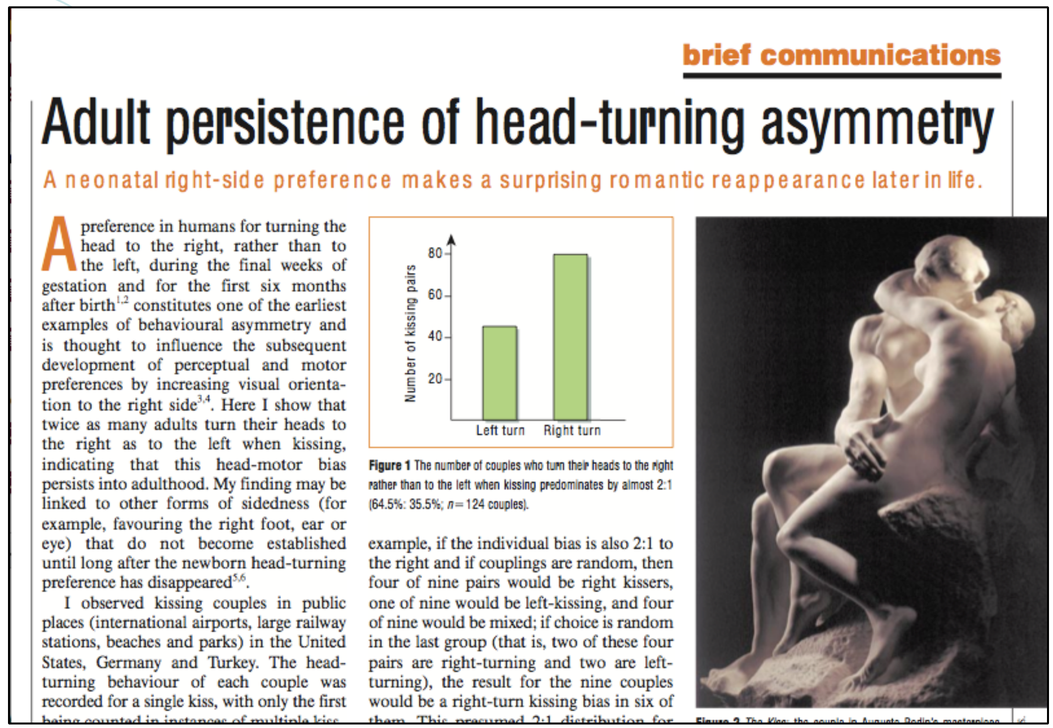
Sampling & Estimation

Kissing the "Right" Way

- **Course Location:** Unit 3: Sampling and Estimation
- **Learning Goals:** (1) Use TinkerPlots™ to carry out a bootstrap analysis to estimate a standard error; (2) Compute an interval estimate for a proportion
- **Preparation:**
  - *Standard Deviation Tutorial*. [web article]

Research Question

What percentage of couples lean their heads to the right when kissing?



brief communications

Adult persistence of head-turning asymmetry

A neonatal right-side preference makes a surprising romantic reappearance later in life.

A preference in humans for turning the head to the right, rather than to the left, during the final weeks of gestation and for the first six months after birth[1,2] constitutes one of the earliest examples of behavioural asymmetry and is thought to influence the subsequent development of perceptual and motor preferences by increasing visual orientation to the right side[3,4]. Here I show that twice as many adults turn their heads to the right as to the left when kissing, indicating that this head-motor bias persists into adulthood. My finding may be linked to other forms of sidedness (for example, favouring the right foot, ear or eye) that do not become established until long after the newborn head-turning preference has disappeared[5,6].

I observed kissing couples in public places (international airports, large railway stations, beaches and parks) in the United States, Germany and Turkey. The head-turning behaviour of each couple was recorded for a single kiss, with only the first

Figure 1 The number of couples who turn their heads to the right rather than to the left when kissing predominates by almost 2:1 (64.5%: 35.5%; $n = 124$ couples).

example, if the individual bias is also 2:1 to the right and if couplings are random, then four of nine pairs would be right kissers, one of nine would be left-kissing, and four of nine would be mixed; if choice is random in the last group (that is, two of these four pairs are right-turning and two are left-turning), the result for the nine couples would be a right-turn kissing bias in six of them.

Study Design

Observation of 124 couples between the ages of 13 and 70

Couples were observed in public places (e.g., airports, beaches, train stations, parks) and were not holding any luggage

Direction of head lean was recorded.

44 Couples
Lean Left

80 Couples
Lean Right

How were the groups determined: Self-selection

Bootstrapping

What percentage of all couples lean to the right?

64.5% of the couples observed leaned to the right. This is a good initial guess (a point estimate).

Problem: A different sample of 124 couples would have given a slightly different answer to the question. How do we account for this?

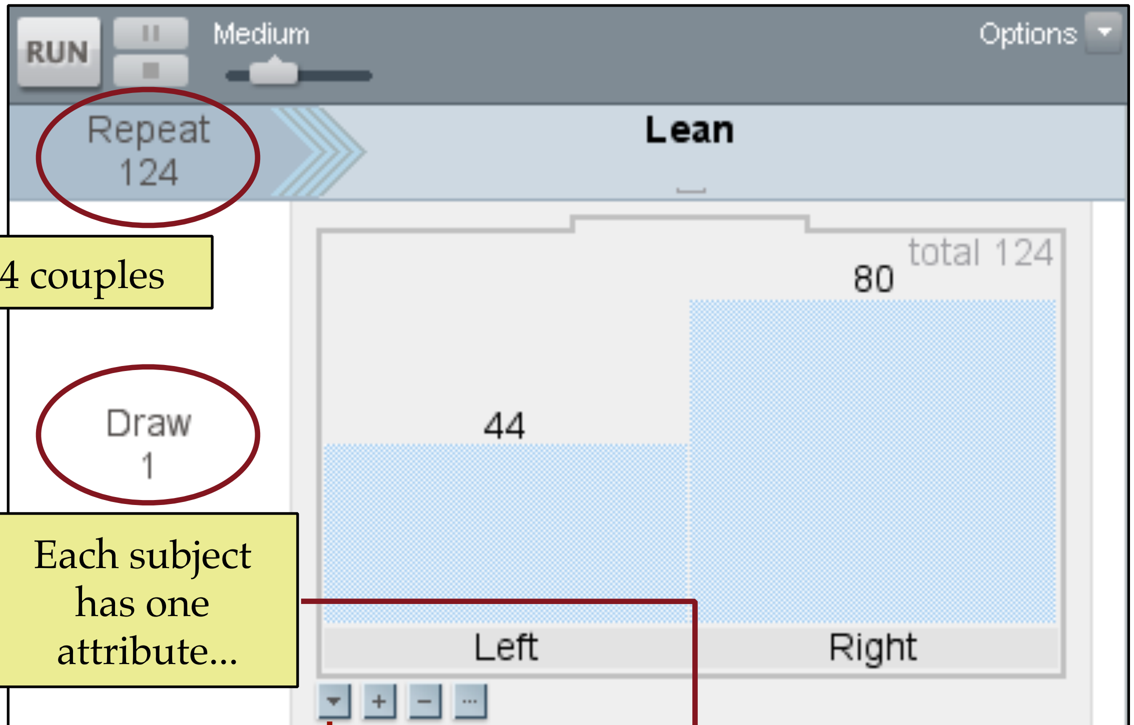
Statistical Problem: How much variation is there in the estimate from sample-to-sample?

What we would *like* to do...

Re-sample 124 couples from the population, observe the percentage that lean to the right when kissing, do this many times, compute the amount of variation in the estimate...*but we do not have access to the population* (if we did, why estimate?)

Solution: Use sample as a substitute for the population, and resample from the sample (with replacement, of course!)

Model



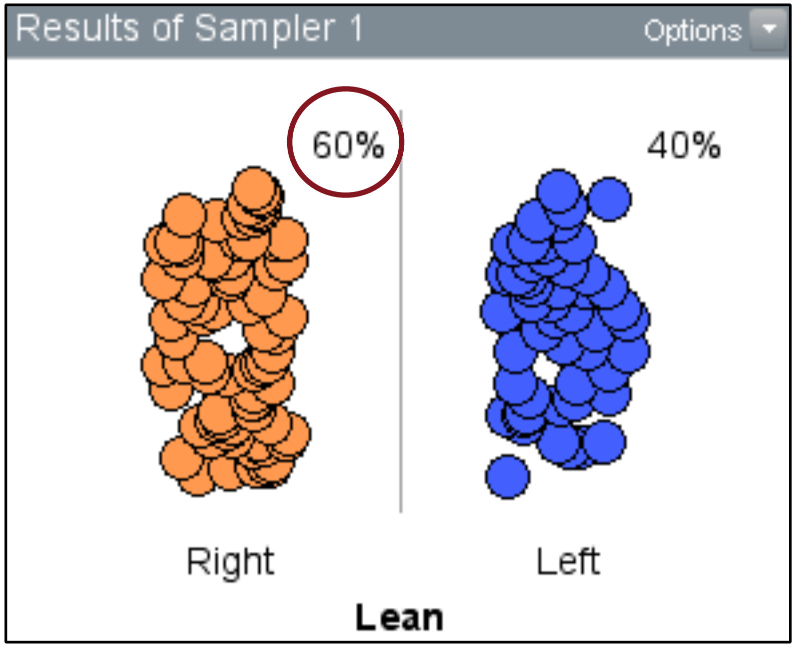
124 couples

Each subject has one attribute...

WITH REPLACEMENT

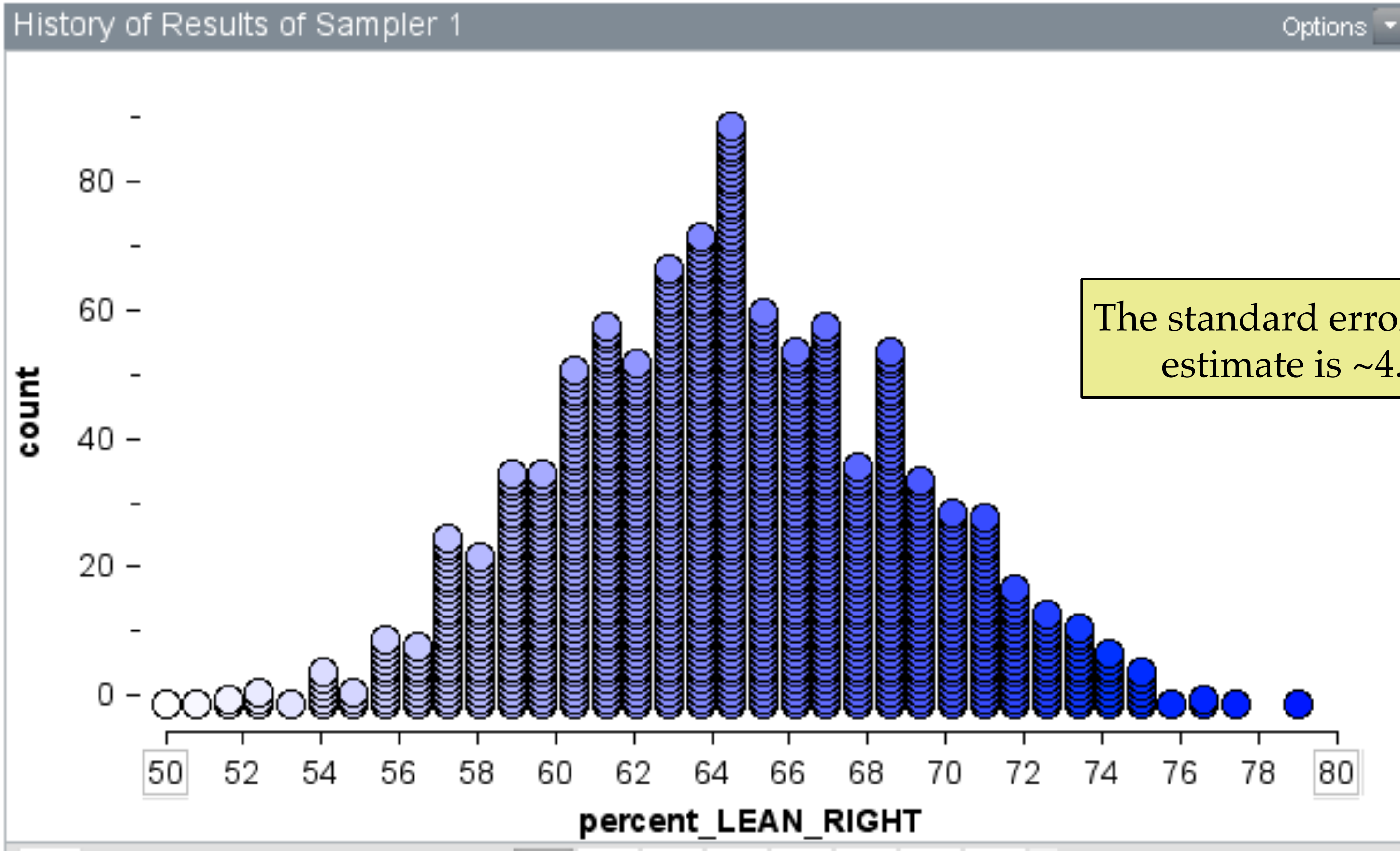
Lean Direction
Randomly sampled WITH replacement.

Simulate



Trial Result
60% of the couples in the bootstrapped sample lean to the right when kissing.

Evaluate



The standard error of the estimate is ~4.527

The percentage of couples who lean to the right when kissing is between 55.45% and 73.55%. (64.5% ± 9.05)

Sample Wrap-Up Questions

- What is the standard error?
- What is the margin of error?
- How would you interpret the margin of error in this context? In general?
- What is the purpose of an interval estimate?
- How would you interpret your interval estimate?

CATALST Curriculum Page

http://www.tc.umn.edu/~catalst/

Aids for Students' Learning

- TinkerPlots™ Instructions
- Visualizations
- Homework Assignments
- Readings

Aids for Teaching

- Lesson Plans

Questions & Discussion

Thank You!

If you have any questions about the CATALST course, please contact anyone of the PIs:

Joan Garfield (jbg@umn.edu)

Robert delMas (delma001@umn.edu)

Andrew Zieffler (zief0002@umn.edu)

CATALYST

CHANGE AGENTS FOR TEACHING AND LEARNING STATISTICS

Accelerating the change of content and pedagogy in introductory statistics