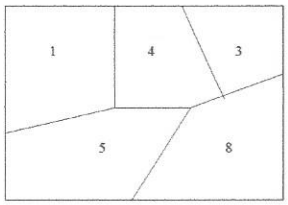
What Every High School Teacher Needs to Know about Statistics

Roxy Peck
Cal Poly, San Luis Obispo
rpeck@calpoly.edu

NCTM 2105

1

Consider the map of counties shown below. The number in each county is last month's incidence rate for a disease in cases per 100,000 population.
(Dick Schaeffer, 2005)



NCTM 2105

2

Statistical Versus Mathematical Thinking

- Mathematical Thinking
  - Explain patterns
  - Often a deterministic way of thinking
- Statistical Thinking
  - Search for patterns in the presence of variability
  - Acknowledge role of chance variation
- Statisticians ask:
  Could this have happened by chance?

NCTM 2105

3

Statistical Thinking

- *Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.'*
  Samuel S. Wilks, 1951
- I keep saying the sexy job in the next ten years will be statisticians… The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it—that's going to be a hugely important skill in the next decades, not only at the professional level but even at the educational level for elementary school kids, for high school kids, for college kids. Because now we really do have essentially free and ubiquitous data. So the complimentary scarce factor is the ability to understand that data and extract value from it.
  Hal Varian, Chief Economist, Google 2009

NCTM 2105

4

That Day is Here! BUT…

- Statistics recommended in HS curriculum for many years
  - *Curriculum and Evaluation Standards* (NCTM 1989)
  - *Principles and Standards for School Mathematics* (NCTM 2000)
- Inconsistent implementation of these standards
- **Common Core State Standards in Mathematics** give statistics a more prominent role

NCTM 2105

5

And Still…

- Statistics often presented as a loose collection of graphical and numerical methods
- Little or no underlying theory, applications, or connections between concepts
- This is at Odds with Common Core State Standards, which
  - go beyond mechanical and computational aspects
  - include a focus on conceptual understanding necessary for sound statistical reasoning

NCTM 2105

6

A Challenge, An Opportunity, and A Concern

- Challenge: Implementing a curriculum for the statistics and data standards of the common core will be challenging for teachers who have not previously been expected to include this material.
- Opportunity: We have an opportunity to help students develop important statistical thinking skills.
- Concern: Taking the easy way out, by focusing on what is easy to teach—the mechanics and computations—rather than focusing on developing conceptual understanding and statistical thinking.

NCTM 2105

7

Exploring Four Essential Understandings

- The difference between statistical thinking and mathematical thinking.
- The role of variability.
- The difference between sample variability and sampling variability.
- The need to rule out chance.

NCTM 2105

8

The Role of Variability

- Variability plays a role in each phase of the data analysis process:
  - Pose a question that can be answered by collecting data.
    Data collection plan must anticipate variability in the data.
  - Describe the data distribution.
    Summarizing data numerically and graphically involves describing variability in the data distribution.
  - Analyze the data.
    Data analysis must take variability into account.
  - Interpret results in context and draw conclusions based on data,
    Requires acknowledging role of variability in the data (sample variability) and the role of sampling variability.

NCTM 2105

9

Distributions Describe Variability

Being able to think about data in terms of distributions and to distinguish between the different ways distributions are used (to describe the variability in a population, to describe variability in a sample, and to describe variability in the values of a statistic for different possible samples) are key to understanding statistical inference.

NCTM 2105

10

Life would be simple if…

- there was no variability!
  - Simple because it would be easy to draw conclusions based on data.
  - Simple, but BORING!
- As Statistics people we
  - LOVE variability
  - Are OK with being wrong 5% of the time!

Many math folks don't get this!

NCTM 2105

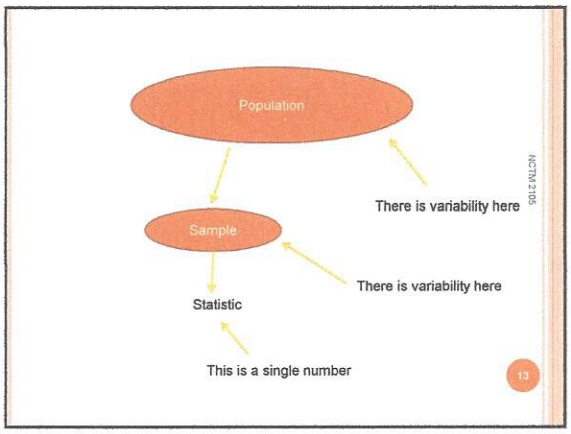
11

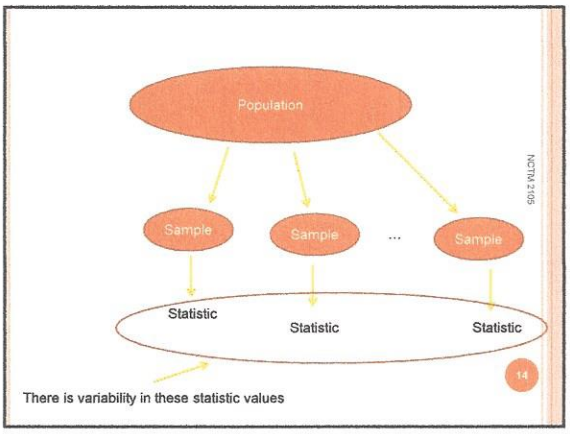
Distributions Describe Variability

- Variability in a population
- Variability in a sample
- Sample-to-sample variability in the values of a statistic

NCTM 2105

12





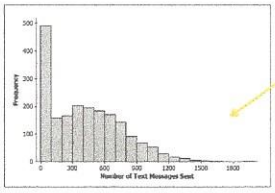
Distributions Describe Variability

- Variability in a population
  - Population Distribution
- Variability in a sample
  - Sample Distribution
- Sample-to-sample variability in the values of a statistic
  - Sampling Distribution

NCTM 2105
15

Example

- Population of 2000 students
- Distribution of number of text messages sent in one month

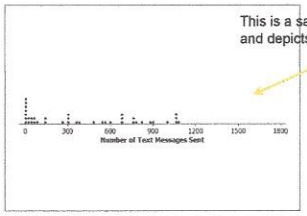


This is a population distribution

NCTM 2105
16

Example continued

- Random sample of 40 students



This is a sample distribution and depicts SAMPLE VARIABILITY

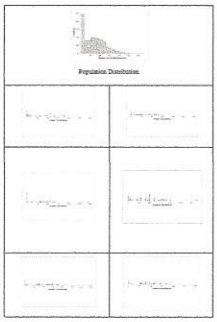
NCTM 2105
17

Example continued

IF a sample is selected at random from the population, we expect the sample distribution to resemble the population distribution

NCTM 2105
18

Sample-to-Sample Variability



19

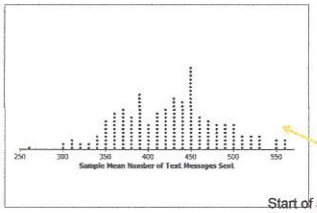
Example Continued

For this population and for these six samples:

Population mean	420.6
Sample 1 mean	422.2
Sample 2 mean	418.2
Sample 3 mean	411.7
Sample 4 mean	564.2
Sample 5 mean	367.2
Sample 6 mean	484.7

20

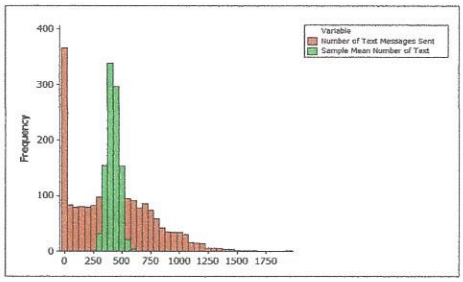
For 200 Random Samples



Start of a Sampling Distribution
This reflects SAMPLING VARIABILITY

21

For 1000 Random Samples



22

Helping Students Develop Statistical Thinking

- Learning to ask: Could this have happened by chance?
  - A Simple Game
  - A Nurse Accused

23

A Simple Game

(with apologies to those who have seen this before)

24

A Simple Game

- Statistical Thinking Involved
  - Could this have happened by chance if…
  - Convincing evidence versus proof
  - Acknowledging the risk of an incorrect conclusion

NCTM 2105

25

A Nurse Accused

- From *Statistics: A Guide to the Unknown*
  - *Kristen Gilbert was a nurse at the Veteran's Administration (VA) hospital in Northampton, Massachusetts.*
  - *She had a reputation of being particularly good in a crisis. If a person went into cardiac arrest, she was often the first to notice something was wrong, signal code blue and administer epinephrine. Often the patient's life was saved.*
  - *Other nurses became suspicious both because of the number of cardiac arrests when Gilbert was on duty and also the number of deaths.*
  - *Eventually, Gilbert was accused and prosecuted for the murder of a number of patients by administering a powerful heart stimulant.*

NCTM 2105

26

The Evidence

	Death on Shift		
Gilbert Present	Yes	No	Total
Yes	40	217	257
No	34	1350	1384
Total	74	1567	1641

Proportion of shifts with a death: 74/1641 = 0.045
Proportion of Gilbert shifts with a death: 40/257 = 0.156

NCTM 2105

27

Could this have happened by chance??

- If probability of a death on a shift is 0.045, what is the chance that Gilberts 257 shifts would have 40 deaths?
- Could do a physical simulation using 955 white beads and 45 red beads.
  - Mix beads and select one at random to represent a shift. If the selected bead is white, no death occurred. If the selected bead is red, a death occurred. Replace the selected bead.
  - Continue for a total of 257 "shifts", counting number of shifts with a death.
  - Repeat whole process a large number of times to get a sense of likely and unlikely values for number of shifts with a death out of 257 shifts if probability of a death is the same for Gilbert shifts as it is for all shifts..
  - Ask question, is it likely that we would observe 40 or more shifts with a death? Could this have happened by chance?

NCTM 2105

28

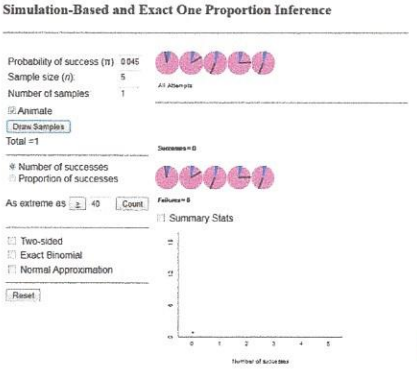
Let's investigate the Gilbert data

- Technology

http://www.rossmanchance.com/applets/OneProp/OneProp.htm

NCTM 2105

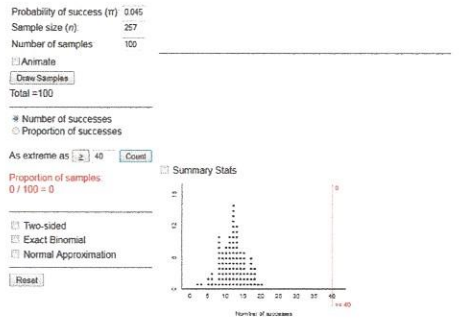
29



NCTM 2105

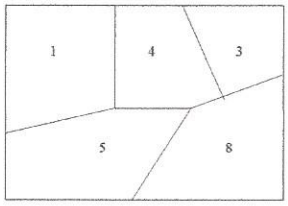
30

Simulation-Based and Exact One Proportion Inference



NCTM 2105

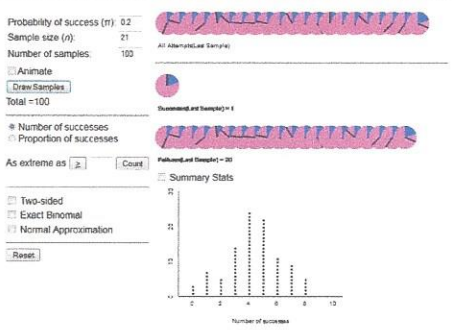
REVISITING THE FIRST EXAMPLE



NCTM 2105

32

Simulation-Based and Exact One Proportion Inference



NCTM 2105

RECAP: FOUR ESSENTIAL UNDERSTANDINGS

- The difference between statistical thinking and mathematical thinking.
- The role of variability.
- The difference between sample variability and sampling variability.
- The need to rule out chance.

NCTM 2105

34

THANK YOU

- Thank you for attending this talk.
- Questions or comments to

  rpeck@calpoly.edu

NCTM 2105

35