

# The Changing Face of Statistics: Implications for our Classrooms

## Teaching Inference in the Common Core

Katherine Halvorsen

Department of Mathematics and Statistics

Smith College, Northampton, MA

# The Structure of the Common Core Standards for Mathematics:

- **Domains:** The largest grouping or related standards
- **Clusters:** Groups of related standards
- **Standards:** What students should understand and be able to do.

# The Domains and Clusters for High School Statistics & Probability

- Interpreting Categorical and Quantitative Data
  - Summarize, represent, and interpret data on a single count or measurement variable
  - Summarize, represent, and interpret data on two categorical and quantitative variables
  - Interpret linear models
- **Making Inferences and Justifying Conclusions**
  - **Understand and evaluate random processes underlying statistical experiments**
  - **Make inferences and justify conclusions from sample surveys, experiments and observational studies**
- Conditional Probability and the Rules of Probability
  - Understand independence and conditional probability and use them to interpret data
  - Use the rules of probability to compute probabilities of compound events in a uniform probability model
- Using Probability to Make Decisions
  - Calculate expected values and use them to solve problems
  - Use probability to evaluate outcomes of decisions

# Domain: Making Inferences and Justifying Conclusions

## Cluster: Understand and evaluate random processes underlying statistical experiments

- S-IC.1. Understand statistics as a process for making inferences about population parameters based on a random sample from that population.
- S-IC.2. **Decide if a specified model is consistent with results from a given data-generating process, e.g., using simulation.**

# **Cluster: Making inferences and justifying conclusions from sample surveys, experiments and observational studies**

- S-IC.3. Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.
- S-IC.4. Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling.
- **S-IC.5. Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.**
- S-IC.6. Evaluate reports based on data.

# How randomization is used in making inferences from data

- Random sampling from a population for a survey allows valid conclusions to be drawn from sample to population
- Random assignment of subjects to treatments for an experiment creates a fair comparison of treatment effectiveness
- Randomization tests allow us to judge how unusual a result would be if a given model holds

# The concept of a statistical test

- An hypothesis test is an inferential method for assessing evidence provided by data to decide whether the data supports or fails to support some claim (hypothesis) about a characteristic (parameter) of the population.

# Example: In a criminal trial in a U.S. court

- The proceedings start with the **null hypothesis** that the defendant is innocent
- The **alternative hypothesis** is that the defendant is guilty
- The prosecution provides evidence against the hypothesis
- The defense provides evidence for the hypothesis
- The jury must be convinced “beyond a reasonable doubt” (their  **$\alpha$ -level**) that the hypothesis is false.

How do we convince ourselves that the difference in outcomes between two treatments or conditions is “statistically significant?”

- The Null Hypothesis: The outcomes of the two treatments do not differ on average
- The Alternative Hypothesis: The outcomes are significantly different

**Example:** A poll for Life magazine entitled *If Women Ran America* reports that two thirds of the women interviewed say that the problem of unequal pay for equal work is a serious one while only half of the men have the opinion. Suppose the survey is based on 44 people: 20 men and 24 women.

The table shows outcome of the Life Magazine poll.

	Yes	No	Total
Women	16	8	24
Men	10	10	20
Total	26	18	44

- What are the observed proportions of men and women who say yes in this sample?
- For women the proportion saying YES = .67
- For men the proportion saying YES = .50
- Given this sample, do you believe there is a difference between the way men and women feel about the issue of unequal pay for equal work?

To construct a statistical test, first state the hypotheses to be tested:

- **NULL HYPOTHESIS:** men and women feel the same way about the issue (If this is true in the population, the proportions who say YES would be expected to be about the same)
- **ALTERNATIVE HYPOTHESIS:** men and women feel differently about the issue (if this is true in the population, the proportions who say YES would be expected to differ in the two groups)

How different do the two sample proportions have to be to make us doubt the null hypothesis?

We use simulation to provide some evidence about this question.

# Assume the marginal totals are given:

- Suppose the numbers of men (20) and women (24) that we survey are fixed, and that the total numbers of yes (26) and no (18) counts are also fixed.
- What other table counts might have occurred given these fixed marginal totals?

Is this table possible? Are there others?

	Yes	No	Total
Women	24	0	24
Men	2	18	20
Total	26	18	44

# Choose a test statistic

- Notice that a consequence of fixed margins is that by choosing the number of women who say Yes the rest of the table entries are determined
- We use this fact to compare tables. The number of women who say Yes is our **test statistic**

Use a deck of cards to simulate this poll.

- Count out 24 red cards to represent the women in your sample
- Count out 20 black cards to represent the men in your sample.

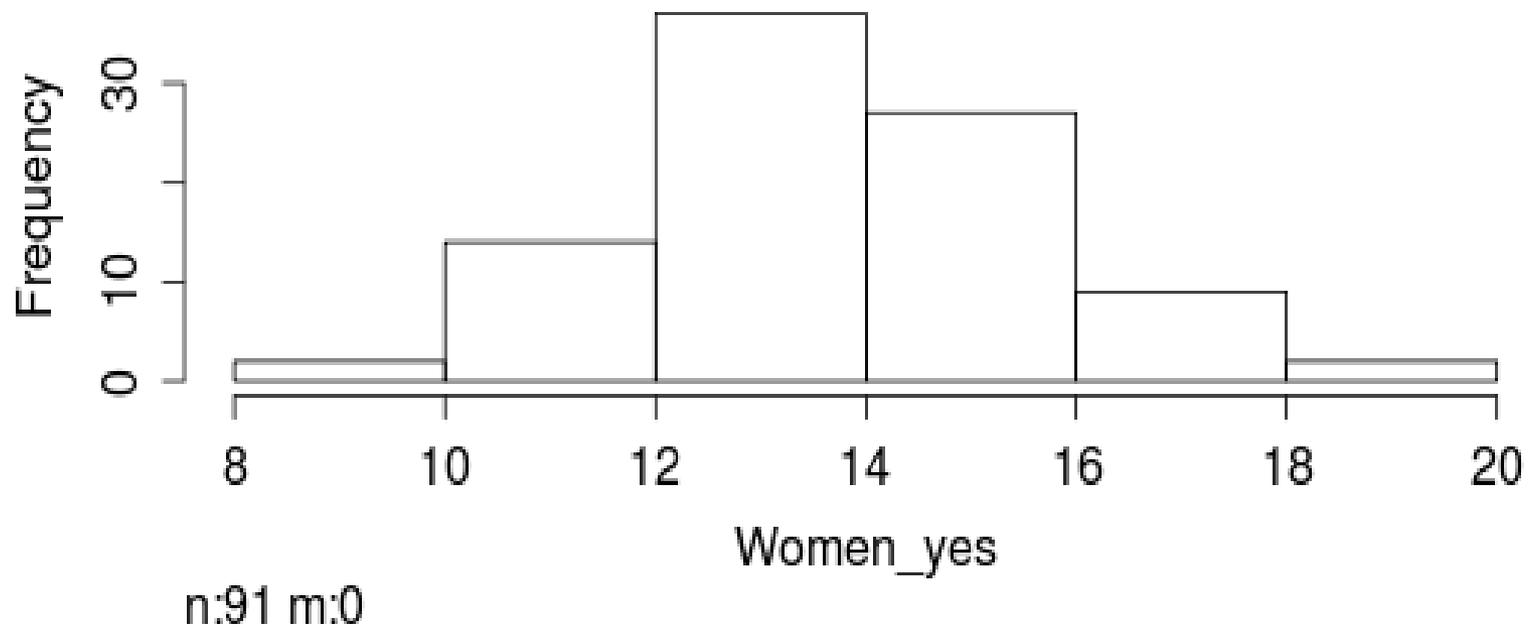
Generate data you would expect to see if the null hypothesis were true:

- By shuffling these 44 cards and dealing out 26 cards we can simulate **randomly selecting** the people who said yes...without regard to their sex.
- If men and women feel similarly about this issue we would expect about the same proportion of each group to say yes.

Tally your results and repeat the experiment many times.

- Count the red cards among the 26 cards you dealt. The red cards represent the women who said yes.
- Repeat this process many times to generate the **sampling distribution** of the number of women who said yes.

# Here are results from 91 replications



# The tabulated results

No. YES's	9	10	11	12	13	14	15	16	17	18	19
No. of times	1	1	5	9	15	22	16	11	7	2	2

# Analyze the results

- What counts are "likely?"
- How did you decide to call a particular count "likely?"
- What counts are "unlikely?"
- What do our simulation results tell us about how *unusual* the actual poll results are?

# Draw a conclusion

- What does the result of the randomization test imply about the difference between men's and women's views in the population?
- The probability of getting 16 or more women saying yes under the null hypothesis is:  $22/91$
- Thus, the p-value of our test is: 0.2418
- We fail to reject the null and conclude that the evidence we have suggests men's and women's opinions do not differ on this question.

# A conceptualization of this randomization test

- Each subject has two characteristics: their sex and their response to the question
- We can imagine that if men and women do not differ in their responses then we can reassign the 26 yes's at random to the individuals in our sample.
- By doing this re-randomization many times we generate the pattern of yes's we would expect to see if there were no difference
- We compare what we expect to see if the null is true with what we actually observed.
- If what we observed would rarely be seen when the null is true, we reject the null.

# Summarizing the Randomization Test in the two-sample setting

- Null hypothesis: The two groups have the same proportion who say yes
- Alternative: The proportions differ
- Choose a test statistic: Number of women who say yes
- Generate the sampling distribution of the test statistic under the null
- Compute the p-value of the observed value of the test statistic

**S-IC.2. Decide if a specified model is consistent with results from a given data-generating process, e.g., using simulation.**

- Randomization tests provide a way you can teach this Common Core Standard without presenting the underlying distributional structure of the data (the hypergeometric).