

Making Sense of Statistical Inference: A Simulation Approach

Daren Starnes The Lawrenceville School (NJ) dstarnes@lawrenceville.org

A “few” inference formulas

| | |
|--|--|
| $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ | $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ |
| $\bar{x} \pm t^* \frac{s_x}{\sqrt{n}}$ | $t = \frac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}}$ |
| $(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ | $z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}}$ |
| $(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ | $t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ |
| $b \pm t^* SE_b$ | $t = \frac{b - \beta_0}{SE_b}$ |
| $\chi^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$ | |

Questions:

- (1) What percent of students in introductory statistics can reliably perform the appropriate calculation in a given inference problem?
- (2) Of those who can do the calculations correctly, what percent understand the meaning of their result?

Goals for the Session:

- Use a web-based applet to estimate a margin of error from a sample survey.
- Determine statistical significance in a randomized experiment using a hands-on simulation and a web-based applet.
- Discuss the appropriate scope of inference for a statistical study based on how the data were produced.

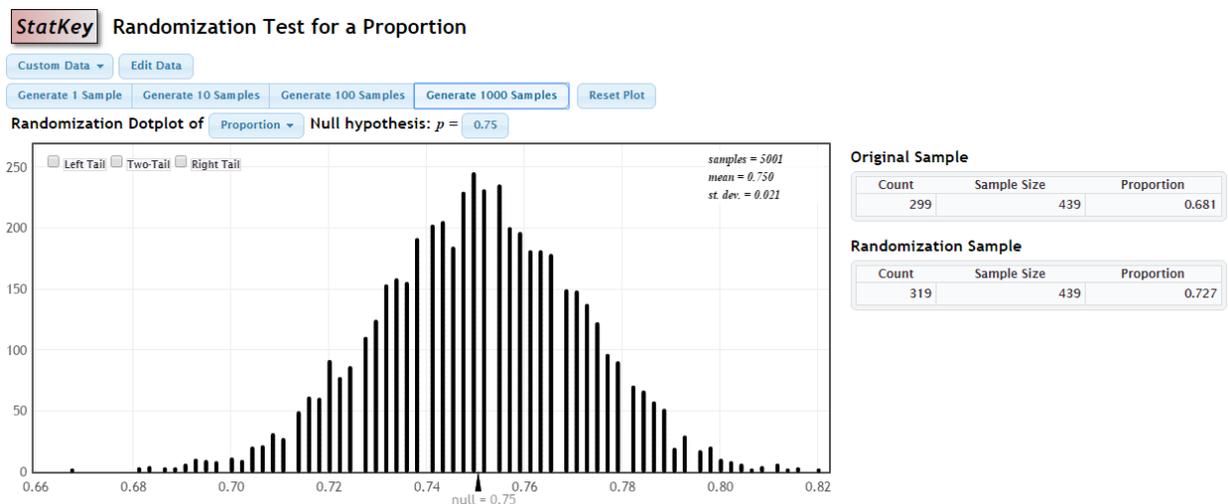
I. Do parents monitor teen internet use?

Estimating margin of error

The Pew Research Center asked a random sample of 439 parents of U.S. 13- to 14-year-olds, “Have you ever checked which websites your child visited?” 68% said “Yes.” Suppose that we want to estimate the true proportion p of all parents of U.S. 13- to 14-year-olds who would say that they have checked which websites their child has visited.

- What is our “best guess” for the value of the population proportion p based on the sample data?
- Do you believe that the population proportion p is exactly equal to this value?
- Do you believe that the population proportion p is close to this value?
- What values of p are plausible (believable) based on the sample result?
Is $p = 0.75$ a plausible value of the population proportion? <Don't answer yet!>

1. Point your browser to www.lock5stat.com/statkey. Then click on “Test for Single Proportion”.
2. Click on “Edit Data”. Enter 299 for the count of successes and 439 as the sample size.
3. Change the “Null hypothesis” value of p to 0.75.
4. Click on Click on Generate 1 Sample. Repeat several times. What is happening?
5. Click on Generate 1000 Samples several times. What do you notice?



Is it plausible that $p =$

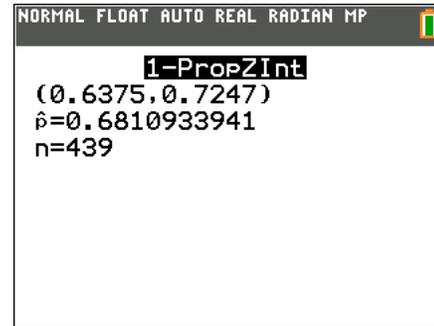
- | | | | | |
|-------|-------|-------|-------|-------|
| 0.74? | 0.73? | 0.72? | 0.71? | 0.70? |
| 0.69? | 0.68? | 0.67? | 0.66? | 0.65? |
| 0.64? | 0.63? | 0.62? | | |

What is the maximum distance that the sample proportion is likely to be from the true population proportion due to random sampling variability? In other words, what's the *margin of error*?

A Traditional One-sample z interval for a proportion

Still not comfortable with this simulation approach? Let's compare to our traditional inference formula:

$$\begin{aligned} \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ = 0.681 \pm 1.96 \sqrt{\frac{0.681(0.319)}{439}} \\ = 0.681 \pm 0.0436 \\ = (0.6374, 0.7246) \end{aligned}$$



II. Does sleep deprivation linger?

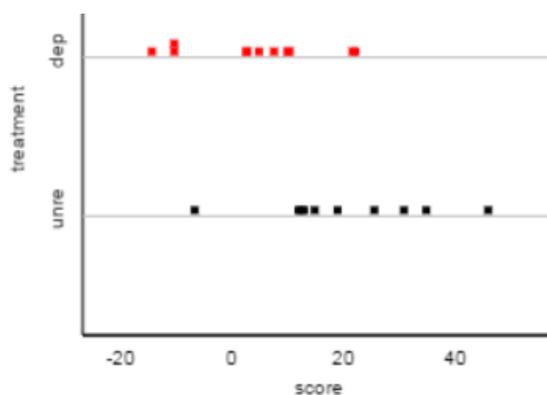
Determining statistical significance

(from Rossman, Cobb, Chance, and Holcomb's NSF project shared at JMM 2008 in San Diego)

Researchers have established that sleep deprivation has a harmful effect on visual learning. But do these effects linger for several days, or can a person "make up" for sleep deprivation by getting a full night's sleep on subsequent nights? A recent study (Stickgold, James, and Hobson, 2000) investigated this question by randomly assigning 21 subjects (volunteers between the ages of 18 and 25) to one of two groups: one group was deprived of sleep on the night following training and pre-testing with a visual discrimination task, and the other group was permitted unrestricted sleep on that first night. Both groups were then allowed as much sleep as they wanted on the following two nights. All subjects were then re-tested on the third day. Subjects' performance on the test was recorded as the minimum time (in milliseconds) between stimuli appearing on a computer screen for which they could accurately report what they had seen on the screen. The sorted data and dotplots presented here are the improvements in those reporting times between the pre-test and post-test (a negative value indicates a decrease in performance):

Sleep deprivation ($n = 11$): -14.7, -10.7, -10.7, 2.2, 2.4, 4.5, 7.2, 9.6, 10.0, 21.3, 21.8

Unrestricted sleep ($n = 10$): -7.0, 11.6, 12.1, 12.6, 14.5, 18.6, 25.2, 30.5, 34.5, 45.6



Summary Statistics:

| | n | Mean | SD |
|--------|----|-------|-------|
| dep | 11 | 3.90 | 12.17 |
| unre | 10 | 19.82 | 14.73 |
| pooled | 21 | 11.48 | 13.44 |

Compare the improvement scores for the two groups. Does it appear that subjects who got unrestricted sleep on the first night tended to have higher improvement scores than subjects who were sleep deprived on the first night? Explain briefly.

Question: Is it plausible (believable) that there's really no harmful effect of sleep deprivation, and random chance alone produced the observed differences between these two groups?

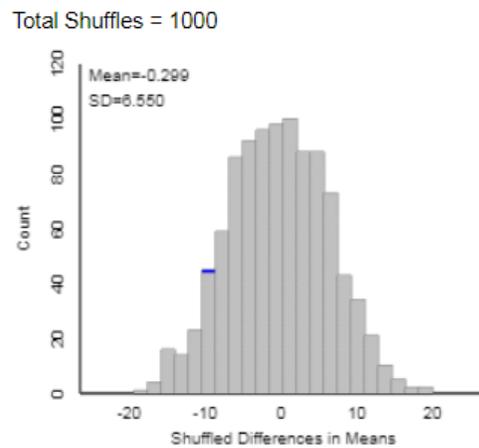
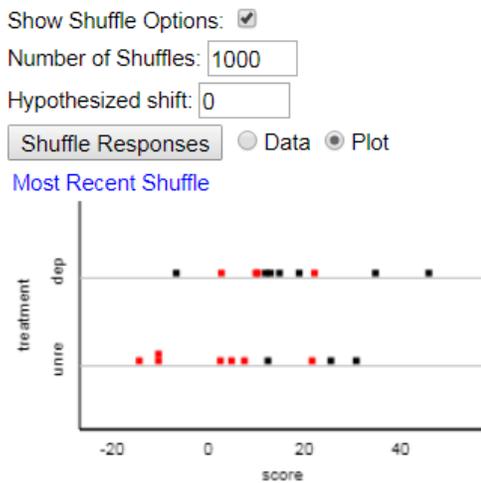
Physical simulation of the random assignment

- If no difference in treatment effects, then values will be the same as in the original study.
- Write each of the 21 data values on a separate card.
- Place all of the cards (subjects) in a bag.
- How large a difference in group means with different random assignments?
- Mix your cards and deal two groups—one with 10 cards (unrestricted sleep) and one with 11 cards (sleep deprived).
- Calculate the difference (unrestricted – sleep deprived) in mean time improvement for the two groups.
- Use a sticky dot to record your value on the poster.

Discussion:

Computer simulation of the random assignment

- Point your browser to www.rossmanchance.com/applets.
- Choose the Randomization test for quantitative response—two means applet



III. Scope of inference

The way in which data are produced determines the types of conclusions we can make.

- Random sampling allows us to make an inference about _____.
- Random assignment allows us to make an inference about _____.

| | | Were subjects randomly selected? | |
|----------------------------------|-----|---|--|
| | | YES | NO |
| Were subjects randomly assigned? | YES | The researcher: —may infer cause and effect; and —may generalize findings to the population. | The researcher: —may infer cause and effect; but —may <i>not</i> generalize findings to the population. |
| | NO | The researcher: —may <i>not</i> infer cause and effect; but —may generalize findings to the population. | The researcher: —may <i>not</i> infer cause and effect; and —may <i>not</i> generalize findings to the population. |

Adapted from *The Statistical Sleuth*, by Ramsey and Schafer