Math 361

Day 17

Confidence Intervals – Inv. 1.11

Announcements

HW 4 is a group assignment:

- Please turn in one copy per **2 to 3 students**.
- The last question relates to the group project that is 15% of your total grade.

Learning Objectives

1. Decide between using the ***"Plus Four" confidence interval*** and a ***"Wald" (or "z") confidence interval***.

2. Compute a ***confidence interval*** for a population proportion π**.**

3. Interpret a ***confidence interval*** without statistical jargon for a particular situation.

4. Determine whether a ***test of significance*** or a ***confidence interval*** is more appropriate to answer a particular research question.

Recall St. George's Hospital in Inv. 1.3

Parameter of interest: π = probability of death

Test H_0: $\pi=0.15$ vs. H_a: $\pi > 0.15$

Observed 8 of 10 patients die so the p-value ≈ 0 and we reject H_0.

*It looks like the death rate is not 15%...**what is it?***

Recall St. George's Hospital in Inv. 1.3

*It looks like the death rate is not 15%...**what is it?***

What if we test $\pi=0.16$, or $\pi=0.20$, or …

Using H_0: $\pi=0.50$ leads to a p-value > 0.05 so 50% is a plausible death rate…and so is everything above 50%

Based on our sample, it looks like the death rate is somewhere between 50% and 100%.

Recall St. George's Hospital in Inv. 1.3

*It looks like the death rate is not 15%...**what is it?***

What if we test $\pi=0.16$, or $\pi=0.20$, or …

Using H_0: $\pi=0.50$ leads to a p-value > 0.05 so 50% is a plausible death rate…and so is everything above 50%

Based on our sample, it looks like the death rate is somewhere between 50% and 100%.

This is a mis-use of a test of significance so let's make a new procedure.

Estimating π

Goal: find a set of plausible values for π, the probability of success or population proportion

$$P(L < \pi < U) = 0.95$$

↑ formulas ↑

Estimating π

Goal: find a set of plausible values for π, the probability of success or population proportion

Idea: Find formulas for L and U that we can compute from a sample so that the probability that π is between L and U is 0.95.

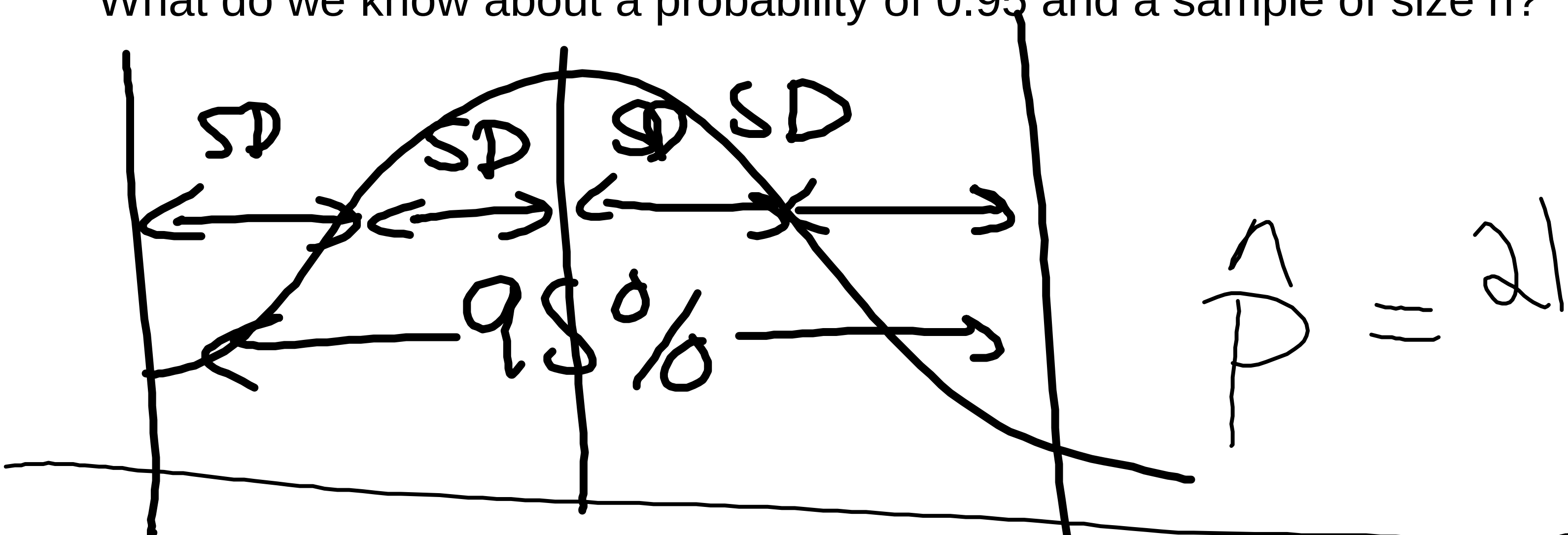
The set [L, U] is called a ***95% confidence interval***

Finding L and U

Given a sample of size n, we want to find L and U so that

$$P(L< \pi<U)=0.95$$

What do we know about a probability of 0.95 and a sample of size n?



$\hat{p} = 21/29$

Sample proportions $= \hat{p}$

Finding L and U

Given a sample of size n, we want to find L and U so that

$$P(L < \pi < U) = 0.95$$

What do we know about a probability of 0.95 and a sample of size n?

- The Empirical Rule says 95% of data from a normal distribution is within 2 SDs of the mean
- The CLT says that if $n\pi \geq 10$ and $n(1-\pi) \geq 10$ then the sample proportion $\hat{p}$ is approximately normal with mean π and SD = $\sqrt{\pi(1-\pi)/n}$

Finding L and U

Given a sample of size n, we want to find L and U so that

$$P(L< \pi<U)=0.95$$

What do we know about a probability of 0.95 and a sample of size n?

- The Empirical Rule says 95% of data from a normal distribution is within 2 SDs of the mean
- The CLT says that if $n\pi \geq 10$ and $n(1-\pi) \geq 10$ then the sample proportion $\hat{p}$ is approximately normal with mean π and SD = $\sqrt{\pi(1-\pi)/n}$

Putting these facts together means that

P(mean-2SD < $\hat{p}$ < mean +2SD) = 0.95
P(π -2 SD< $\hat{p}$ <π +2 SD) = 0.95
…some algebra…

P($\hat{p}$ -2 SD< π <$\hat{p}$+2 SD) = 0.95.
Use SD = $\sqrt{\hat{p}(1-\hat{p})/n}$ and we have formulas for L and U!

One Proportion *z*-interval ("Wald")

- General form: statistic $\pm$ "margin-of-error"
- An approximate 95% confidence interval for π

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- Conditions: binomial process and that *n* is large enough for CLT to kick in, that is

$$n\pi \geq 10 \text{ and } n(1-\pi) \geq 10$$

“Plus Four” 95% CI for π

What if $n\hat{p}<10$ or $n(1-\hat{p})<10$?

No problem, just add 2 success and 2 failures to your sample!

Definition: Plus Four 95% confidence interval for π:

- Determine the number of successes (X) and sample size (n) in the study
- Increase the number of successes by two and the sample size by four. Make this value the midpoint of the interval: $\tilde{p} = (X + 2)/(n + 4)$
- Use the z-interval procedure as above for the augmented sample size of $(n + 4)$:

$$\tilde{p} \pm 1.96\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$$

Which formula, Wald or Plus Four?

Suppose you're estimating π with a confidence interval using a sample of size n from a Binomial Random Process:

1. If the confidence level is 95%, then always use the "Plus Four" formula
2. If the confidence level is not 95% and there are at least 10 "successes" and 10 "failures" in the sample, then you may use the "Wald" (a.k.a. "z") formula

If 2. is not true, ask a statistician for help…

1. Decide between using the ***"Plus Four" confidence interval*** and a ***"Wald" (or "z") confidence interval***.

Estimating probability of "heads"

Let π = probability of "heads" in a coin toss.

Suppose we observe 16 heads in 20 tosses of a coin.

What are the plausible values of π? Calculate a 95% CI

2. Compute a ***confidence interval*** for a population proportion π**.**

What do we mean by 95% confidence?

- We say a confidence interval procedure is “95% confident” if, in the long run, 95% of intervals created with this method succeed in capturing the value of the parameter

- To test this, you can create a process where you know π, generate 1000s of samples, calculate the corresponding interval for each sample, compute the percentage of the intervals that success in capturing π

Estimating probability of "heads"

Let π = probability of "heads" in a coin toss.

Suppose we observe 16 heads in 20 tosses of a coin.

Interpret the 95% CI

Using the "Plus Four" formula, we found that L = 0.625 and U = 0.975.

These values may be interpreted as follows:

I am 95% confident that the probability of getting "heads" with this coin is between 0.625 and 0.975.

"95% confident" means that if I was to repeatedly toss the coin 20 times, record the number of "heads" and compute a "Plus Four" CI, then 95% of these intervals would contain the actual probability of getting "heads".

3. Interpret a ***confidence interval*** without statistical jargon for a particular situation.

Test or Confidence Interval?

If a research question asks:

- whether π is a specific value or set of values then use a **test of significance** (i.e. Binomial test)

Is a coin fair?

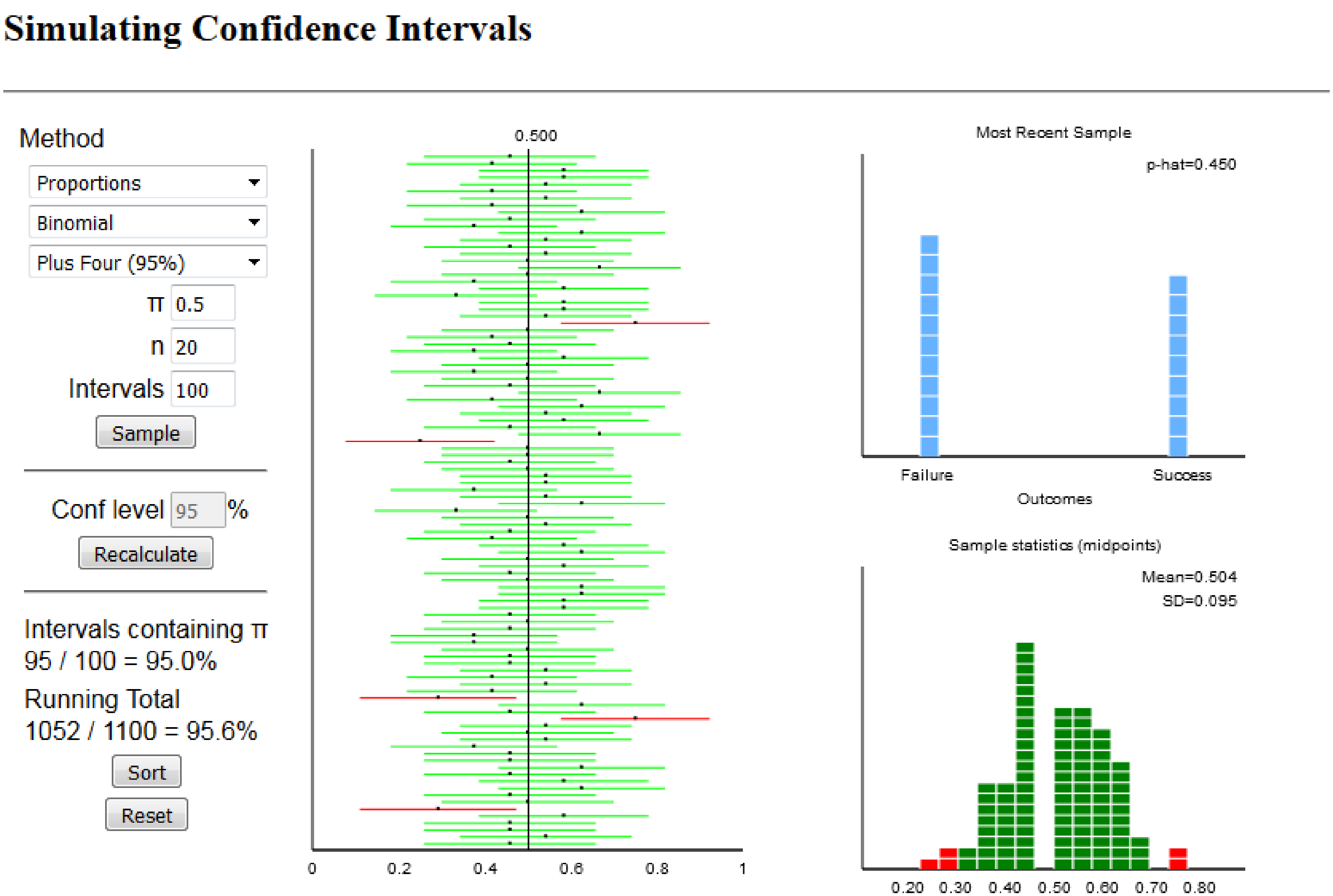
Do most students prefer cookies over brownies?

- "what is" the value of π then use a **confidence interval**

What is the probability this coin lands "heads"?

What proportion of all students prefer cookies over brownies?

Simulating CI Applet



Inv. 1.11: Estimating the Death Rate

Try parts d, f, and g:

- Compute both the one-sample z-interval (“wald”) and the “Plus Four” interval by hand and by applet
- Determine which one method is better by simulating sample data in an applet
- Interpret the “Plus Four” interval

One Proportion *z*-interval ("Wald")

Can choose any level of confidence

- An approximate C% confidence interval for π

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

  - *z** is called the "critical value" and is the number such that the probability of between –z* and z* is C in the Normal distribution.
  - Larger confidence level means larger multiplier means wider interval