# Math 243

Day 15

Confidence Intervals – Inv. 1.11

# Announcements

- HW 4 posted, due Oct. 22
- **Pick a group** for the team project
  - Each team should consist of 2-3 students
  - The project is 20% of your course grade

*Each project group may turn in a single set of solutions for HW 4 on Oct. 22.*

# Recall St. George's Hospital in Inv. 1.3

**Parameter of interest:** $\pi$ = 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

# 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?

# 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 $\pi$

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

- 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 $(n + 4)$:

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

# Estimating probability of “heads”

Let $\pi$ = probability of “heads” in a coin toss.

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

**What are the plausible values of $\pi$? Calculate a 95% CI**

# 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 = 0.577 and U = 0.923.

These values may be interpreted as follows:

I am 95% confident that the probability of getting “heads” with this coin is between 0.577 and 0.923.

“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”.

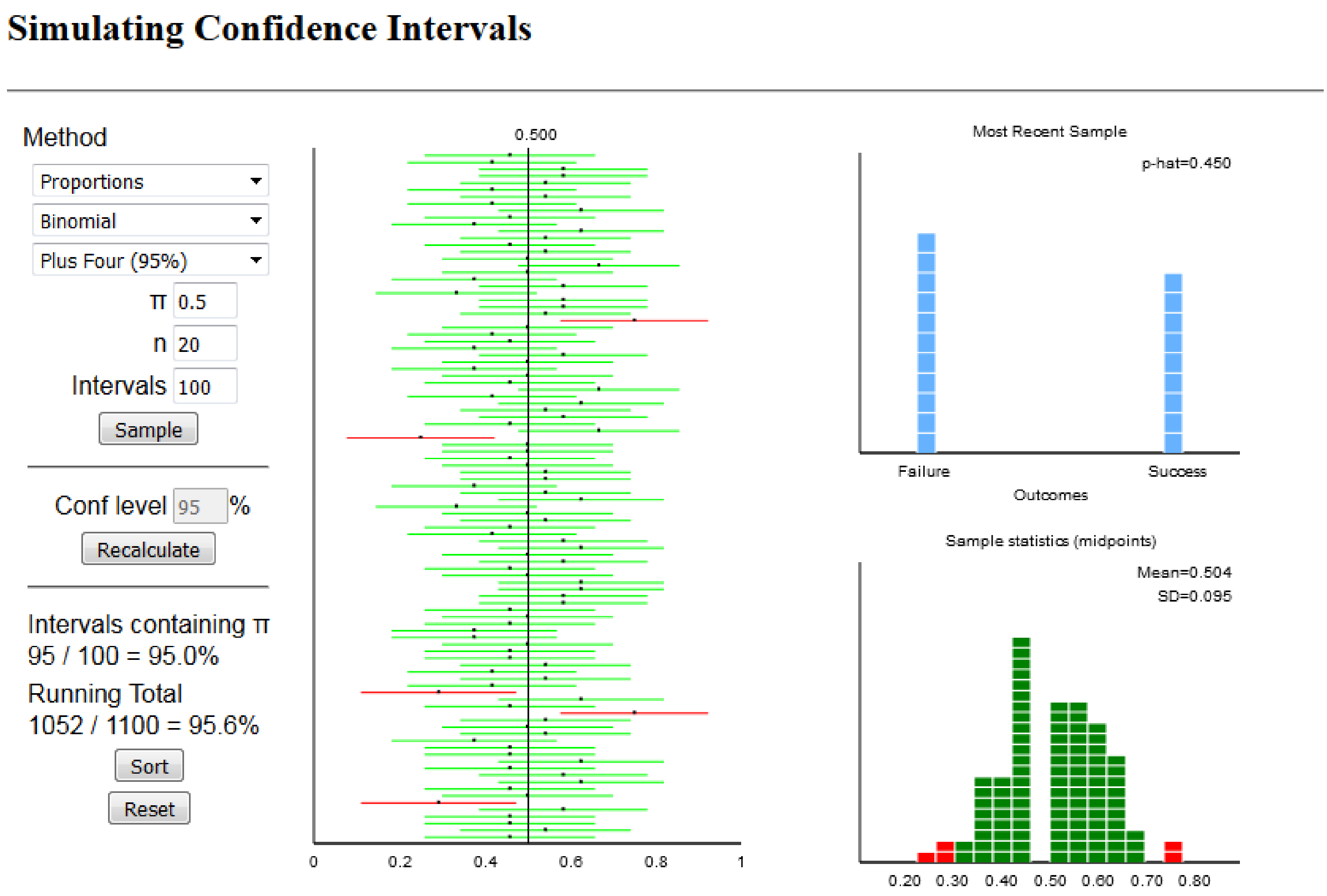
# Group Work – Inv. 1.11 page 83

Recall Investigation 1.3, where you learned that 8 of the 10 most recent heart transplantation operations at St. George's Hospital resulted in a death.

(a) Use the one sample $z$-interval method to find a 95% confidence interval for the probability of a heart transplantation death at St. George's hospital. Does anything bother you about doing this?

(f) Use the Plus Four procedure to determine and interpret a 95% confidence interval for the probability of a death during a heart transplant operation at St. George's hospital. [*Hints*: You can do the calculation either by hand, first finding $\tilde{p}$ and $z^*$, or with the Theory-Based Inference applet or software by telling the technology there were 4 more operations consisting of two more deaths than in the actual sample.]

# Simulating CI Applet



# 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 $\pi$, generate 1000s of samples, calculate the corresponding interval for each sample, compute the percentage of the intervals that success in capturing $\pi$

# Inv. 1.11: Estimating the Death Rate

**Try parts d, g:**

- Determine which one method is better by simulating sample data in an applet

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

Can choose any level of confidence

- An approximate C% confidence interval for $\pi$

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

  - $z*$ 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