Math 243

Inv. 1.7 – Power, cont’d

Inv. 1.9 – One Proportion Z-test, by hand and applet

How can we compute the **Power** of a test?

Power: probability of correctly rejecting H_0 when a specific H_a is true

1. Find the rejection region for a given level of significance.
2. Simulate the distribution assuming the *alternative* hypothesis is true.
3. Compute the probability of the rejection region assuming H_a is true.

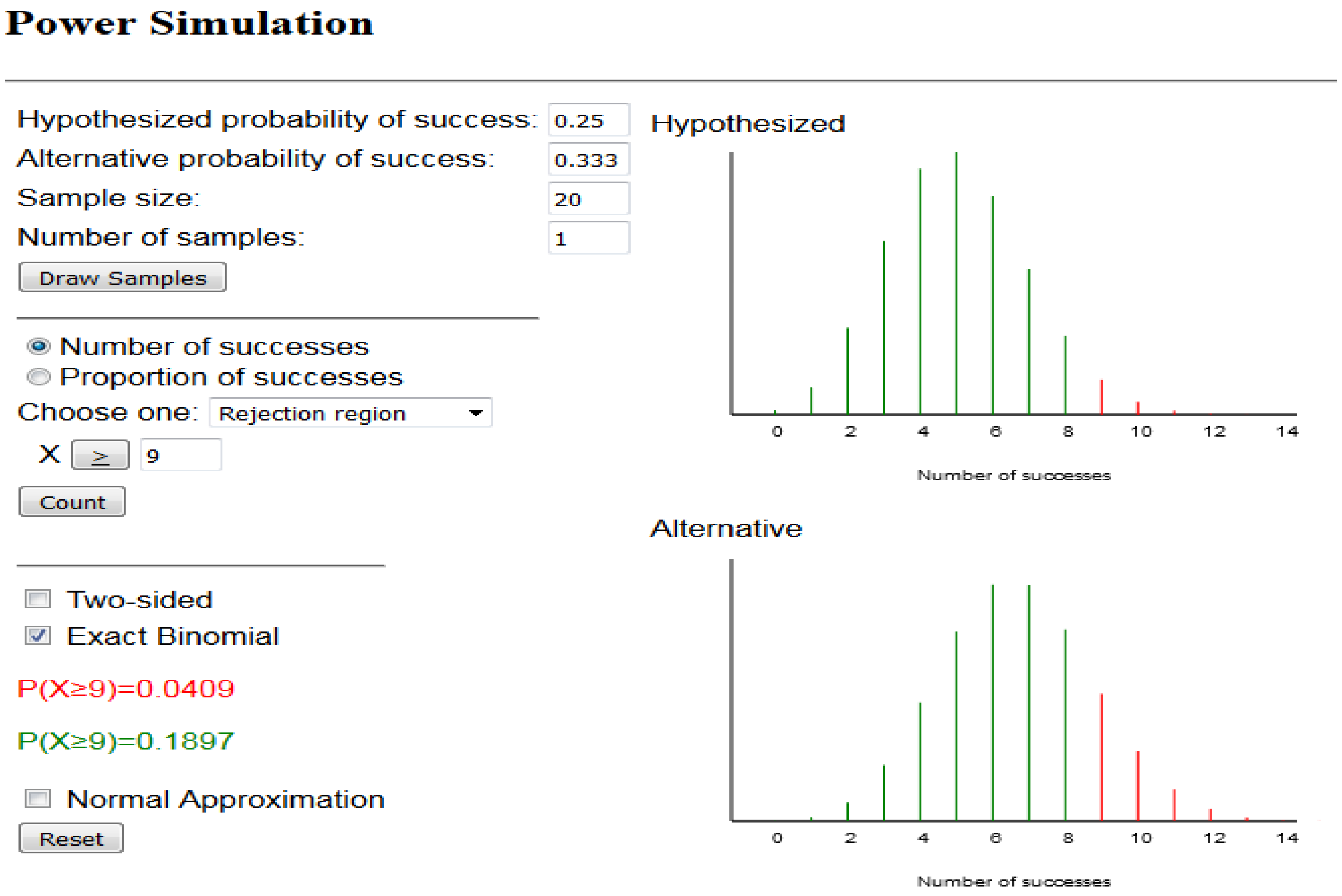
Sound hard? It's easy when you use an applet!

Power Simulation Applet (batting averages)

Power is given in green

The probability that the manager correctly decides the 0.333 player improved is 0.19…

…not very likely.



How can **power** be improved?

Try increasing number of at-bats the manager observes.

If the manager watches 250 at-bats then the probability he will decide the 0.333 hitter has improved is about 0.91…

… much better from the player's perspective

Power Simulation

Hypothesized probability of success: 0.250

Alternative probability of success: 0.333

Sample size: 250

Number of samples: 1000

Draw Samples

Total =1000

- Number of successes
- Proportion of successes

Choose one: Level of significance

α = 0.05

Count

Hypothesized: Proportion of samples: 46 / 1000 = 0.0460

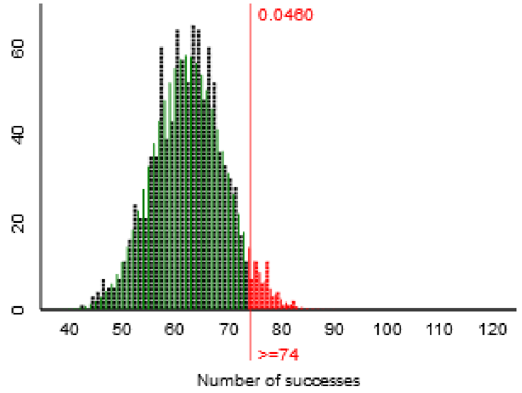
Alternative: Proportion of samples: 920 / 1000 = 0.9200

- Two-sided
- Exact Binomial

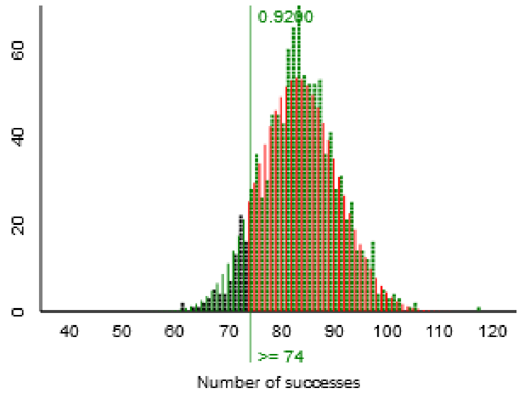
P(X≥74)=0.0560

P(X≥74)=0.9056

Hypothesized



Alternative



What is the probability of a **Type I error**?

The manager is worried about making a **Type I error**, that is, deciding the player improved when really he hasn't.

The probability Type I error is set by choosing a cutoff value so that if the p-value is below this value then the null hypothesis will be rejected.

This value is called the **level of significance** and is notated α.

The manager could decrease the probability of a type I error by using $\alpha = 0.01$ instead of $\alpha = 0.05$.

Controlling Type I & II error

Control the probability of Type I error by fixing the level of significance α:

If you only reject H_0 when the p-value is less than α then the probability of making a type I error is at most α

Control the probability of Type II error (β) by adjusting your study design:

Design your study so that power = 1-β is high (close to 1).

Try increasing sample size or decreasing α

Try practice problem 1.7D on page 61

Practice Problem 1.7D

Suppose you want to test a person's ability to discriminate between two types of soda. You fill one cup with Soda A and two cups with Soda B. The subject tastes all 3 cups and is asked to identify the odd soda. You record the number of correct identifications in 10 attempts. Assume a one-sided alternative.

(a) If the subject's actual probability of a correct identification is 0.50, what is the power of this test for a level of significance of $\alpha = 0.50$? [*Hint*: What is the null hypothesis?]

(b) Write a one-sentence interpretation of the power you calculated in (a) in context.

(c) What is the power if you give the subject 20 attempts?

Solution

Practice Problem 1.7D

Suppose you want to test a person's ability to discriminate between two types of soda. You fill one cup with Soda A and two cups with Soda B. The subject tastes all 3 cups and is asked to identify the odd soda. You record the number of correct identifications in 10 attempts. Assume a one-sided alternative.

(a) If the subject's actual probability of a correct identification is 0.50, what is the power of this test for a level of significance of $\alpha = 0.50$? [*Hint*: What is the null hypothesis?]

(b) Write a one-sentence interpretation of the power you calculated in (a) in context.

(c) What is the power if you give the subject 20 attempts?

(a) P(X>=4) = 0.8281

(b) If a subject actually has a 50% chance of selecting the "odd" cola, then the probability of correctly deciding the subject is not just guessing is 0.83.

(c) P(X>=8) = 0.8684

Performing a test of significance **by hand**

We'll develop some tools that will allow us to test

$$H_0: \pi = \text{(some number)}$$

vs.

$$H_a: \pi \neq \text{(some number)}$$

by hand

Example: Is a coin fair?

Suppose we have a coin that we suspect of being biased. Let's test

$$H_0: \pi = 0.5$$

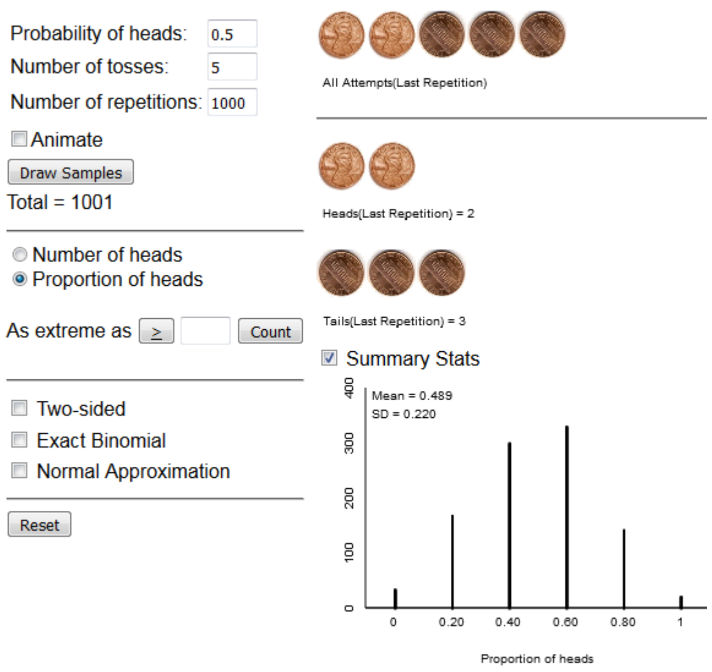
vs.

$$H_a: \pi \neq 0.5$$

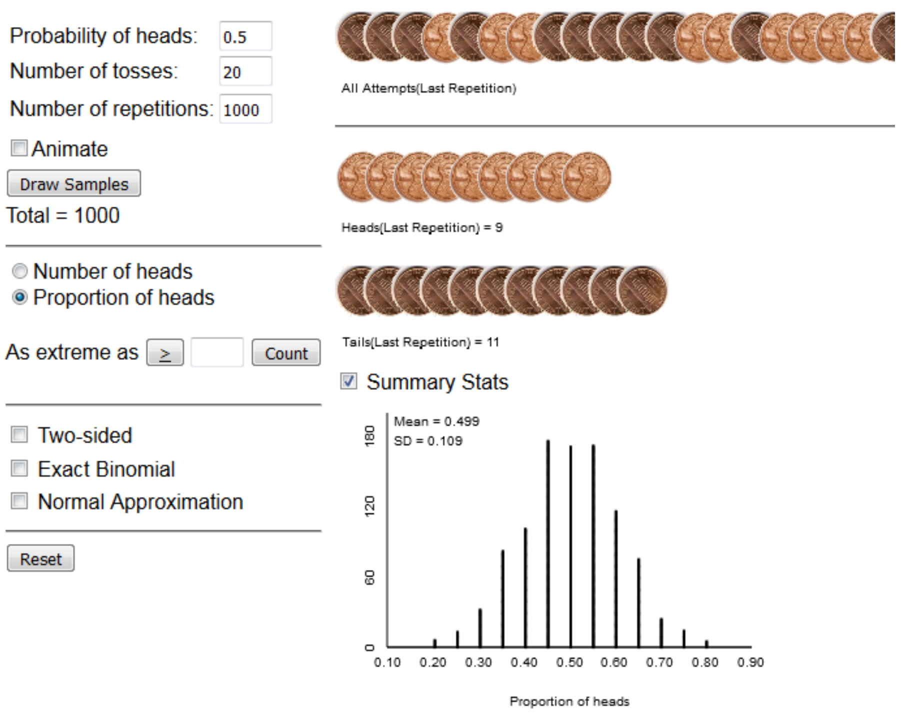
where π is the probability of the coin landing "heads"

Notice the distribution of the proportion of heads appears to have a specific form if n is large enough

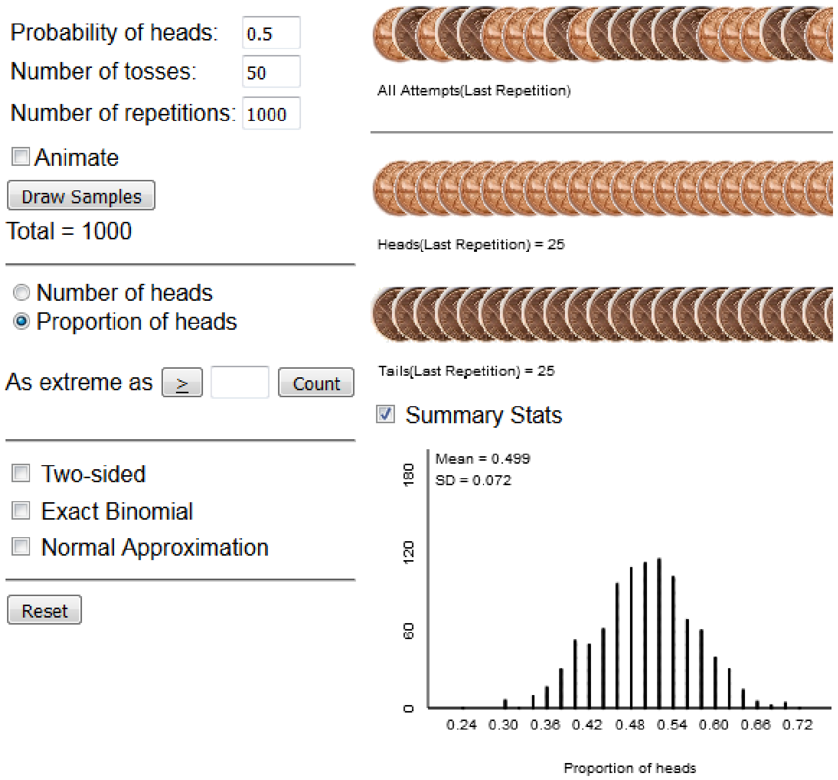
n =5



n=20

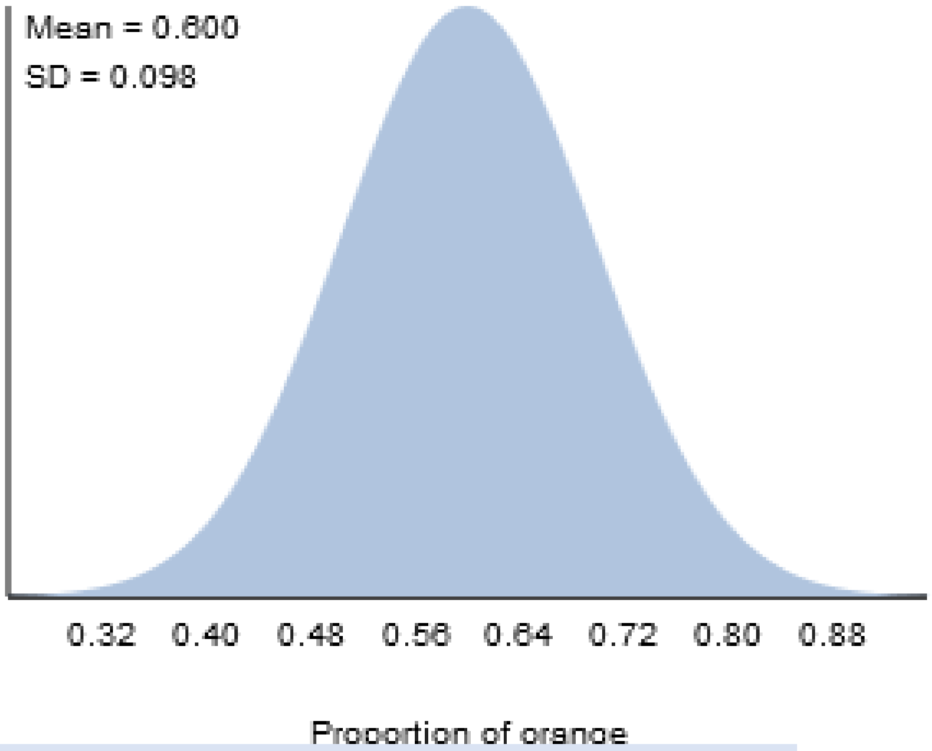


n=50



The Normal Distribution

Described by two parameters:
- mean
- standard deviation



The formula for the curve is messy…

The Normal Approximation to the Binomial

Equations relating parameters:

- mean =π
- Standard deviation = $\sqrt{\frac{\pi(1-\pi)}{n}}$

The Normal Approximation to the Binomial

Equations relating parameters:

- mean $=\pi$
- Standard deviation $=\sqrt{\frac{\pi(1-\pi)}{n}}$

But the approximation only works well when

$\pi n \geq 10$ and $(1-\pi)n \geq 10$

Central Limit Theorem: CLT

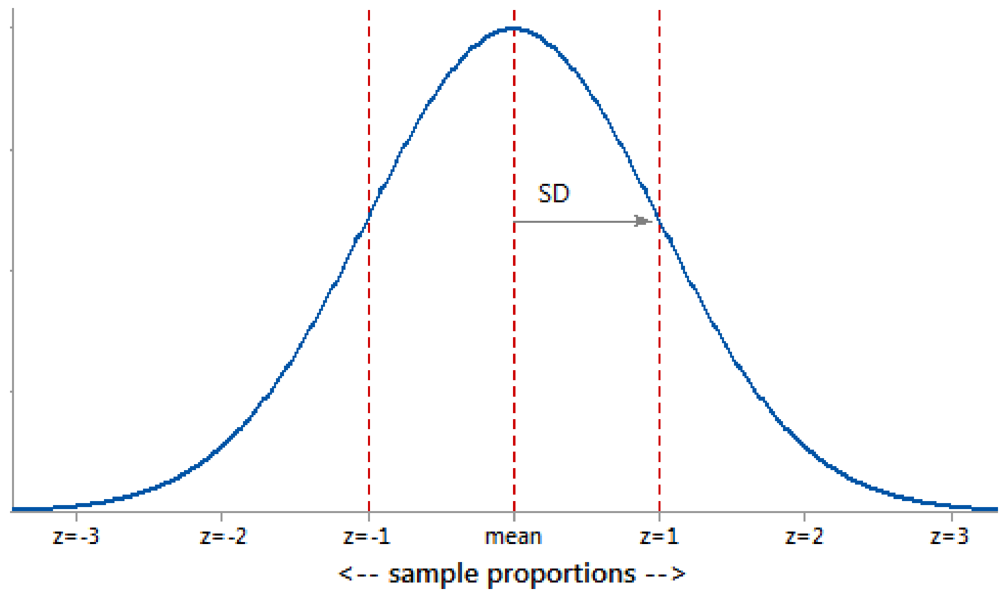
The distribution of sample proportions (stemming from a binomial process) will be approximately normal

If the sample size is large relative to the value of π

(that is, $n\pi \geq 10$ and $n(1-\pi) \geq 10$)

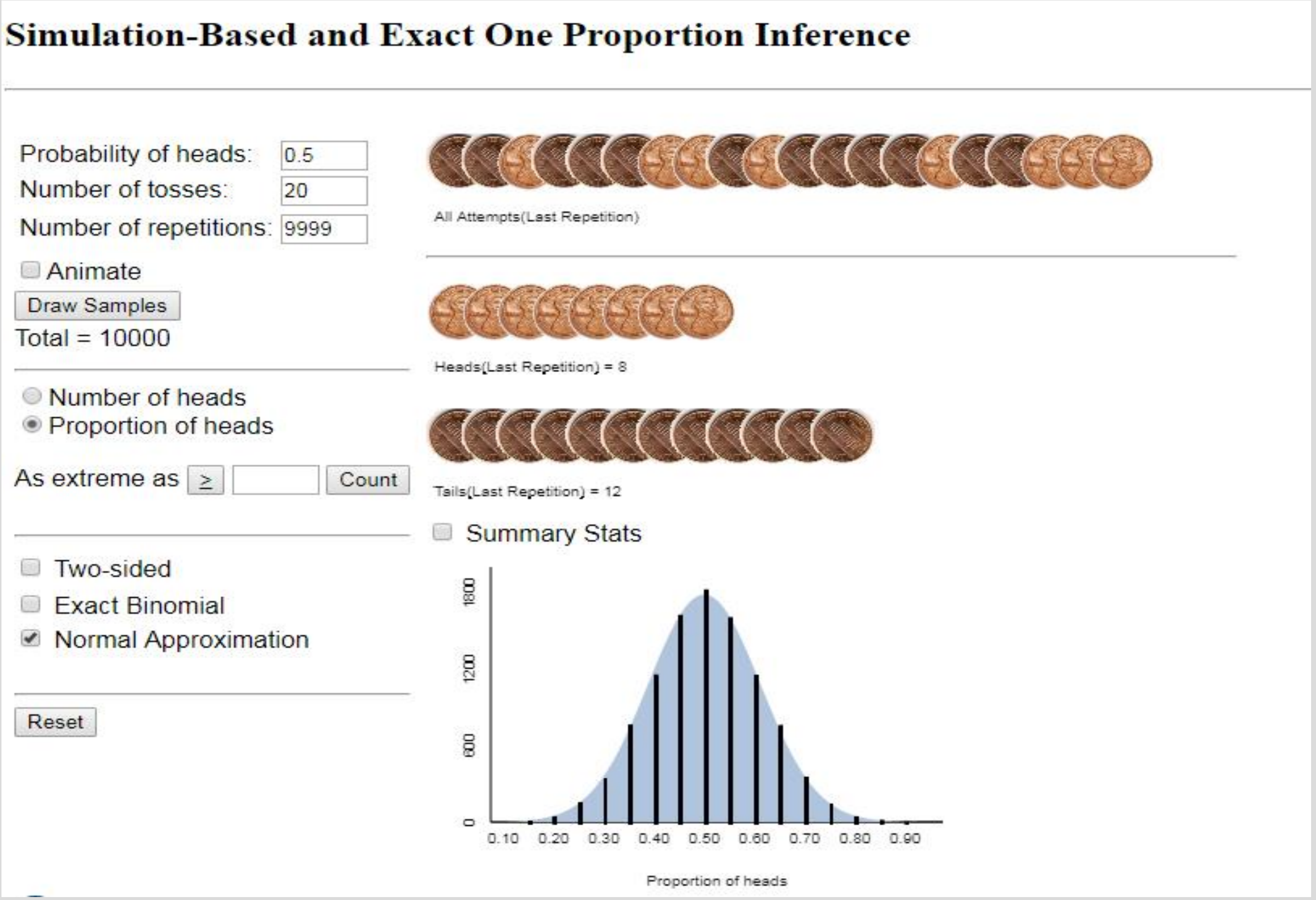
Then

- the mean $\approx \pi$
- and SD $\approx \sqrt{\pi(1-\pi)/n}$



Applying CLT…

Count # heads out of 20 coin tosses, then repeat 10,000 times…



…the distribution of X = “proportion of heads” is ≈ Normal!

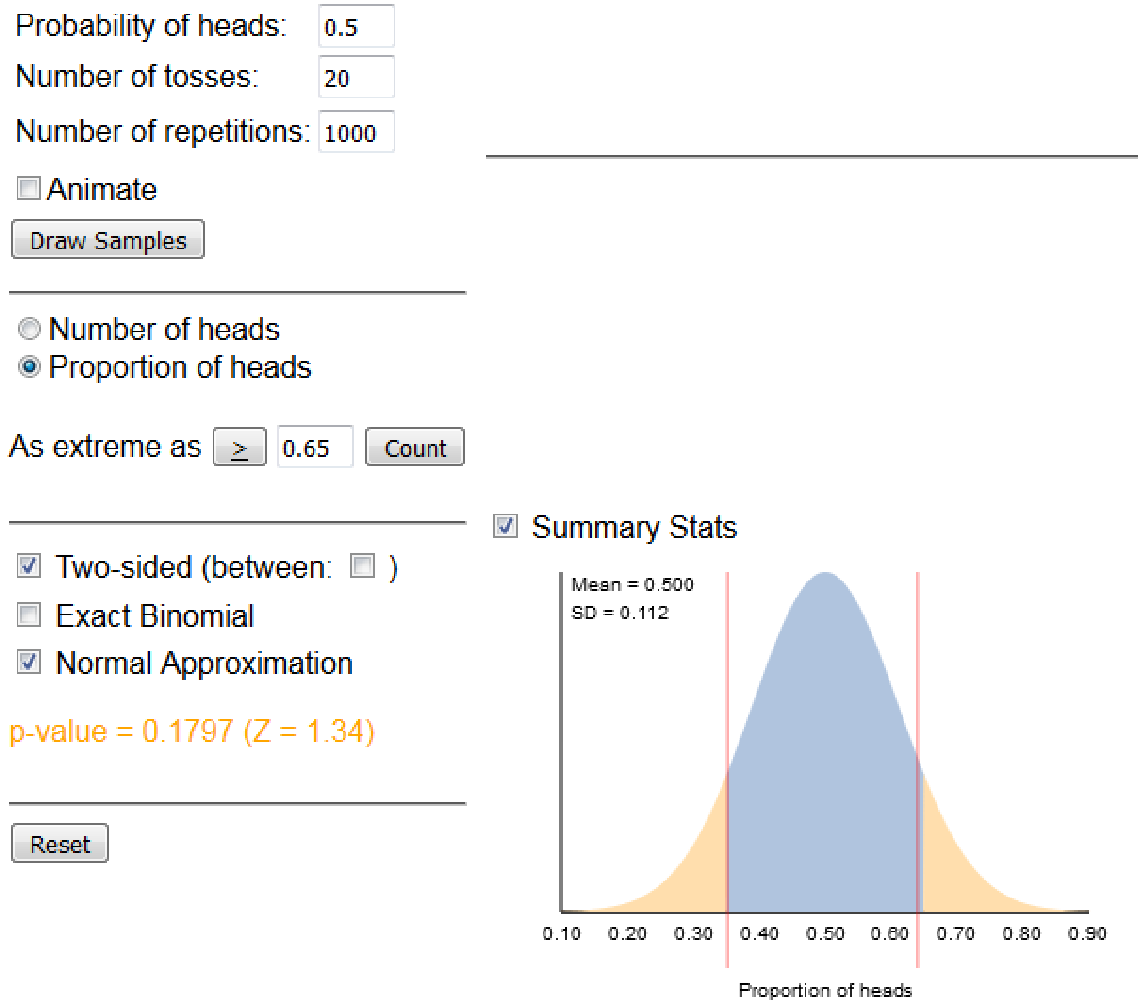
Suppose we observe 13 out of 20 tosses land heads…

Sample proportion is 13/20=0.65

Check “two-sided” box and “Normal Approximation”

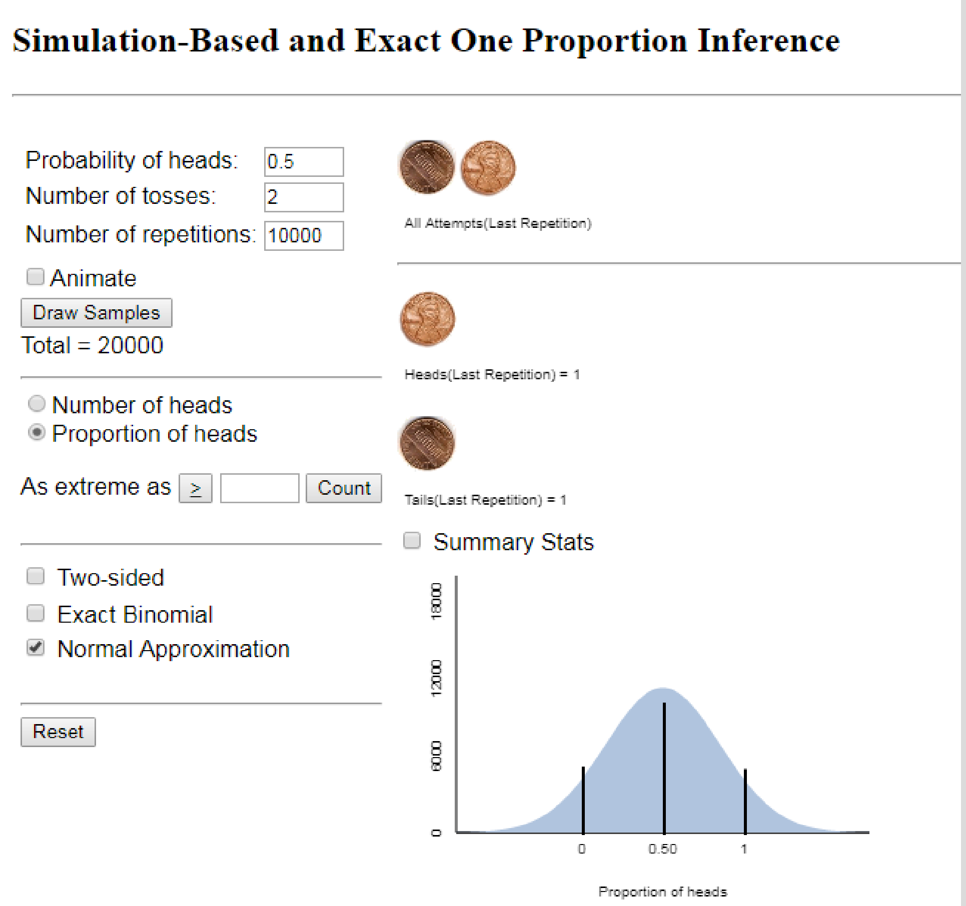
The p-value is 0.18, so fail to reject the null.

There is no evidence our coin is biased.



Must check conditions before applying CLT

Count # heads out of 2 coin tosses, then repeat 10,000 times…

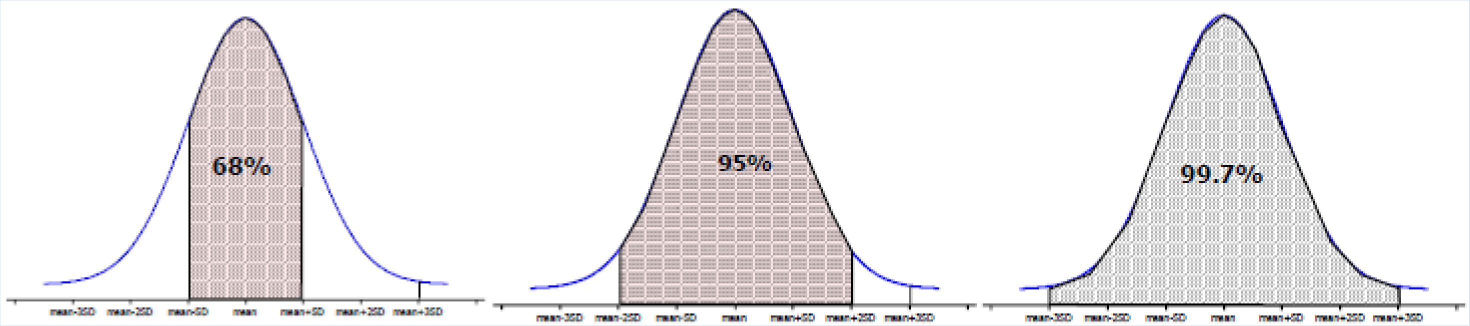


…the distribution of X = "proportion of heads" is not very Normal!

Advantages to Normal Distribution

- Historically was more convenient for calculating probabilities (area under the curve)
  - Could standardize (z-score) and look up on a table
  - Especially useful for calculating power
- Still useful for some of today's "big data" problems
- Empirical rule (68/95/99.7)

- the interval ($\mu - \sigma$, $\mu + \sigma$) should capture approximately 68% of the distribution.
- the interval ($\mu - 2\sigma$, $\mu + 2\sigma$) should capture approximately 95% of the distribution.
- the interval ($\mu - 3\sigma$, $\mu + 3\sigma$) should capture approximately 99.7% of the distribution.



The Empirical Rule allows us to perform a “two-sided” test by hand!

- Draw distribution using CLT
- Compute **mean +2SD** and **mean – 2SD**
- Find rejection region using ER
- If the sample proportion is in rejection region, reject the null otherwise fail to reject the null

Example: observed 13 “heads” in 20 tosses.

One Sample *z*-test for proportions

1. Define parameter (process probability or population proportion)
2. State null and alternative hypotheses
3. Check whether CLT applies: $n\pi$, $n(1-\pi) \geq 10$
4. Calculate test statistic (*z*-score)
   - Interpretation: Number of SDs from null value of π

   OR calculate Mean + 2SD and Mean -2SD to find rejection region
5. Calculate p-value using normal distribution

OR check whether the sample proportion is within 2 SDs of the mean

6. State conclusions

Inv. 1.9: toys or candy on Halloween?

a) **Obs. Units:** children

Variable: Did a child choose the toy?

b) The parameter of interest is the proportion of all children who prefer toys to candy on Halloween (π)

c) Test H_0: $\pi = 0.5$ vs. H_a: $\pi \neq 0.5$

d) Of n=284 children, 135 chose the toy so $\hat{p} = 0.475$.

Could compute the p-value using simulations or Exact Binomial, but let's try applying the CLT and use the Normal Distribution instead. This is called the…

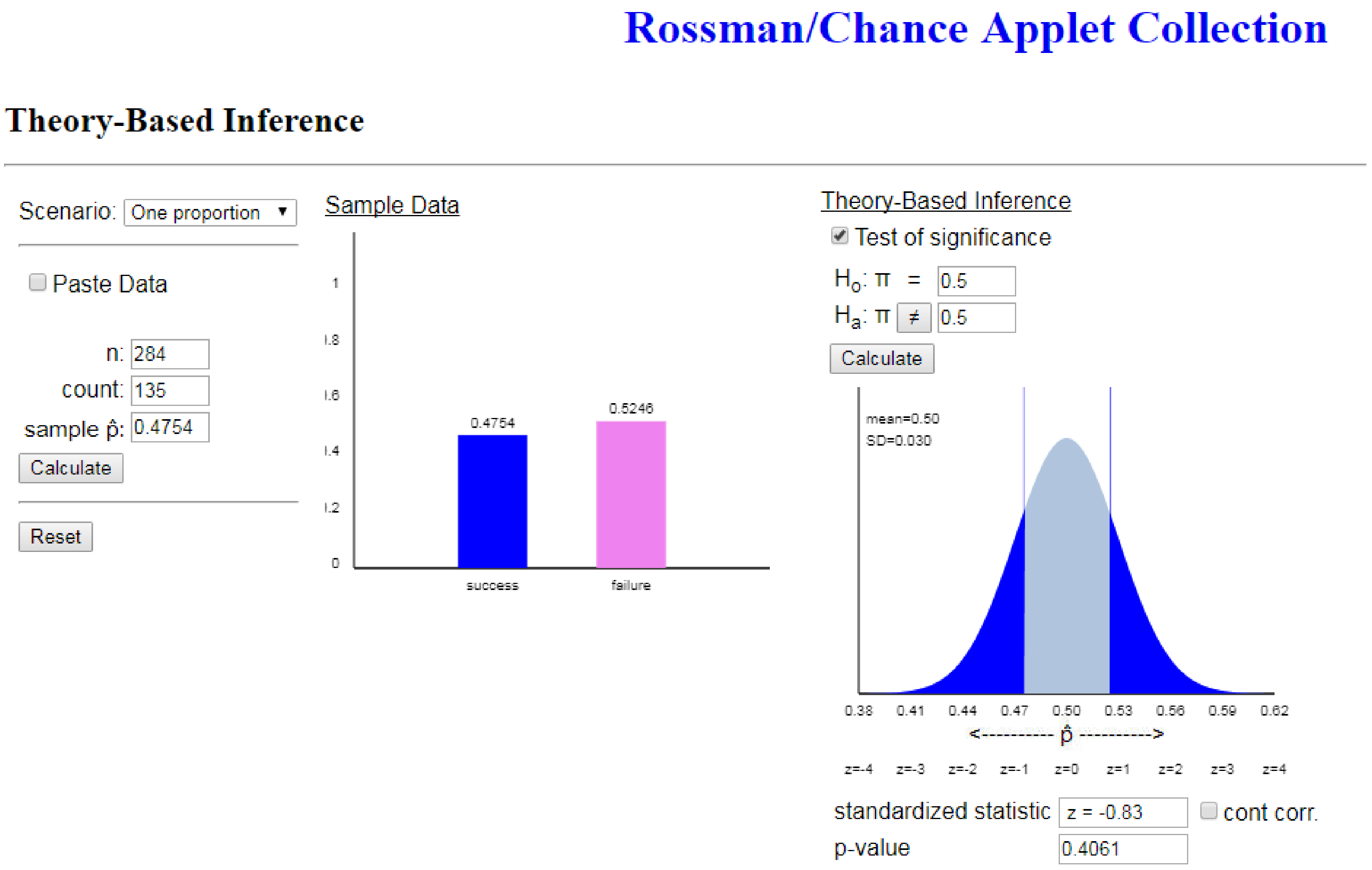
Inv. 1.9 – by hand via Empirical Rule

(e) Check conditions $n\pi \geq 10$ and $n(1-\pi) \geq 10$:

(f) Draw the normal distribution of $\hat{p}$ assuming H_0 is true:
mean= π and SD = $\sqrt{\pi(1-\pi)/n}$ and add our value of $\hat{p}$.

(i) z-score = how many SD's from the mean is our $\hat{p}$? More than 2 is "extreme" by ER.

Inv. 1.9 – via applet



Investigation 1.9

- p-value ≈ 0.40
- Conclusions?