Revised: 3/20/2021

Inference I

Michael C. Hackett

Assistant Professor, Computer Science

Community College of Philadelphia

Inference

- *Descriptive statistics* was primarily concerned with describing and summarizing data
- *Probability* was primarily concerned with understanding and the likelihood of events happening
- **Inference** is primarily concerned with understanding uncertainty

Point Estimates

- A **point estimate** an estimation of what might be observed if an entire population was sampled.
  - For example, a poll suggests that a politician's approval rating is 53%
  - 53% is a point estimate; The entire population was not surveyed, only a sample was.

- The entire population, p, is called the ***parameter*** *of interest*.
  - It's what we what hope to glean information about based on a sample of that population, $\hat{p}$

- Unless the entire population is sampled, p remains unknown and $\hat{p}$ is an estimate for p

Point Estimates

- Two types of errors in point estimates:
  - Sampling error: The variability of point estimates from the actual parameter of interest

  - Bias: A systematic tendency to under- or over-estimate the true population value.

Sampling Variability

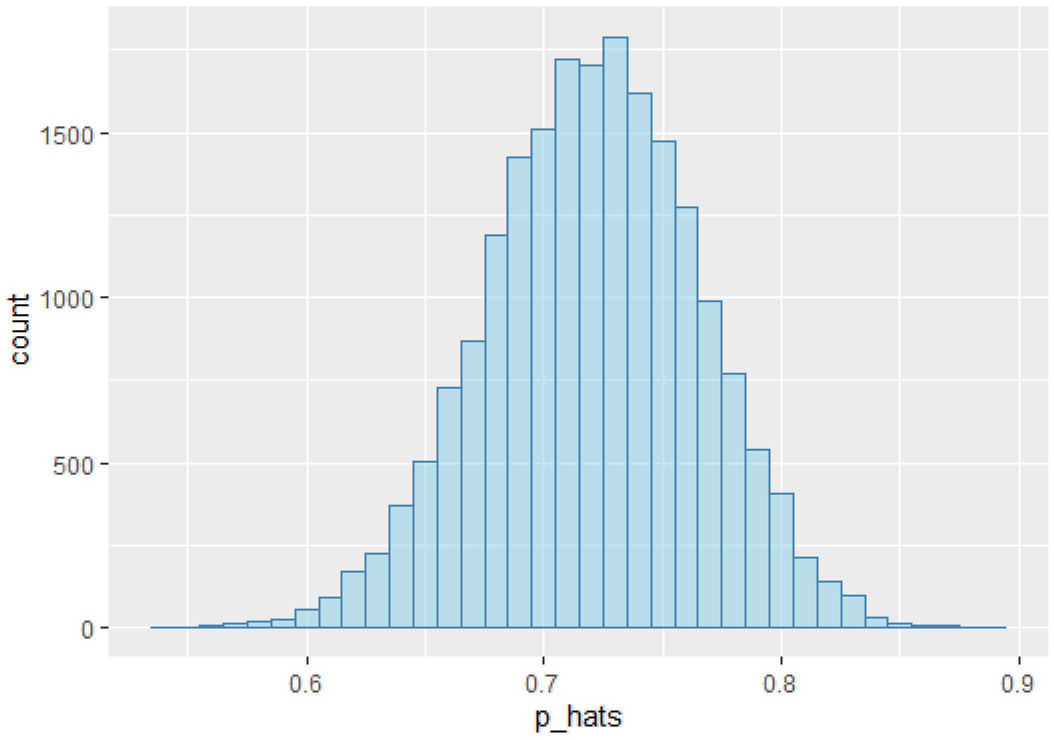
- The point estimate likely won't match the true parameter of interest.
  - For example, 72% of CCP students are in favor of a new school policy.
    - No way to know this without surveying every student
  - A survey of 100 students will likely not result in exactly 72%, but it should be close provided there were no major sampling errors.
- The question becomes how would $\hat{p}$ behave if p really is 0.72?

Sampling Variability

- A simulation can provide insight:
  1. Create a vector the size of the population (assumes to be 15000 for this example). 72% should be “approve”, 28% should be “disapprove”
  2. Randomly sample 100 entries from the vector
  3. Find the sample proportion of “approve” entries
  4. Add the sample proportion to a vector
  5. Repeat Step 1…

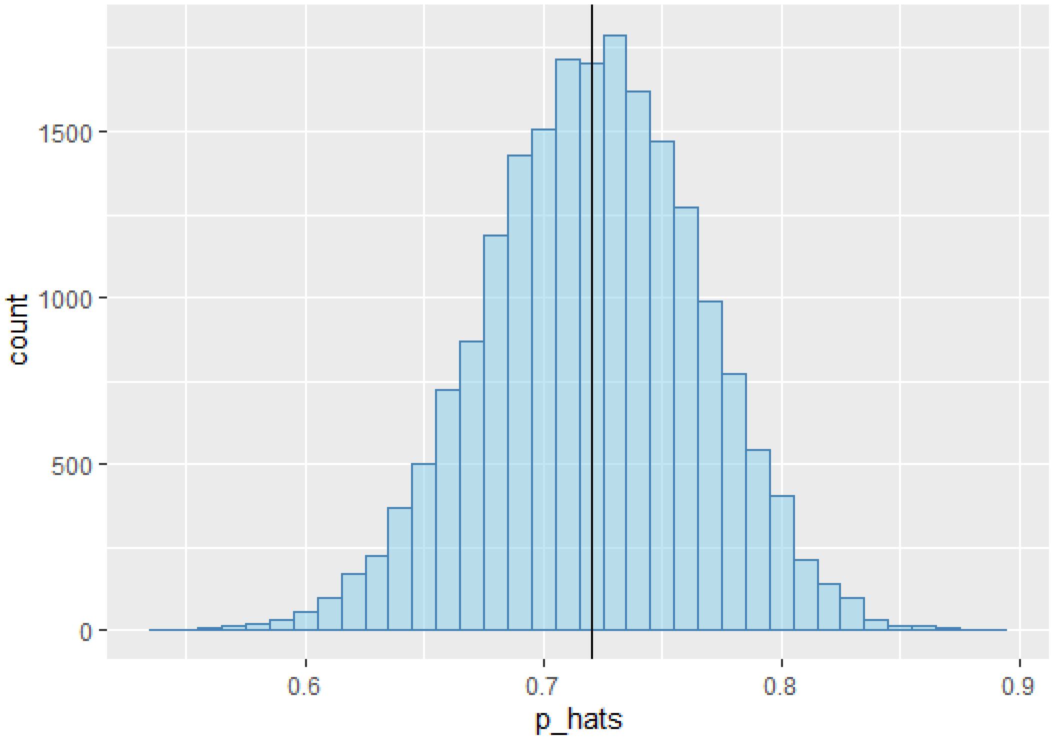
Sampling Variability

- Histogram of 20000 simulations:
  - Called a **sampling distribution**



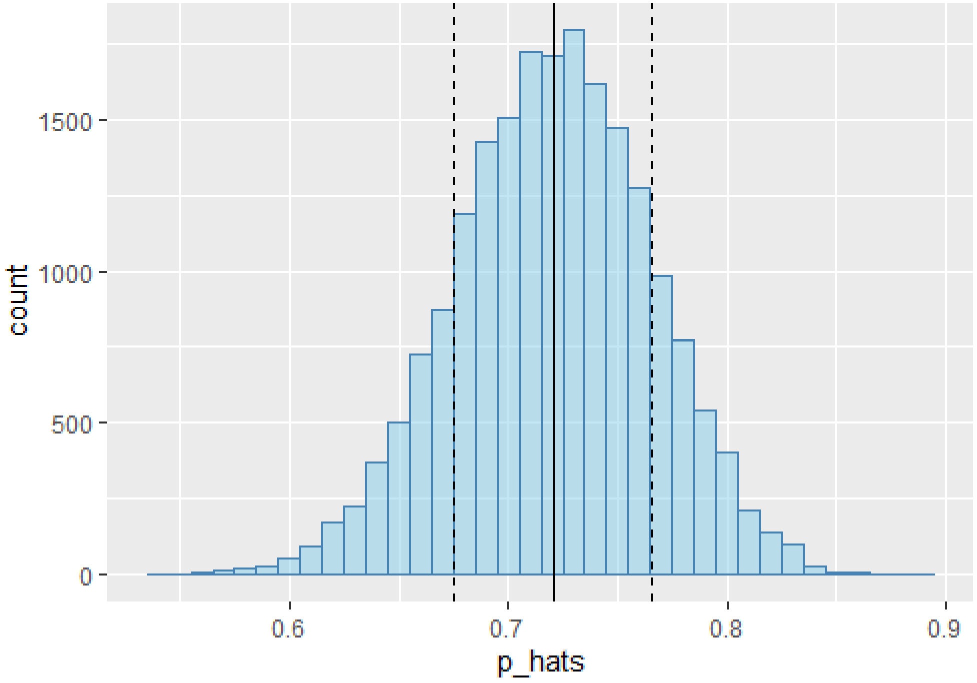
Sampling Variability

- The center of the distribution (mean $\hat{p}$) will be close to p
  - The more simulations, the closer it will get



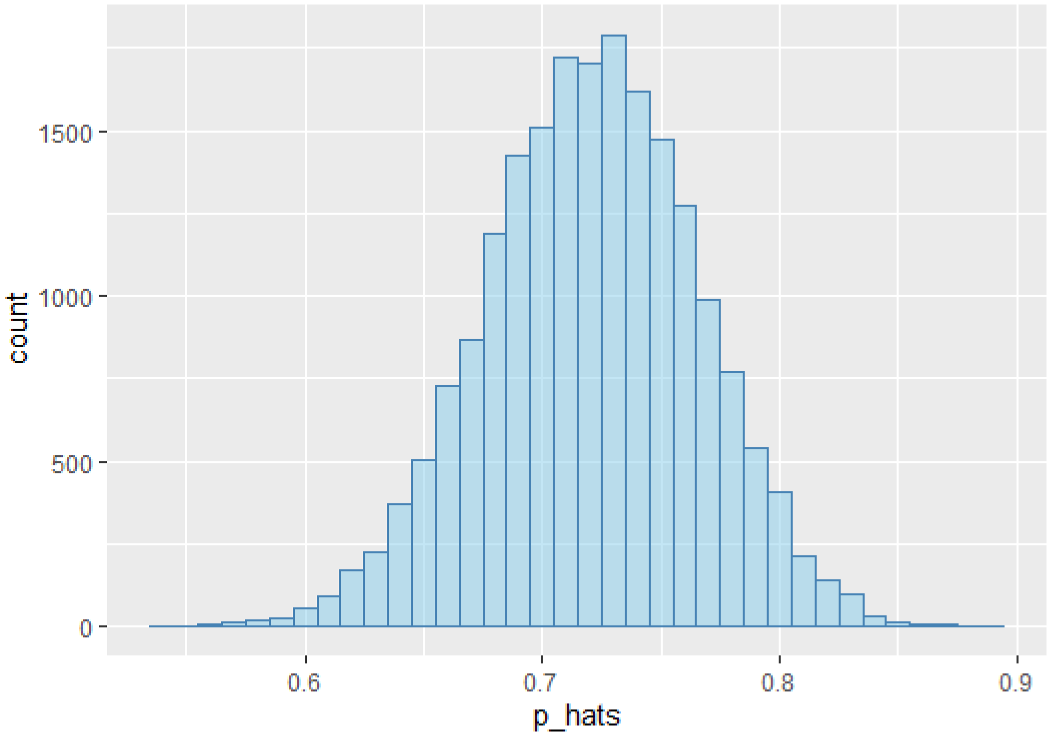
Sampling Variability

- The standard deviation for this distribution is about 0.045
  - In sampling distributions, it's called the **standard error,** $SE_{\hat{p}}$



Sampling Variability

- Shape is symmetric, bell-shaped, and resembles a normal distribution



Central Limit Theorem

- The Central Limit Theorem is that when observations are independent and n (the sample size) is sufficiently large, $\hat{p}$ will tend to follow a normal distribution with:
  - $\mu_{\hat{p}} = p$
  - $SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

- n is sufficiently large when $np \geq 10$ and $n(1-p) \geq 10$
  - The *success-failure condition*

Central Limit Theorem

- It might sound a bit like the Law of Large Numbers from probability theory
  - Law of Large Numbers: One sample
  - Central Limit Theorem: *Distribution* of samples

Central Limit Theorem

- Applying the Central Limit Theorem to the previous example
  - Observations are independent and n (the sample size) is sufficiently large

- $\mu_{\hat{p}} = 0.72$
- $SE_{\hat{p}} = \sqrt{\frac{0.72(1-0.72)}{100}} = 0.044$

- $100(0.72) \ \geq 10$ and $100(1 - 0.72) \ \geq 10$
  - $72 \ \geq 10$ and $2.8 \geq 10$
  - The success-failure condition is NOT satisfied
  - N is not sufficiently large enough

Central Limit Theorem

- A survey of 100 Americans resulted in 67% in favor of a new legislation
  - Does the Central Limit Theorem apply?
  - Independent if from the respondents were randomly selected
  - Not sure of what p really is
  - Can substitute $\hat{p}$ in place of p
    - $100(0.67) \geq 10$ and $100(1 - 0.67) \geq 10$
    - $67 \geq 10$ and $33 \geq 10$
    - The success-failure condition is satisfied
  - Central Limit Theorem applies, meaning $\hat{p}$ approximately follows a normal distribution.

- $\mu_{\hat{p}} = 0.67$
- $SE_{\hat{p}} = \sqrt{\frac{0.67(1-0.67)}{100}} = 0.267$

Confidence Intervals

- The sample proportion provides a plausible value for the population proportion.
  - The point estimate provides a plausible value for the population parameter.
  - Will have a standard error

- Better to provide a range of plausible values for the population parameter, called a **confidence interval**
  - The larger the interval, the more confident you can be that the range contains the exact population parameter.

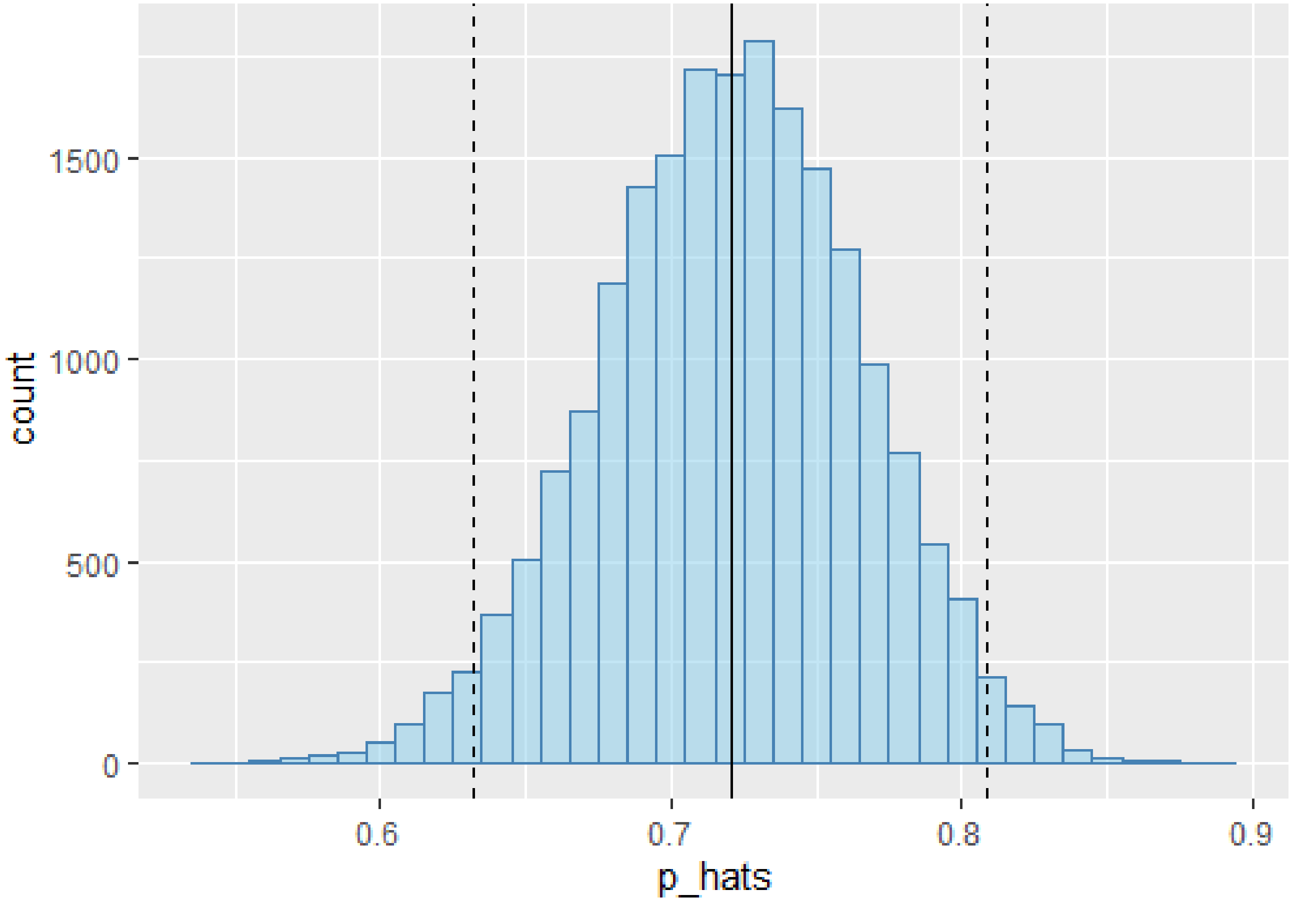
Confidence Intervals – 95%

- Like standard deviations in a probability distribution, standard errors in a sampling distributions determine confidence intervals.
  - Point estimates follow a normal distribution if the Central Limit Theorem applies
  - In a normal distribution, 95% of the data is within 1.96 standard deviations of the mean

$$\hat{p} \pm z^* \times SE_{\hat{p}} = 0.67 \pm 1.96 \times \sqrt{\frac{0.67(1-0.67)}{100}} = (.577, .762)$$

- 95% confident that the actual population parameter is between 57.7% and 76.2 percent
  - $z^* \times SE_{\hat{p}}$ is called the **margin of error**
  - $1.96 \times \sqrt{\frac{0.67(1-0.67)}{100}} = 0.092 = 9.2\%$

Confidence Intervals – 95%



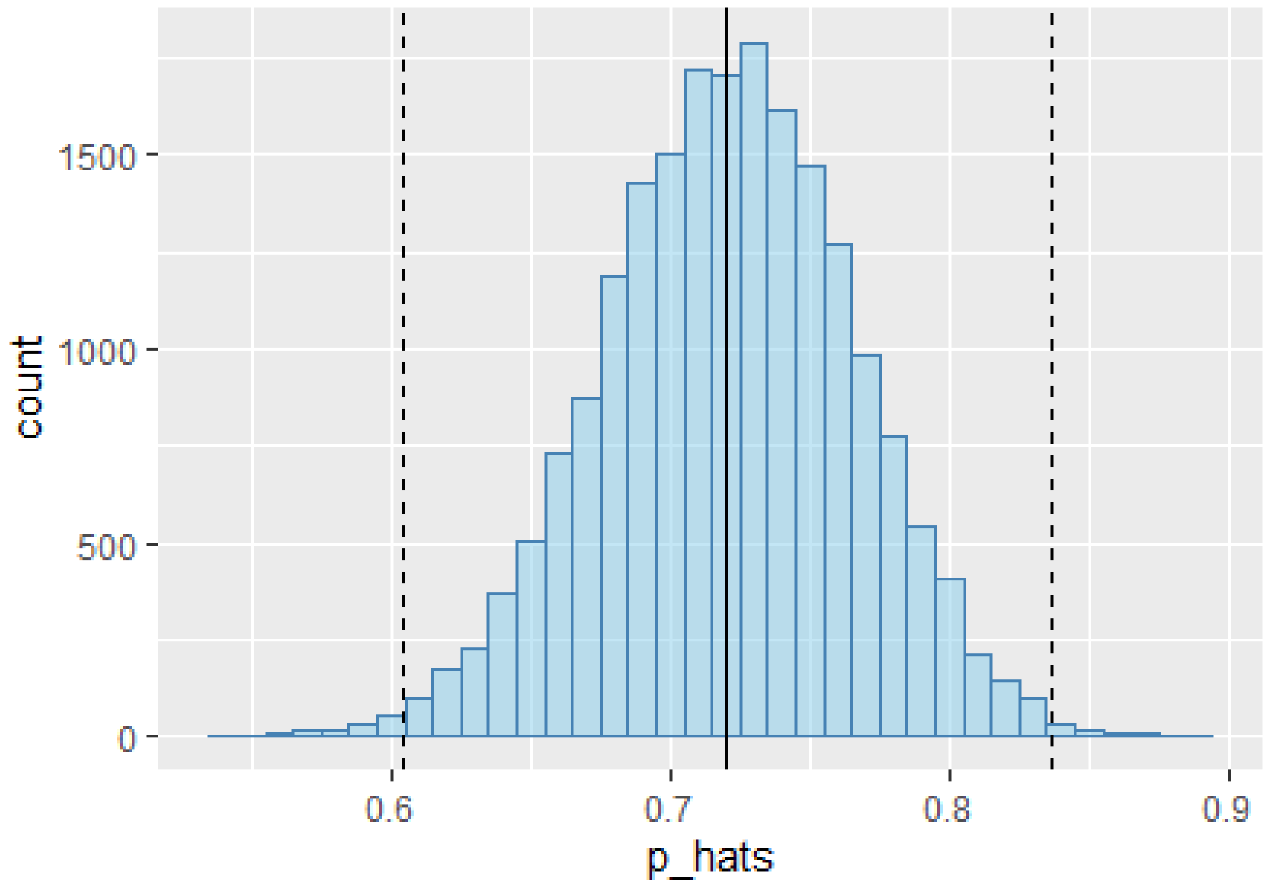
Confidence Intervals – 99%

- In a normal distribution, 99% of the data is within 2.58 standard deviations of the mean

$$\hat{p} \pm 2.58 \times SE_{\hat{p}} = 0.67 \pm 2.58 \times \sqrt{\frac{0.67(1-0.67)}{100}} = (.549, .791)$$

- 99% confident that the actual population parameter is between 54.9% and 79.1%

Confidence Intervals – 99%



Confidence Intervals

In New York City on October 23rd, 2014, a doctor who had recently been treating Ebola patients in Guinea went to the hospital with a slight fever and was subsequently diagnosed with Ebola.

Soon thereafter, an NBC 4 New York/The Wall Street Journal/Marist Poll found that 82% of New Yorkers favored a "mandatory 21-day quarantine for anyone who has come in contact with an Ebola patient". This poll included responses of 1,042 New York adults between Oct 26th and 28th, 2014.

Confidence Intervals

In New York City on October 23rd, 2014, a doctor who had recently been treating Ebola patients in Guinea went to the hospital with a slight fever and was subsequently diagnosed with Ebola.

Soon thereafter, an NBC 4 New York/The Wall Street Journal/Marist Poll found that 82% of New Yorkers favored a "mandatory 21-day quarantine for anyone who has come in contact with an Ebola patient". This poll included responses of 1,042 New York adults between Oct 26th and 28th, 2014.

- What is the point estimate?
- Is it reasonable to use a normal distribution to model that point estimate?

Confidence Intervals

- Based on a sample size (*n*) of 1042, the point estimate is $\hat{p} = 0.82$

- Independence: Simple random sampling
- Success-failure condition: $1042 \times 0.82 \ \geq 10$ and $1042(1 - 0.82) \ \geq 10$
  - $854.44 \ \geq 10$ and $187.56 \ \geq 10$

- Both conditions are met; $\hat{p}$ can reasonably be modeled using a normal distribution

Confidence Intervals

In New York City on October 23rd, 2014, a doctor who had recently been treating Ebola patients in Guinea went to the hospital with a slight fever and was subsequently diagnosed with Ebola.

Soon thereafter, an NBC 4 New York/The Wall Street Journal/Marist Poll found that 82% of New Yorkers favored a "mandatory 21-day quarantine for anyone who has come in contact with an Ebola patient". This poll included responses of 1,042 New York adults between Oct 26th and 28th, 2014.

- What is the estimated standard error?
  - $SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.82(1-0.82)}{1042}} = 0.012$

Confidence Intervals

In New York City on October 23rd, 2014, a doctor who had recently been treating Ebola patients in Guinea went to the hospital with a slight fever and was subsequently diagnosed with Ebola.

Soon thereafter, an NBC 4 New York/The Wall Street Journal/Marist Poll found that 82% of New Yorkers favored a “mandatory 21-day quarantine for anyone who has come in contact with an Ebola patient". This poll included responses of 1,042 New York adults between Oct 26th and 28th, 2014.

- Construct a 95% confidence interval for p, the proportion of New York adults who supported a quarantine for anyone who has come into contact with an Ebola patient.

Confidence Intervals

$$\hat{p} \pm z^* \times SE_{\hat{p}} = 0.82 \pm 1.96 \times \sqrt{\frac{0.82(1-0.82)}{1042}} = (.797, .843)$$

- 95% confident that the actual population parameter is between 79.7 and 84.3 percent

Hypothesis Tests

- *Hypotheses* are claims about a population.
  - A **hypothesis test** tests the validity of the statement

- The **null hypothesis** (H_0) is a claim to be tested
  - Usually, a skeptical perspective
- The **alternative hypothesis** (H_1) is an alternative claim to be tested
  - Usually new perspectives
  - Often represented by a range of possible parameter values

Hypothesis Tests

- Fifty people with a college degree were given multiple-choice questions on world health, such as:

  How many of the world's 1 year old children today have been vaccinated against some disease:
  a. 20%
  b. 50%
  c. 80%

- H_0= People never learn these topics in school and their answers are random guesses
- H_A= People do have knowledge on these topics and do better than random guessing; People have wrong knowledge on these topics and do worse than randomly guessing

Hypothesis Tests

- H_0
  - Respondents are as accurate as randomly guessing; The proportion of respondents who pick the correct answer is 33.3% (1 in 3)
  - $p = 0.333$

- H_A
  - The proportion of respondents who pick the correct answer is something other than 33.3%
    - People did worse than randomly guessing (< 33.3%) or people did better than randomly guessing (>33.3%)
  - $p \neq 0.333$

Hypothesis Tests

- It’s unlikely that $\hat{p}$ will equal p exactly
  - That is, it is unlikely the proportion of the sample (the test takers) will perform exactly equal to what is expected by blindly guessing.
- The result was 24% of test takers answered the question correctly.
  - Does this provide strong enough evidence that $p \neq 0.333$?
- We measure confidence intervals for hypothesis tests the same way as for a sample distribution.

Hypothesis Tests

- Confidence interval for a hypothesis test
  - Check if $\hat{p}$ is approximately normal (Central Limit Theorem)
    - Data was from a simple random sample
    - $n\hat{p} = 50 \times .24 = 12 \geq 10$
    - $n(1-\hat{p}) = 50 \times .76 = 38 \geq 10$

  - Construct confidence interval
    - $\hat{p} \pm z^* \times SE_{\hat{p}} = 0.24 \pm 1.96 \times \sqrt{\frac{0.24(1-0.24)}{50}} = 0.24 \pm 1.96 \times 0.0604 = (.12, .36)$

- 95% confident that the proportion of college educated adults that will answer the question correctly is between 12% and 36%

Hypothesis Tests

- H_0 $(p = 0.333)$ or 33.3% falls within that 95% confidence interval, H_0 is not implausible and we fail to reject the null hypothesis.
  - **There is not sufficient evidence** to say that the performance of test takers was any different from them guessing randomly.

- This does not necessarily mean the null hypothesis is true.
  - It was a very small sample size of 50 people

Errors

- In hypothesis tests, assertions are made about which hypothesis (H_0 or H_A) might be true, but its possible that assertion is wrong.

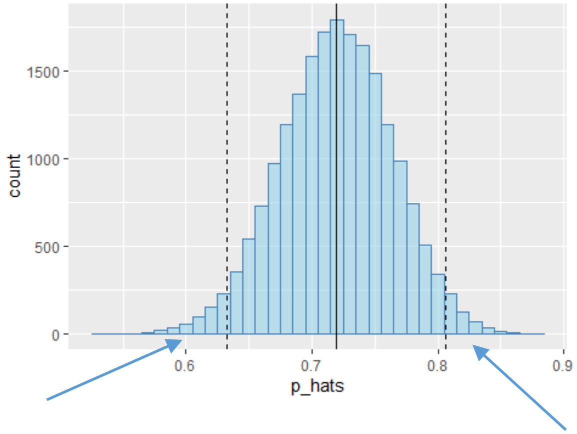
- Four possible scenarios:

		Test Conclusion	
		Fail to reject H_0	Reject H_0
Truth	H_0 is true	Correct	Type I Error
	H_A is true	Type II Error	Correct

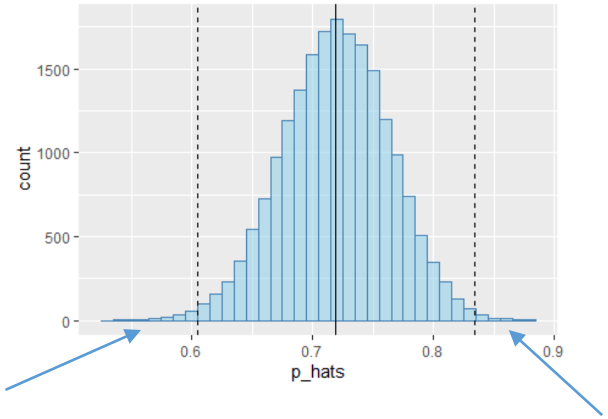
- A **Type I Error** is rejecting H_0 when it is true
- A **Type II Error** is failing to reject H_0 when H_A is true

Errors

- Using a 95% confidence interval and if the null hypothesis happens to be true, an error will be made whenever the point estimate is at least 1.96 standard errors away from the population parameter.
  - This will happen about 5% of the time
  - 2.5% in each tail.



- Using a 99% confidence interval, an error will be made whenever the point estimate is at least 2.58 standard errors away
  - This will happen about 1% of the time
  - 0.5% in each tail.



p-values

- A **p-value** is the probability of observing data at least as favorable to the alternative hypothesis as our current data set, provided the null hypothesis is true

- p-values quantify the strength of the evidence against the null hypothesis and in favor of the alternative hypothesis.
  - Statistical hypothesis tests normally use p-values instead of deciding based on a confidence interval.

p-values

- 5000 Americans are randomly polled on if they approve of president's policies
- H_0
  - Half approve, half disapprove
  - $p = 0.5$
- H_A
  - Something other than half approving, and half disapproving
  - $p \neq 0.5$

p-values

- The poll find 61% of Americans approve of president's policies
  - Is 61% a real difference from H_0?
  - What would the sampling distribution look like if H_0 was true?
- If the null hypothesis were true, the population proportion would be the null value, $p_0 = 0.5$
- Need to check if the distribution is approximately normal

p-values

- Independence: Random sampling
- Success-Failure:
  - Checked using the null value, p_0, instead of p
  - $5000 \times 0.5 \ \geq 10$ and $5000 \times (1 - 0.5) \ \geq 10$
  - $2500 \ \geq 10$ and $2500 \geq 10$

- Approximately normal

p-values

- The null value p_0 is again used instead of p for calculating the standard error
- $SE_{p_0} = \sqrt{\frac{0.5(1-0.5)}{5000}} = 0.007$

- If the null hypothesis is true, then the sample proportion should follow a normal distribution with mean 0.5 and a standard error of 0.007
  - This is called the **null distribution**

p-values

- If the null hypothesis were true, determine the chance of finding $\hat{p}$ at least as far into the tails as 0.61 under the null distribution, which is a normal distribution with mean = 0.5 and SE = 0.007
- This is a normal probability problem where x = 0.61. To find the tail area, start by computing the Z-score using the mean of 0.5 and the standard error of 0.007:
  - $Z = \frac{0.61-0.5}{0.007} = 15.71$ (positive Z-score)

```
> 1-pnorm(15.71, mean=.5)
[1] 0
```

This indicates no data is to the right of the Z-score.
In other words, ALL data (100%) is to the left

p-values

```
> 1-pnorm(15.71, mean=.5)
[1] 0
```

- If the null hypothesis was true (Americans are evenly split) then this p-value tells us it would be impossible have poll results of 61% approval.
  - Rejecting the null hypothesis in favor of the alternative hypothesis (Americans were not evenly split) is most plausible.
- **Significance** is strong evidence in favor or against a hypothesis.
  - Represented as α
  - Typical significance is 0.05

p-values

- The p-value in this example is 0
  - There is a zero percent chance for a 61% approval rating in the null distribution

- Since 0 is less than a significance of 0.05, this provides strong evidence against the null hypothesis.

Conducting a Hypothesis Test

- *A simple random sample of 1028 US adults in March 2013 show that 56% support nuclear arms reduction. Does this provide convincing evidence that a majority of Americans supported nuclear arms reduction at the 5% significance level?*

Conducting a Hypothesis Test

1. Prepare
   - Identify the parameter of interest, list hypotheses, identify the significance level, and identify $\hat{p}$ and n .
2. Check
   - Verify conditions to ensure $\hat{p}$ is nearly normal under H_0. For one-proportion hypothesis tests, use the null value to check the success-failure condition.
3. Calculate
   - If the conditions hold, compute the standard error, again using p_0, compute the Z-score, and identify the p-value.
4. Conclude
   - Evaluate the hypothesis test by comparing the p-value to α and provide a conclusion in the context of the problem.

Conducting a Hypothesis Test

- *A simple random sample of 1028 US adults in March 2013 show that 56% support nuclear arms reduction. Does this provide convincing evidence that a majority of Americans supported nuclear arms reduction at the 5% significance level?*

- Prepare:
  - $n = 1028$
  - $\hat{p} = 0.56$
  - $\alpha = 0.05$
  - $H_0 =$ Half support, half against
    - $p = 0.5$
  - $H_A =$ Something other than half supporting and half against
    - $p \neq 0.5$

Conducting a Hypothesis Test

- *A simple random sample of 1028 US adults in March 2013 show that 56% support nuclear arms reduction. Does this provide convincing evidence that a majority of Americans supported nuclear arms reduction at the 5% significance level?*

- Check:
  - Independence: Random sampling
  - Success-Failure:
    - Checked using the null value, p_0, instead of p
    - $1028 \times 0.5 \ \geq 10$ and $1028 \times (1 - 0.5) \ \geq 10$
    - $514 \ \geq 10$ and $514 \geq 10$

Conducting a Hypothesis Test

- Calculate:
  - $SE_{p_0} = \sqrt{\frac{0.5(1-0.5)}{1028}} = 0.0156$
  - $Z = \frac{0.56-0.5}{0.0156} = 3.85$
  - p-value = 0.000059 = 0.0001

```
> 1-pnorm(3.85)
[1] 5.905891e-05
```

Conducting a Hypothesis Test

- *A simple random sample of 1028 US adults in March 2013 show that 56% support nuclear arms reduction. Does this provide convincing evidence that a majority of Americans supported nuclear arms reduction at the 5% significance level?*

- Conclude:
  - P-value is less than 0.05 significance level
  - Reject H_0

  - Provides strong evidence that a majority of Americans did support nuclear arms reduction efforts in March 2013

Choosing a Significance Level

- The traditional significance level is 0.05.
  - It can be helpful to adjust the significance level.

- A level that is smaller or larger than 0.05 may be selected depending on the consequences of conclusions reached from the hypothesis test.

- If making a Type 1 Error is unsafe or very costly, choose a small significance level, like 0.01, to be very cautious about rejecting the null hypothesis.
  - This significance level requires stronger evidence favoring H_A before rejecting H_0.

Choosing a Significance Level

- If a Type 2 Error is relatively more dangerous or more costly than a Type 1 Error, then choose a higher significance level, like 0.10, to exercise more caution about failing to reject H_0 when the alternative hypothesis is true.
  - This significance level allows more evidence for failing to reject H_0