Sampling Distribution of a Sample Proportion

Lecture 25
Sections 8.1 - 8.2

Robb T. Koether

Hampden-Sydney College

Fri, Mar 2, 2012

Outline

Outline

- Suppose that a population is 50% male and 50% female.

- Suppose that a population is 50% male and 50% female.
- We take a sample of 25 and compute the sample proportion of males.

- Suppose that a population is 50% male and 50% female.
- We take a sample of 25 and compute the sample proportion of males.
- The sample proportion could be anything from 0% to 100%, depending on the sample.

- Suppose that a population is 50% male and 50% female.
- We take a sample of 25 and compute the sample proportion of males.
- The sample proportion could be anything from 0% to 100%, depending on the sample.
- But it is probably close to 50%.

- Suppose that a population is 50% male and 50% female.
- We take a sample of 25 and compute the sample proportion of males.
- The sample proportion could be anything from 0% to 100%, depending on the sample.
- But it is probably close to 50%.
- We will simulate this situation.

- Use `randInt(0,1,25)` to randomly select twenty-five 0's and 1's.

- Use `randInt(0,1,25)` to randomly select twenty-five 0's and 1's.
- Interpret 1 as male, 0 as female.

- Use `randInt(0,1,25)` to randomly select twenty-five 0's and 1's.
- Interpret 1 as male, 0 as female.
- Then apply the `sum` function to count the 1's.

- Use `randInt(0,1,25)` to randomly select twenty-five 0's and 1's.
- Interpret 1 as male, 0 as female.
- Then apply the `sum` function to count the 1's.
- Compute the sample proportion of males.

- Use `randInt(0,1,25)` to randomly select twenty-five 0's and 1's.
- Interpret 1 as male, 0 as female.
- Then apply the `sum` function to count the 1's.
- Compute the sample proportion of males.
- A simpler way to do this is to enter `randBin(25,.5)/25`.

- Use `randInt(0,1,25)` to randomly select twenty-five 0's and 1's.
- Interpret 1 as male, 0 as female.
- Then apply the `sum` function to count the 1's.
- Compute the sample proportion of males.
- A simpler way to do this is to enter `randBin(25,.5)/25`.
- Repeat this many times until we can see the shape of the distribution.

Outline

Experimental Results

Count	Samp. Prop.	Frequency
6	0.24	3
7	0.28	6
8	0.32	15
9	0.36	31
10	0.40	52
11	0.44	68
12	0.48	85
13	0.52	73
14	0.56	62
15	0.60	44
16	0.64	33
17	0.68	19
18	0.72	5
19	0.76	3
20	0.80	1

I used `randBin(25,.5,25)` twenty times (seed = 161).

Notation

- The letter p represents the population proportion (a parameter).

Notation

- The letter p represents the population proportion (a parameter).
- In our experiment, $p = 0.50$.

Notation

- The letter p represents the population proportion (a parameter).
- In our experiment, $p = 0.50$.
- The symbol $\hat{p}$ ("p-hat") represents a sample proportion (a statistic).

Notation

- The letter p represents the population proportion (a parameter).
- In our experiment, $p = 0.50$.
- The symbol $\hat{p}$ ("p-hat") represents a sample proportion (a statistic).
- In our experiment, the value of $\hat{p}$ varied from sample to sample.

Notation

- The letter p represents the population proportion (a parameter).
- In our experiment, $p = 0.50$.
- The symbol $\hat{p}$ ("p-hat") represents a sample proportion (a statistic).
- In our experiment, the value of $\hat{p}$ varied from sample to sample.
- $\hat{p}$ was as low as 0.24 and as high as 0.80.

Experimental Results

- What was the *average* value of $\hat{p}$ in my experiment?

Experimental Results

- What was the *average* value of $\hat{p}$ in my experiment?
- It turned out to be 0.4993.

Experimental Results

- What was the *average* value of $\hat{p}$ in my experiment?
- It turned out to be 0.4993.
- What "should" it have been?

Experimental Results

- What was the *average* value of $\hat{p}$ in my experiment?
- It turned out to be 0.4993.
- What "should" it have been?
- Imagine the set of all possible samples of size 25 from the male/female population.

Experimental Results

- What was the *average* value of $\hat{p}$ in my experiment?
- It turned out to be 0.4993.
- What "should" it have been?
- Imagine the set of all possible samples of size 25 from the male/female population.
- That itself is a population.

Experimental Results

- What was the *average* value of $\hat{p}$ in my experiment?
- It turned out to be 0.4993.
- What "should" it have been?
- Imagine the set of all possible samples of size 25 from the male/female population.
- That itself is a population.
- It is the population of *all* samples of size 25.

Experimental Results

- What was the *average* value of $\hat{p}$ in my experiment?
- It turned out to be 0.4993.
- What "should" it have been?
- Imagine the set of all possible samples of size 25 from the male/female population.
- That itself is a population.
- It is the population of *all* samples of size 25.
- I sampled only 500 of its members.

Experimental Results

- What was the *average* value of $\hat{p}$ in my experiment?
- It turned out to be 0.4993.
- What "should" it have been?
- Imagine the set of all possible samples of size 25 from the male/female population.
- That itself is a population.
- It is the population of *all* samples of size 25.
- I sampled only 500 of its members.
- Theory says that the average value of $\hat{p}$ should be p.

Experimental Results

- What was the *average* value of $\hat{p}$ in my experiment?
- It turned out to be 0.4993.
- What "should" it have been?
- Imagine the set of all possible samples of size 25 from the male/female population.
- That itself is a population.
- It is the population of *all* samples of size 25.
- I sampled only 500 of its members.
- Theory says that the average value of $\hat{p}$ should be p.
- Our experimental results were very close!

Experimental Results

- How much did $\hat{p}$ vary?

Experimental Results

- How much did $\hat{p}$ vary?
- That is, what was the standard deviation of $\hat{p}$ in our experiment?

Experimental Results

- How much did $\hat{p}$ vary?
- That is, what was the standard deviation of $\hat{p}$ in our experiment?
- It turned out to be 0.0988.

Experimental Results

- How much did $\hat{p}$ vary?
- That is, what was the standard deviation of $\hat{p}$ in our experiment?
- It turned out to be 0.0988.
- What "should" it have been?

Experimental Results

- How much did $\hat{p}$ vary?
- That is, what was the standard deviation of $\hat{p}$ in our experiment?
- It turned out to be 0.0988.
- What “should” it have been?
- Imagine again the set of all possible samples of size 25 from the male/female population.

Experimental Results

- How much did $\hat{p}$ vary?
- That is, what was the standard deviation of $\hat{p}$ in our experiment?
- It turned out to be 0.0988.
- What "should" it have been?
- Imagine again the set of all possible samples of size 25 from the male/female population.
- I sampled only 500 of its members.

Experimental Results

- How much did $\hat{p}$ vary?
- That is, what was the standard deviation of $\hat{p}$ in our experiment?
- It turned out to be 0.0988.
- What "should" it have been?
- Imagine again the set of all possible samples of size 25 from the male/female population.
- I sampled only 500 of its members.
- Theory says that the standard deviation of $\hat{p}$ should be $\sqrt{\frac{p(1-p)}{n}}$.

Experimental Results

- How much did $\hat{p}$ vary?
- That is, what was the standard deviation of $\hat{p}$ in our experiment?
- It turned out to be 0.0988.
- What "should" it have been?
- Imagine again the set of all possible samples of size 25 from the male/female population.
- I sampled only 500 of its members.
- Theory says that the standard deviation of $\hat{p}$ should be $\sqrt{\frac{p(1-p)}{n}}$.
- That turns out to be

$$\sqrt{\frac{(0.50)(0.50)}{25}} = \sqrt{\frac{0.25}{25}} = \sqrt{0.01} = 0.10.$$

Experimental Results

- How much did $\hat{p}$ vary?
- That is, what was the standard deviation of $\hat{p}$ in our experiment?
- It turned out to be 0.0988.
- What "should" it have been?
- Imagine again the set of all possible samples of size 25 from the male/female population.
- I sampled only 500 of its members.
- Theory says that the standard deviation of $\hat{p}$ should be $\sqrt{\frac{p(1-p)}{n}}$.
- That turns out to be

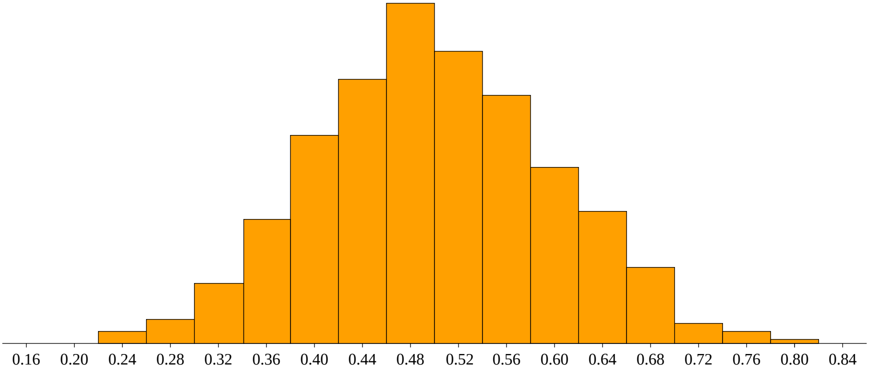
$$\sqrt{\frac{(0.50)(0.50)}{25}} = \sqrt{\frac{0.25}{25}} = \sqrt{0.01} = 0.10.$$

- Our experimental results again were very close!

Experimental Results

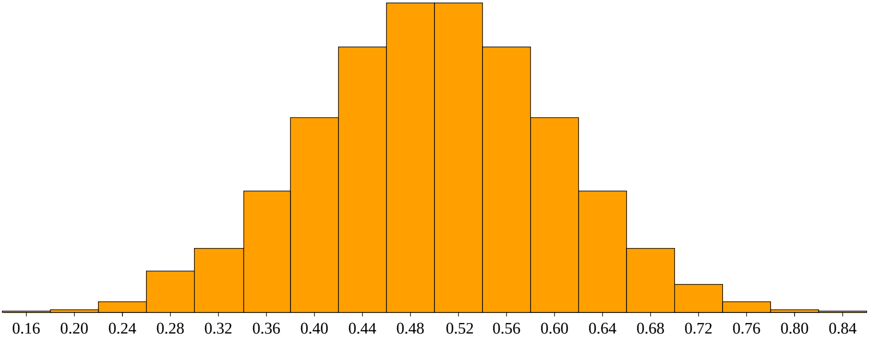
- We also notice that the distribution of values was nearly normal.

Distributions



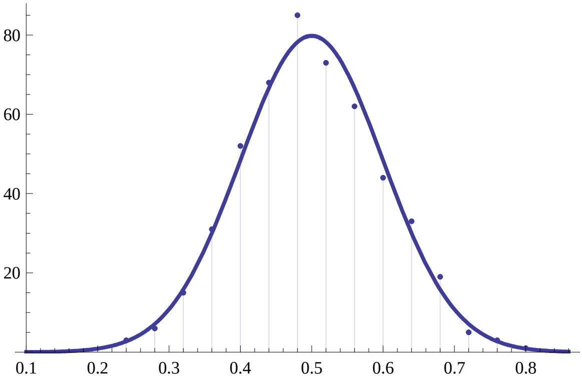
Our results

Distributions



Theoretical results

Distributions



A normal curve

The Sampling Distribution of $\hat{p}$

- Theory tells us that,

$$\mu_{\hat{p}} = p$$

and

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

The Sampling Distribution of $\hat{p}$

- Theory tells us that,

$$\mu_{\hat{p}} = p$$

  and

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

- It also says that for *large* samples, $\hat{p}$ has a normal distribution.

The Sampling Distribution of $\hat{p}$

- Theory tells us that,

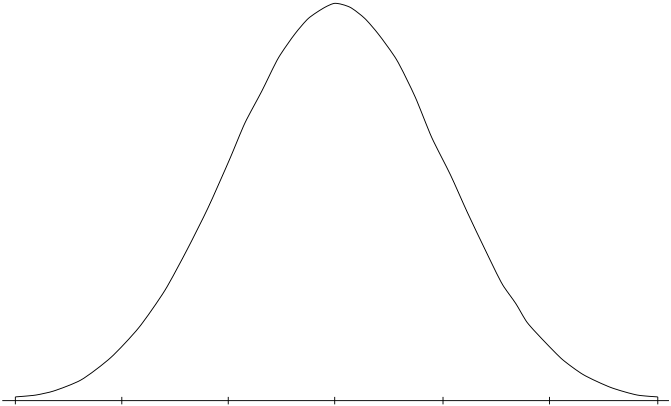
$$\mu_{\hat{p}} = p$$

and

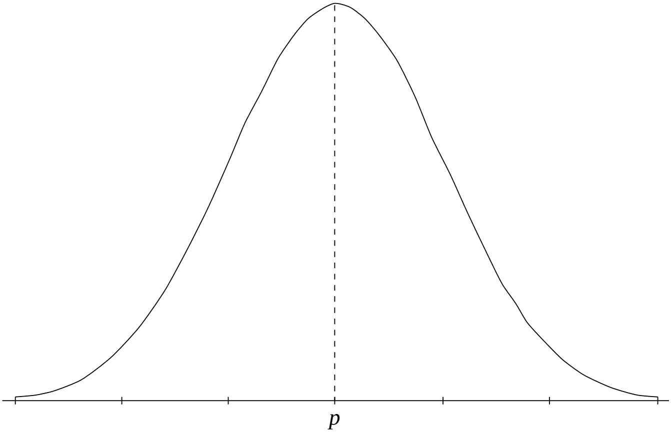
$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

- It also says that for *large* samples, $\hat{p}$ has a normal distribution.
- This is called the Central Limit Theorem for Proportions.

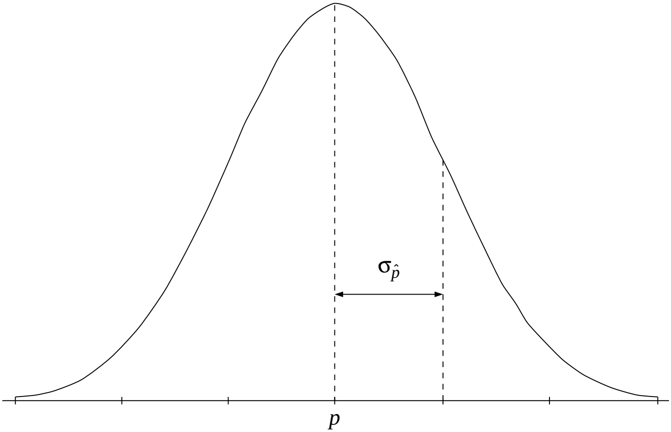
The Sampling Distribution of $\hat{p}$



The Sampling Distribution of $\hat{p}$



The Sampling Distribution of $\hat{p}$



Example (Sampling Distribution of $\hat{p}$)

- Suppose the president's approval rating is 50%.
- If we sample 100 people, how likely is it that our sample proportion will be between 45% and 55%?

Example (Sampling Distribution of $\hat{p}$)

- From the Central Limit Theorem, we know that $\hat{p}$ has a normal distribution with mean 0.50 and standard deviation

$$\sqrt{\frac{(0.50)(0.50)}{100}} = 0.05.$$

- So the probability that $\hat{p}$ will be between 0.45 and 0.55 is

$$\mathit{normalcdf}(.45, .55, .50, .05) = 0.6827.$$

Example (Sampling Distribution of $\hat{p}$)

- Suppose the president's approval rating is 45%.
- Suppose we sample only 10 people, how likely is it that our sample proportion will be between 40% and 60%?
- This is harder.

Sampling Distributions of $\hat{p}$

Example (Sampling Distribution of $\hat{p}$)

- Suppose the president's approval rating is 45%.
- Suppose we sample only 10 people, how likely is it that our sample proportion will be between 40% and 60%?
- This is harder.
- Why?

Outline

Assignment

Homework

1. Given the population in which 25% strongly approve of the president's performance,
   (a) Find the sampling distribution of $\hat{p}$ for samples of size 3.
   (b) Use the Central Limit Theorem to find the mean and standard deviation of that distribution.
2. Repeat the previous exercise for a population in which 10% strongly approve of the president's performance.
3. Use the Central Limit Theorem to find the distribution of $\hat{p}$ when the sample size is $n = 100$.