Sampling Distribution of a Sample Proportion

Lecture 25
Sections 8.1 - 8.2

Robb T. Koether

Hampden-Sydney College

Fri, Feb 26, 2010

Outline

Outline

The Sample Proportion

- The letter p represents the population proportion.
- The symbol $\hat{p}$ ("p-hat") represents the sample proportion.
- We see from our experiment that $\hat{p}$ takes on different values at random, depending on the sample.
- Therefore, $\hat{p}$ is a random variable.

The Random Variable $\hat{p}$

- As a random variable, $\hat{p}$ has a probability distribution.
- Can we describe its pdf?

The Random Variable $\hat{p}$

- As a random variable, $\hat{p}$ has a probability distribution.
- Can we describe its pdf?
- Yes

The Sampling Distribution of $\hat{p}$

- Theory says that
  - The average value of $\hat{p}$ is p and
  - The standard deviation of $\hat{p}$ is

$$\sqrt{\frac{p(1-p)}{n}}.$$

- Furthermore, if the sample size is large enough, the shape of the distribution is normal.

The Sampling Distribution of $\hat{p}$

- That is,

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

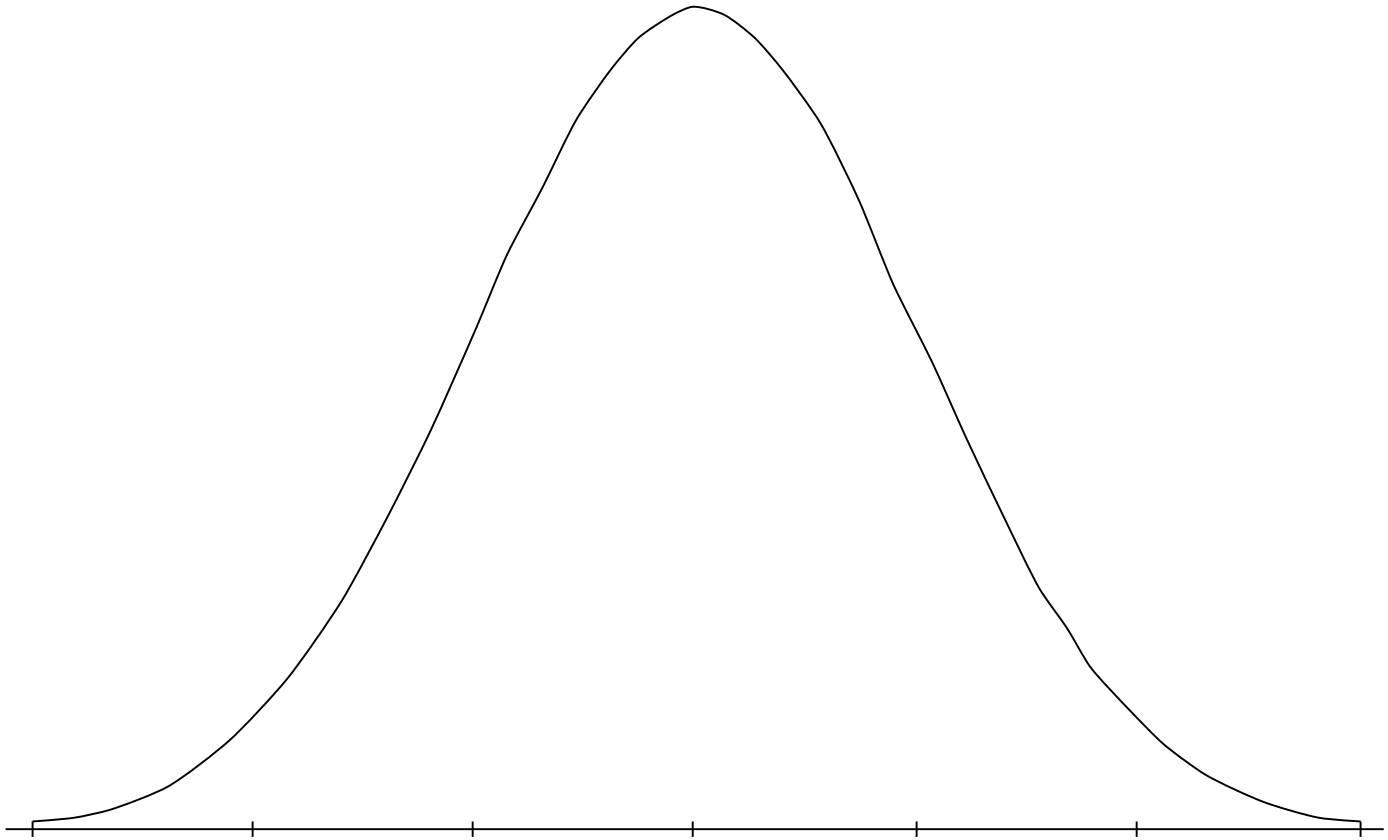
and

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

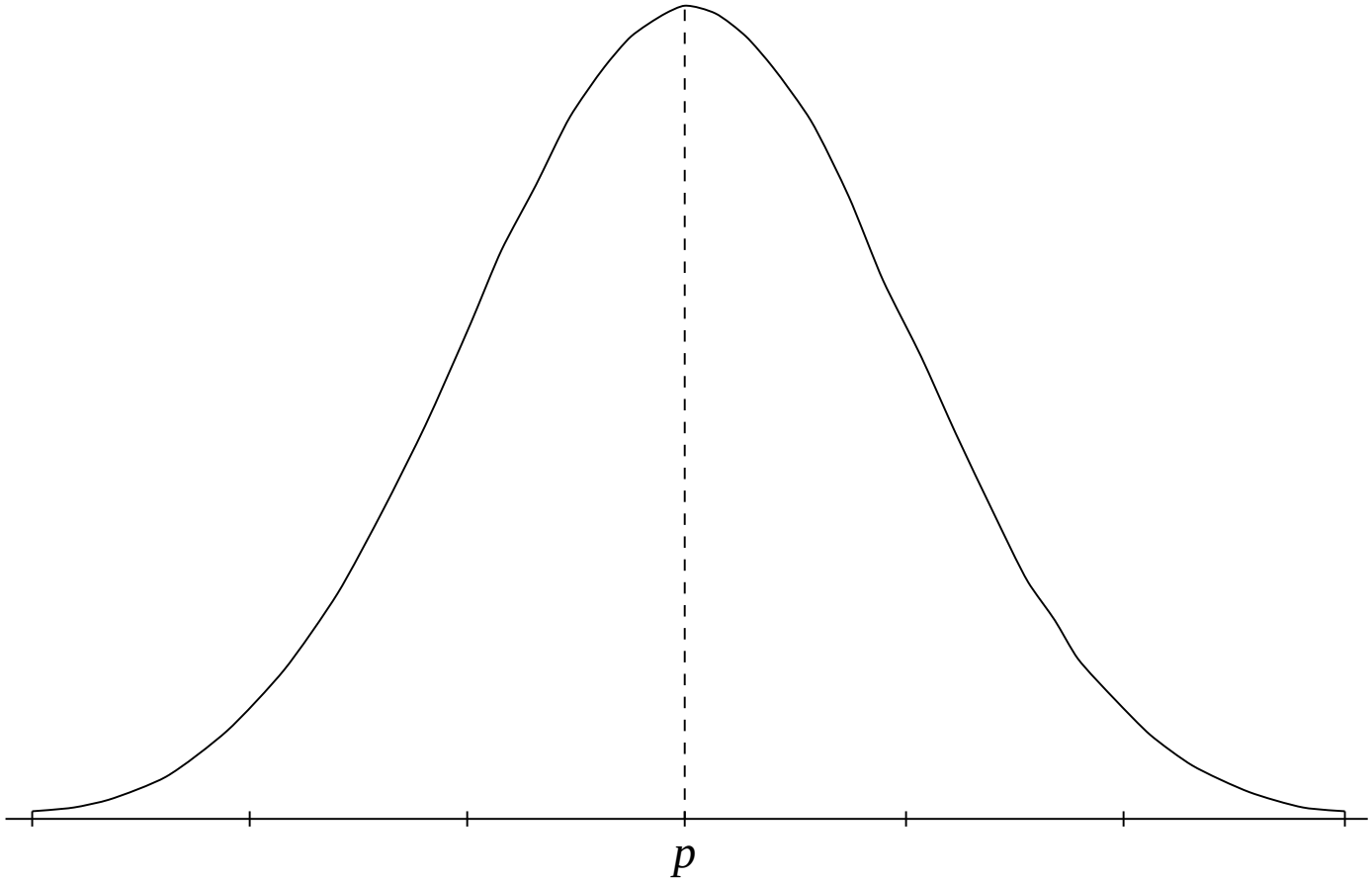
and for large samples,

$$\hat{p} \text{ is } N\left(p, \sqrt{\frac{p(1-p)}{n}}\right).$$

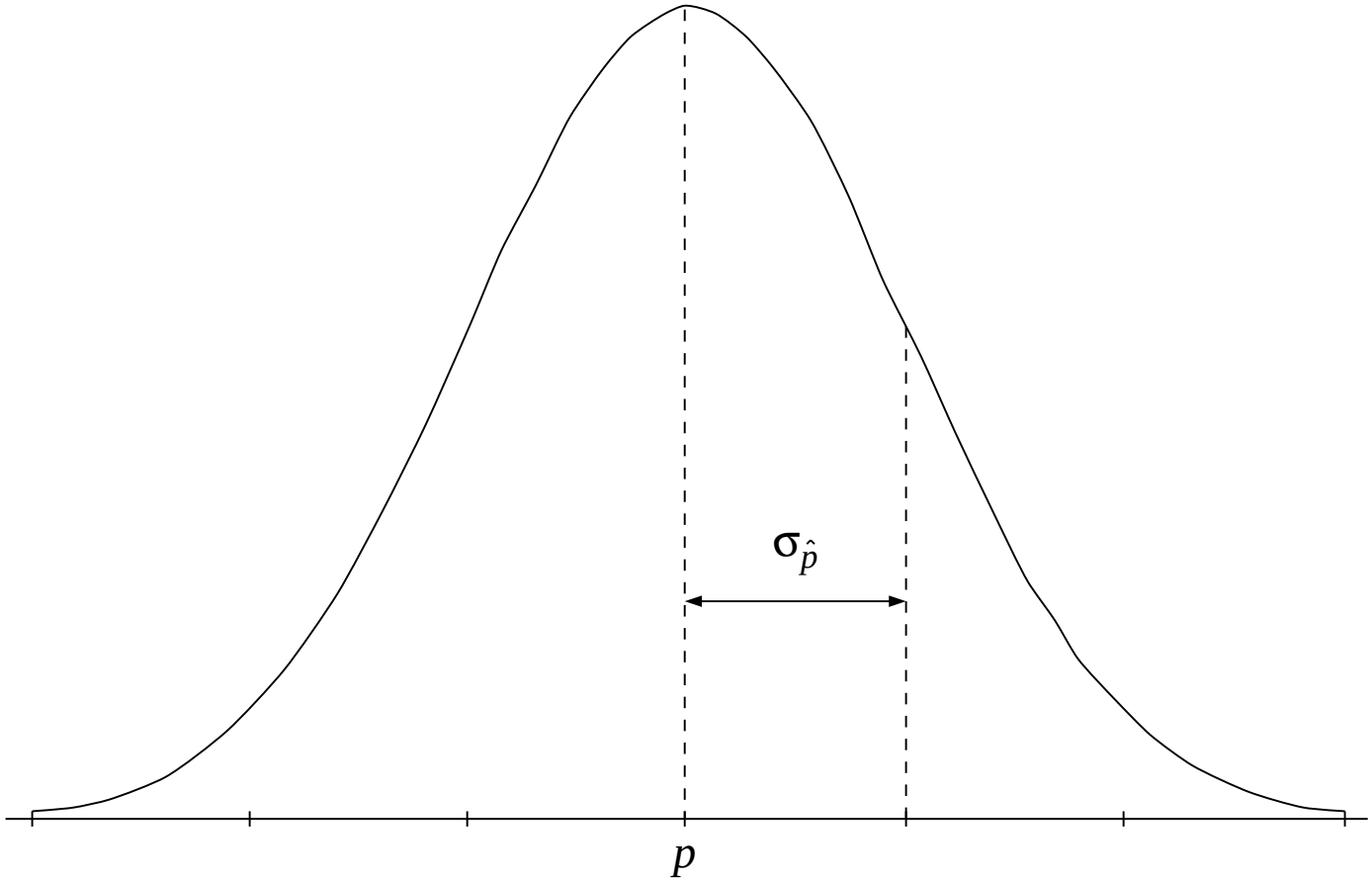
The Sampling Distribution of $\hat{p}$



The Sampling Distribution of $\hat{p}$



The Sampling Distribution of $\hat{p}$



The Sampling Distribution of $\hat{p}$

- Therefore, in our experiment, where $p = 0.40$,

$$\mu_{\hat{p}} = 0.40$$

and

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.40)(0.60)}{10}} = 0.1549.$$

- However, a sample size of 10 is not large enough for the shape to be normal.

Outline

Sampling Distributions

Definition (Sampling Distribution of a Statistic)

The sampling distribution of a statistic is the distribution of values of that statistic over all possible samples of a given size n from the population.

- We may sample with or without replacement.

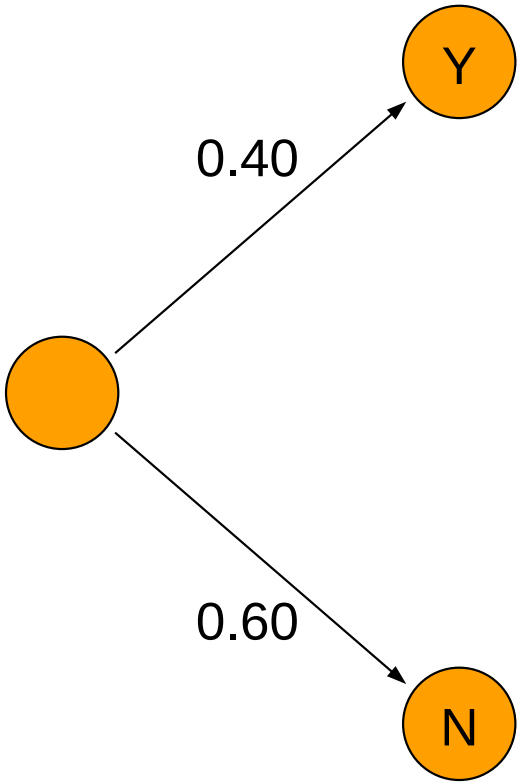
The Sample Proportion

- We will work out the sampling distribution for $\hat{p}$ for sample sizes of 1, 2, and 3.
- Then I will show you the sampling distribution for $\hat{p}$ for sample sizes of 4, 5, and 10.

Example

- Suppose that 40% of all people strongly disapprove of President Obama's performance.
- Suppose that we select one person at random.
- We may diagram the 2 possibilities.

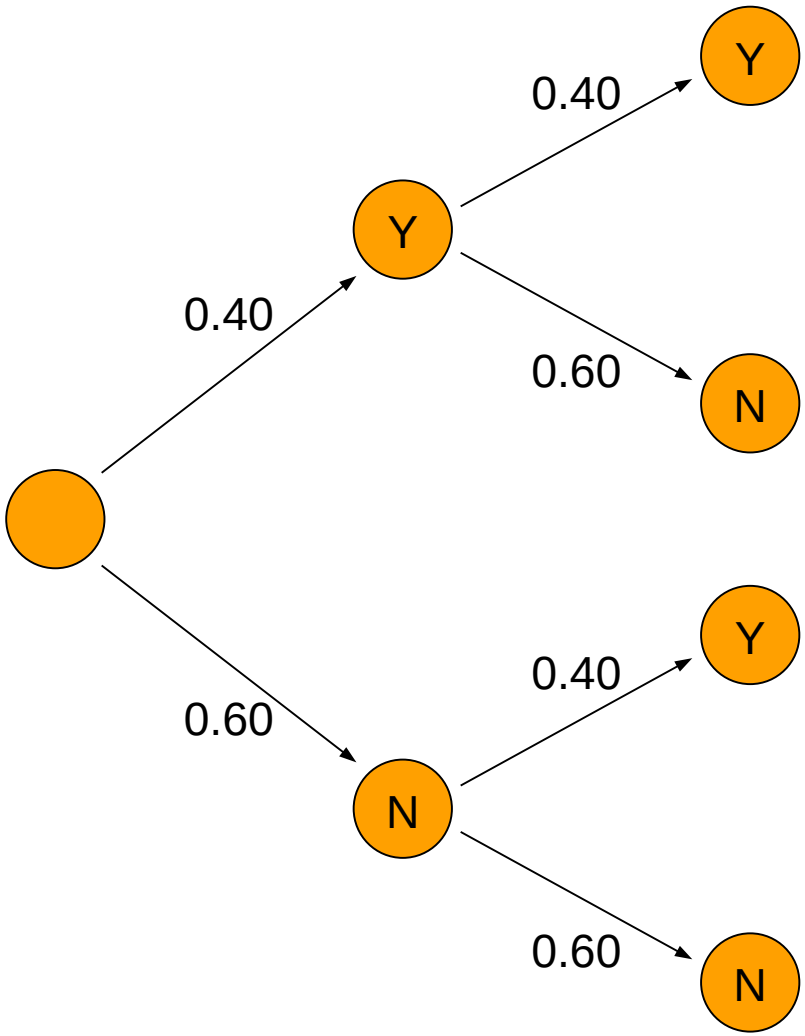
Example



Example

- Now we take a sample of 2 people, sampling with replacement.
- Find the sampling distribution of $\hat{p}$.

Example



YY 2 $(0.40)^2 = 0.16$

YN 1 $(0.40)(0.60) = 0.24$

NY 1 $(0.60)(0.40) = 0.24$

NN 0 $(0.60)^2 = 0.36$

Example

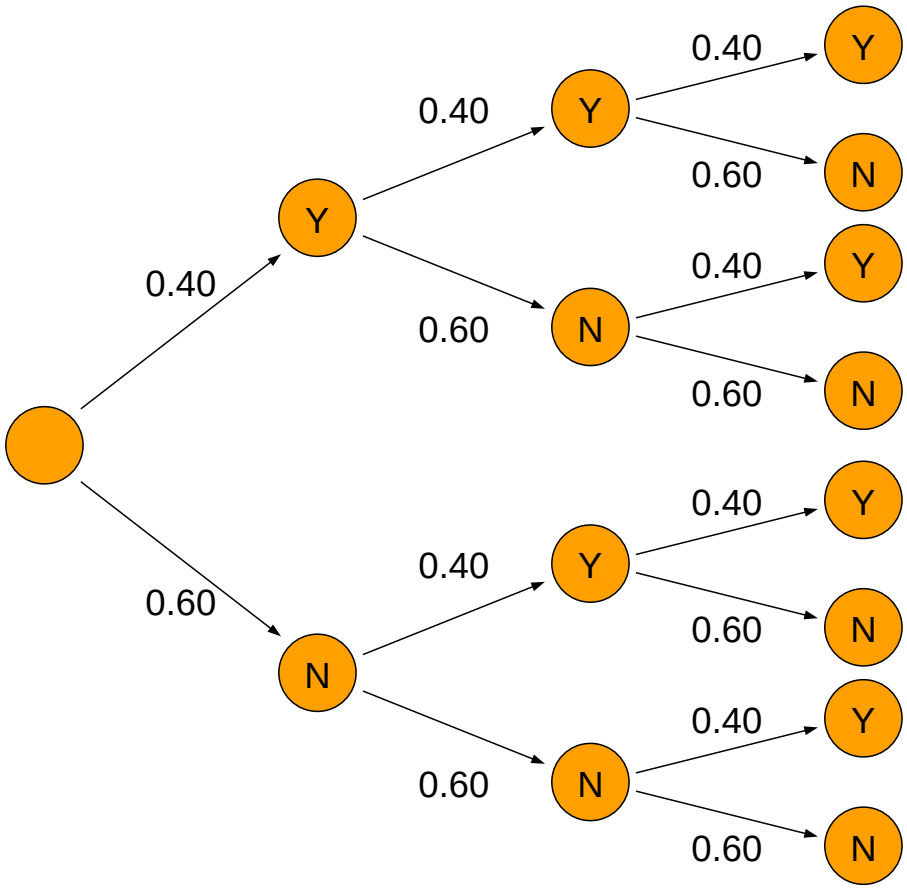
- Let x be the number of people (out of 2) who strong disapprove of President Obama's performance.
- The probability distribution of $\hat{p}$ is

x	$P(x)$
0	0.36
$1/2$	0.48
1	0.16

Example

- Now we take a sample of 3 people, sampling with replacement.
- Find the sampling distribution of $\hat{p}$.

Example



YYY	3	$(0.40)^3 = 0.064$
YYN	2	$(0.40)^2(0.60) = 0.096$
YNY	2	$(0.40)(0.60)(0.40) = 0.096$
YNN	1	$(0.40)(0.60)^2 = 0.144$
NYY	2	$(0.60)(0.40)^2 = 0.096$
NYN	1	$(0.60)(0.40)(0.60) = 0.144$
NNY	1	$(0.60)^2(0.40) = 0.144$
NNN	0	$(0.60)^3 = 0.216$

Example

- Let $\hat{p}$ be the sample proportion of people who strong disapprove of President Obama's performance.
- The sampling distribution of $\hat{p}$ is

$\hat{p}$	$P(\hat{p})$
0	0.216
$1/3$	0.432
$2/3$	0.288
1	0.064

Samples of Size $n = 4$

- If we sample 4 people, then the sampling distribution of the sample proportion is

$\hat{p}$	$P(\hat{p})$
0	0.1296
$1/4$	0.3456
$2/4$	0.3456
$3/4$	0.1536
1	0.0256

Samples of Size $n = 5$

- If we sample 5 people, then the sampling distribution of the sample proportion is

$\hat{p}$	$P(\hat{p})$
0	0.0778
$1/5$	0.2592
$2/5$	0.3456
$3/5$	0.2304
$4/5$	0.0768
1	0.0102

Samples of Size $n = 6$

- If we sample 6 people, then the sampling distribution of the sample proportion is

$\hat{p}$	$P(\hat{p})$
0	0.0467
$1/6$	0.1866
$2/6$	0.3110
$3/6$	0.2765
$4/6$	0.1382
$5/6$	0.0369
1	0.0041

Samples of Size $n = 10$

- If we sample 10 people, then the sampling distribution of the sample proportion is

$\hat{p}$	$P(\hat{p})$
0.00	0.0060
0.10	0.0403
0.20	0.1209
0.30	0.2150
0.40	0.2508
0.50	0.2007
0.60	0.1115
0.70	0.0425
0.80	0.0106
0.90	0.0016
1.00	0.0001

Our Experiment

- Therefore, in our experiment, if we collected 100 samples of size 10, then we would expect

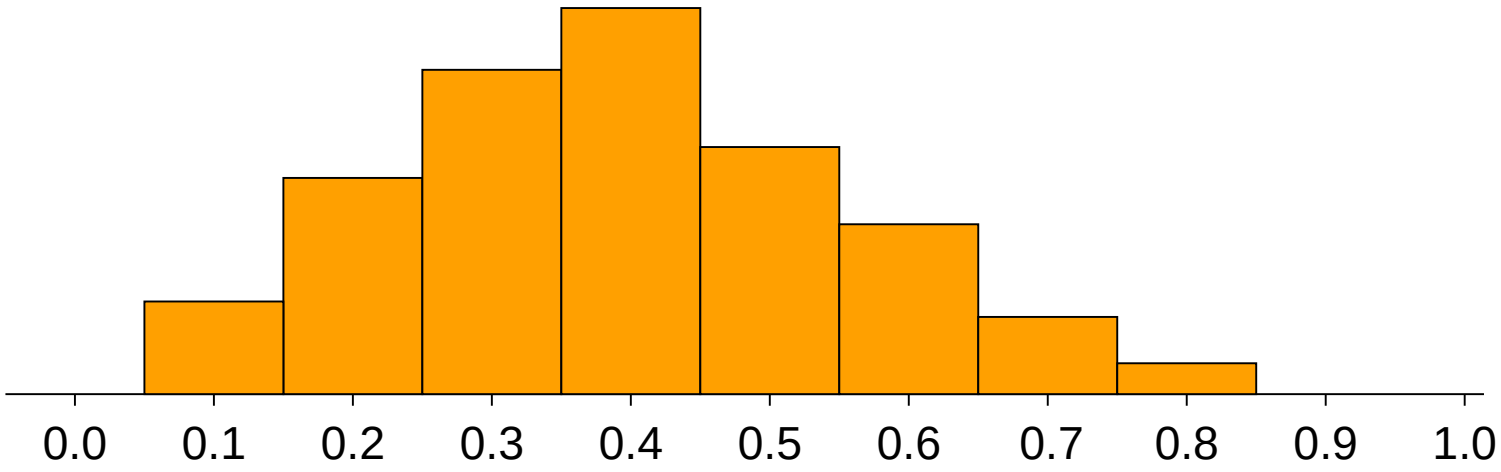
Sample Proportion	Expected No. of Samples	Actual No. of Samples	Difference
0.00	1	0	-1
0.10	4	6	$+2$
0.20	12	14	$+2$
0.30	22	21	-1
0.40	25	25	0
0.50	20	16	-4
0.60	11	11	0
0.70	4	5	$+1$
0.80	1	2	$+1$
0.90	0	0	0
1.00	0	0	0

Our Experiment

- Theory predicts that the average of our data set will be 0.40 and the standard deviation will be 0.1549.
- Find the average and standard deviation of the frequency distribution

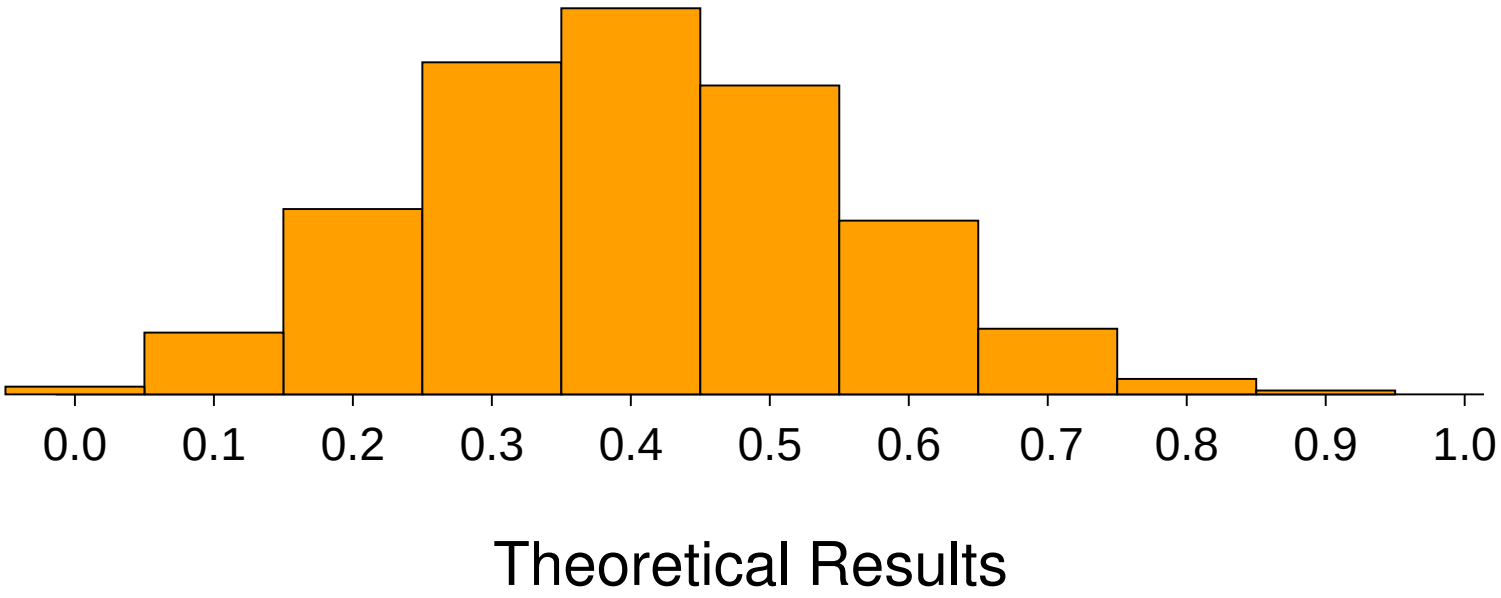
Sample Proportion	No. of Samples
0.00	0
0.10	6
0.20	14
0.30	21
0.40	25
0.50	16
0.60	11
0.70	5
0.80	2
0.90	0
1.00	0

Our Experiment



Our Results

Our Experiment



Outline

Assignment

Homework

- Catch up on old homework.