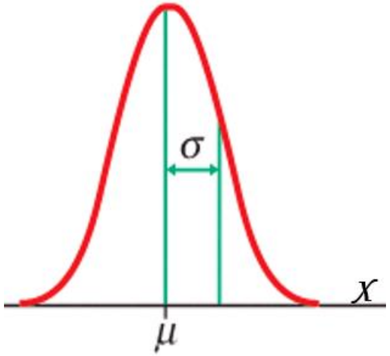
Sampling Distribution of the Sample Proportion

Distribution of X:

$\mu_X = \mu$

$\sigma_X = \sigma$



Let X=height or height.
$X \sim N(\mu,\sigma)$

$$z = \frac{(x-\mu)}{\sigma}$$

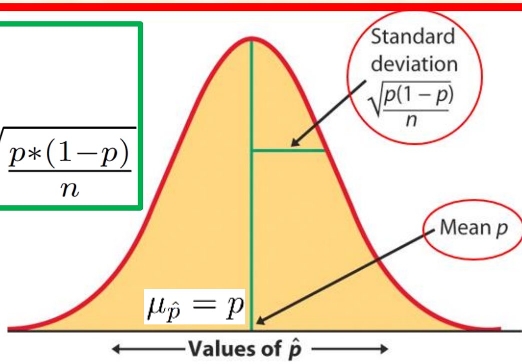
Q: To find Probability with common three steps:

Step 1. Standardize; That is, to find Z-score;
Step2: Draw N(0, 1) and shade;
Step3: Find:
Prob=Area
=NCDF(low, high, 0, 1)

Sampling dist. of sample Prop = Distribution of $\hat{p}$

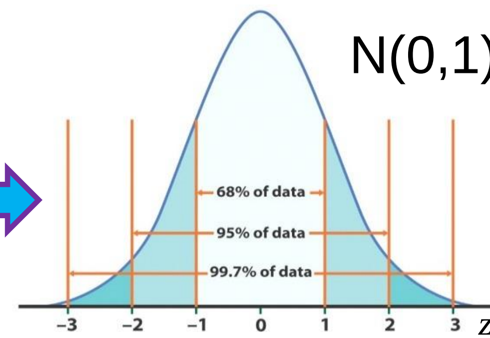
$\mu_{\hat{p}} = p$

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



$\hat{p}$ follows Approximately $N(p, \sqrt{\frac{p(1-p)}{n}})$

Standardize: Z-score of $\hat{p} = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$;

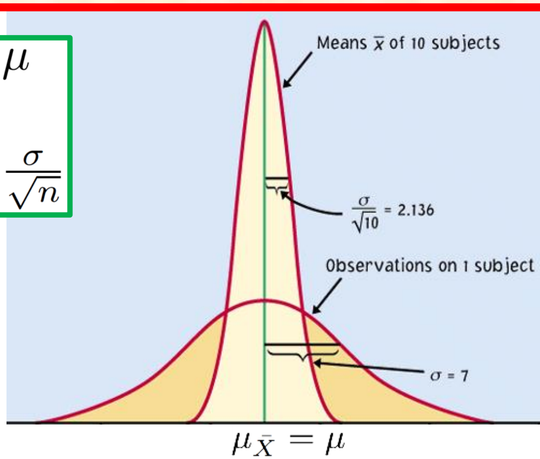


Standardized height (no units)

Sampling dist. of sample Mean = Distribution of

$\mu_{\bar{X}} = \mu$

$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$



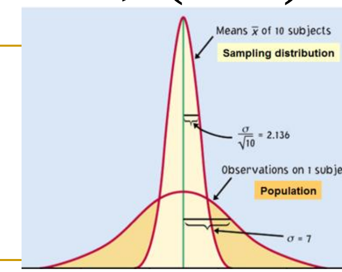
$\bar{X}$ follows Approximately $N(\mu, \frac{\sigma}{\sqrt{n}})$

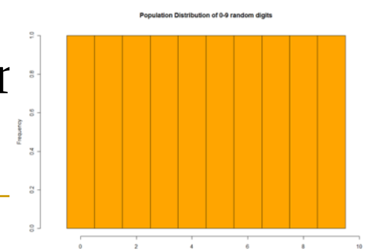
Standardize: Z-score of $\bar{X} = \frac{\bar{X}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$

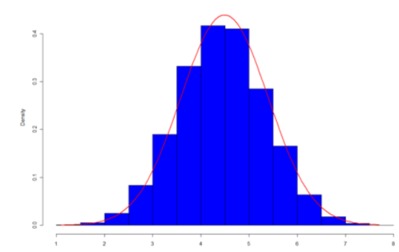
Sampling distribution of a sample mean=distribution of $\bar{X}$

Population

Distribution of X, (n=1):	Sampling distribution of $\bar{X}$, (n>1) :
Exact $N(\mu, \sigma)$	Exact $N(\mu, \frac{\sigma}{\sqrt{n}})$
Not exact Normal, but with Mean μ, and SD σ	Approximately $N(\mu, \frac{\sigma}{\sqrt{n}})$ (By Centr



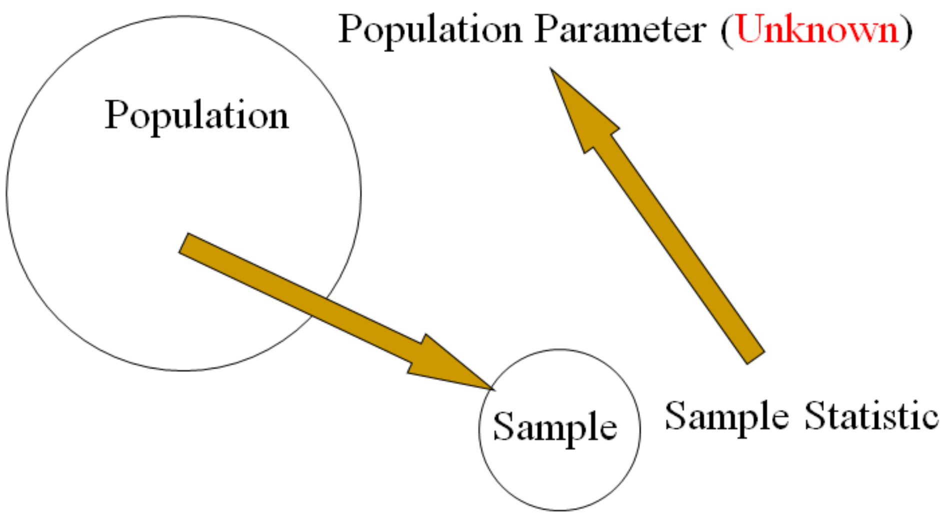




Standardize: Z-score of $\bar{X} = \frac{\bar{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$; Reverse: $\bar{X} = \mu + \left(\frac{\sigma}{\sqrt{n}}\right) * Z$

Simple random sample (SRS)

Sampling Distribution Models



Simple Random Sample (SRS)

- Parameter (Population Characteristics)
  - μ (mean)
  - p (proportion)
- Statistic (Sample Characteristics)
  - $\bar{x}$ (sample mean)
  - $\hat{p}$ (sample proportion)

Data are summarized by statistics (mean, standard deviation, median, quartiles, correlation, etc..)

Concerns:

1) Is sample proportion related to population proportion?

2) If yes, what will be the relationship? Or say, how far or how close is a sample proportion away from the population proportion?

Review: Sampling proportion p-hat

Sample proportion: (p-hat, or **relative frequency**)

$$\hat{p} = \frac{\text{count in the sample}}{Total}$$

Population proportion: p

- Example: At UNCW, p=0.3 *(population proportion of Hispanics)*
- In a random sample of 50 students in an undergrad class, 10 are Hispanic:

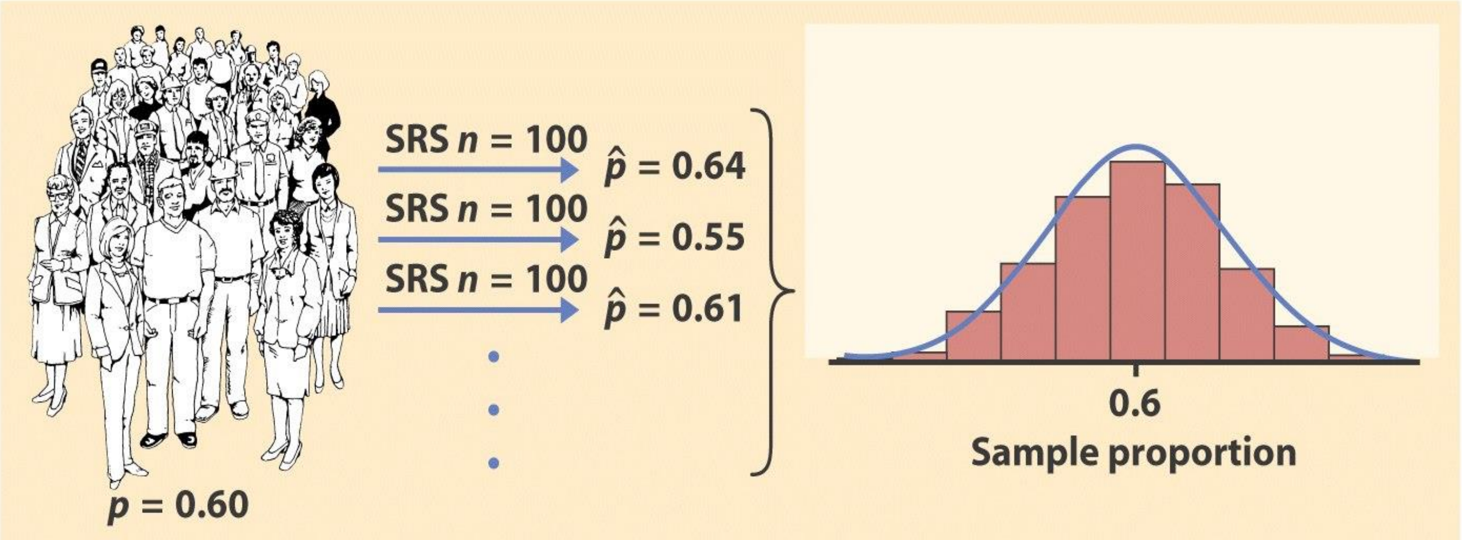
  $\hat{p}$ = (10)/(50) = 0.2 *(sample proportion of Hispanics)*

- The 30 subjects in an SRS are asked to taste an unmarked brand of coffee and rate it "would buy" or "would not buy." Eighteen subjects rated the coffee "would buy."
- $\hat{p}$ = (18)/(30) = 0.6 *(proportion of "would buy")*

Review: Sampling variability

Each time we take a random sample from a population, we are likely to get a different set of individuals and calculate a different statistic. This is called sampling variability.

If we take a lot of random samples of the same size from a given population, the variation from sample to sample—the **sampling distribution**—will follow a predictable pattern.



Sampling Distribution of sample proportion of 10 random digits

(1) Select 10 random digits from Table B, and then take the sample proportion of EVEN numbers;

(2) Repeat this process 4 times for each student from Dr. Chen's class.

More details with illustration:

1. Based on Table B (**random digit table**), we randomly select a line, for example line 106 in this case:

106	68417	35013	15529	72765	85089	57067	50211	47487
107	82739	57890	20807	47511	81676	55300	94383	14893
108	60940	72024	17868	24943	61790	90656	87964	18883
109	36009	19365	15412	39638	85453	46816	83485	41979
110	38448	48789	18338	24697	39364	42006	76688	08708

2. Take sample proportion of EVEN numbers of random digits of (6, 8, 4, 1, 7, 3, 5, 0, 1, 3). We will have sample proportion of EVEN #'s and gives

sample proportion #1 = 4/10=0.4;

Now we move forward to another set of 10 random digits of (1, 5, 5, 2, 9, 7, 2, 7, 6, 5), and we will have sample mean and gives

sample proportion #2 = 3 /10=0.3;

Repeat this procedure 4 times until you get sample proportion #4.

Sampling Distribution of sample mean of 10 random digits

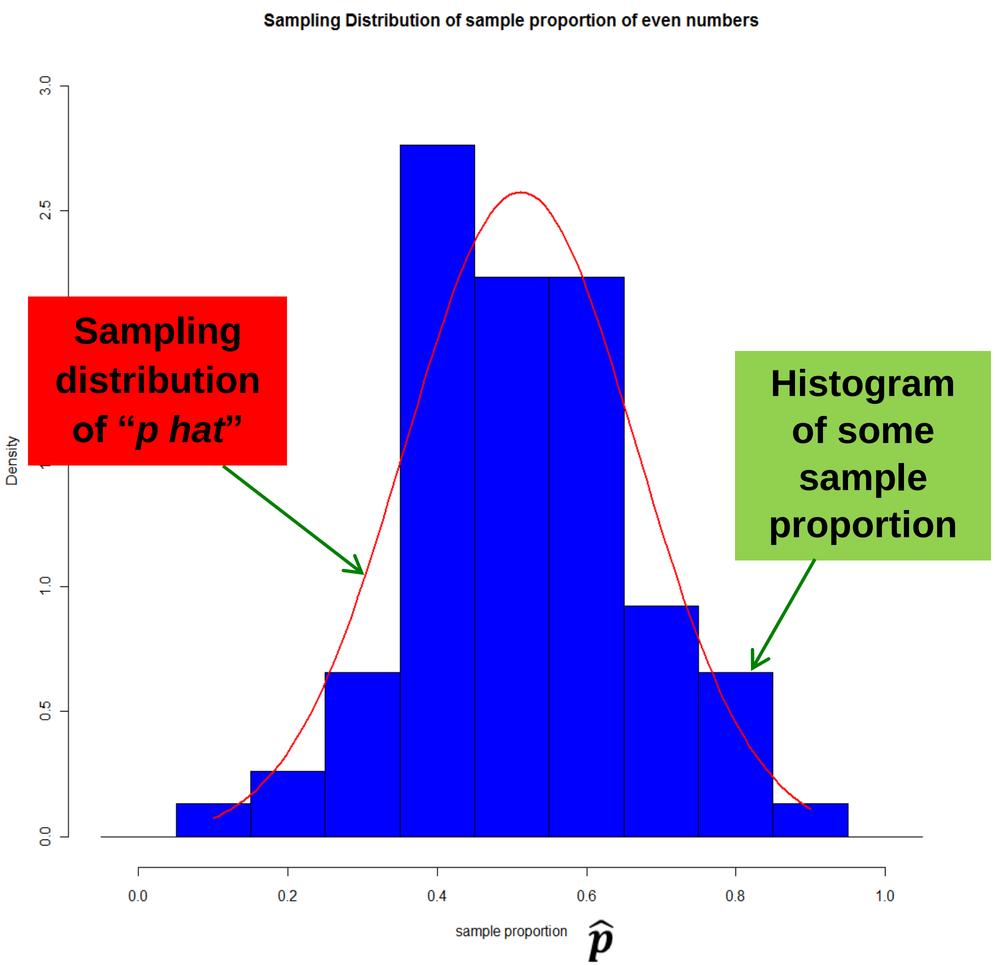
For all of your p-hats:
(1)0.1
(2)0.2 0.2
(5)0.3 0.3 0.3 0.3 0.3
(21)0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4
(17)0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5
(17)0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6
(7)0.7 0.7 0.7 0.7 0.7 0.7 0.7
(5)0.8 0.8 0.8 0.8 0.8
(1)0.9

Q: Draw a histogram with classes as: (for line 101-120 in Table B)

Class	(0, 0.1]	(0.1, 0.2]	(0.2, 0.3]	(0.3, 0.4]	(0.4, 0.5]	(0.5, 0.6]	(0.6, 0.7]	(0.7, 0.8]	(0.8, 0.9]
Counts									

Sampling Distribution of sample mean of 10 random digits

Class	(0, 0.1]	(0.1, 0.2]	(0.2, 0.3]	(0.3, 0.4]	(0.4, 0.5]	(0.5, 0.6]	(0.6, 0.7]	(0.7, 0.8]	(0.8, 0.9]
Counts	1	2	5	21	17	17	7	5	1

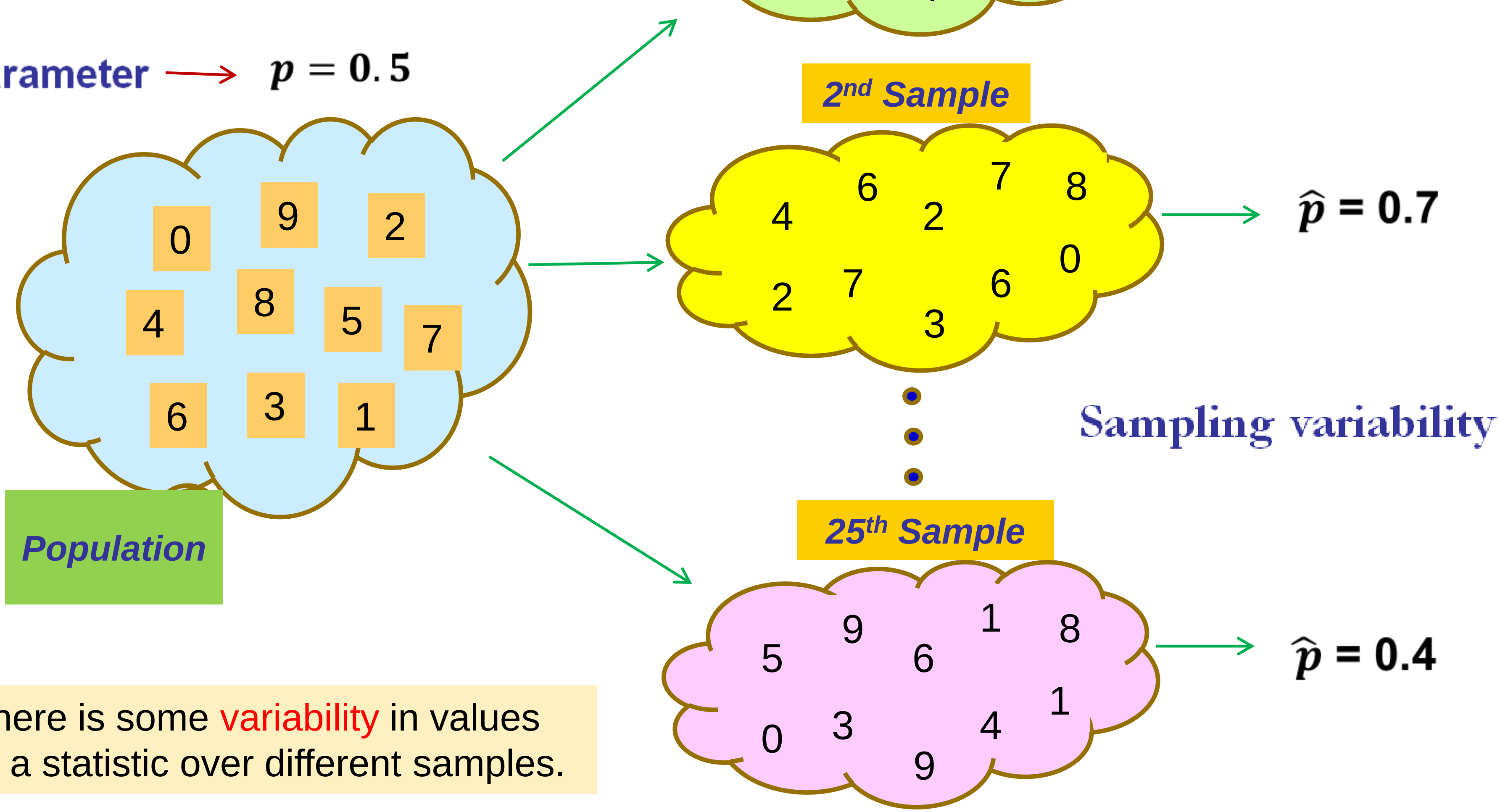


Q: Write a journal about how to get the ***sampling distribution of Sample proportion p-hat*** today, by answering the following questions:

1) How to obtain p-hat's from Table B for each student?
2) How many p-hat's did we have totally in the class?
3) How to make a histogram for p-hat? What is the name of the histogram?
4) What did the smooth curve represent?
5) For the smooth curve, what did the horizontal axis and vertical axis present?

Sampling Distribution

Select 10 random digits from Table B and find sample proportion of even #



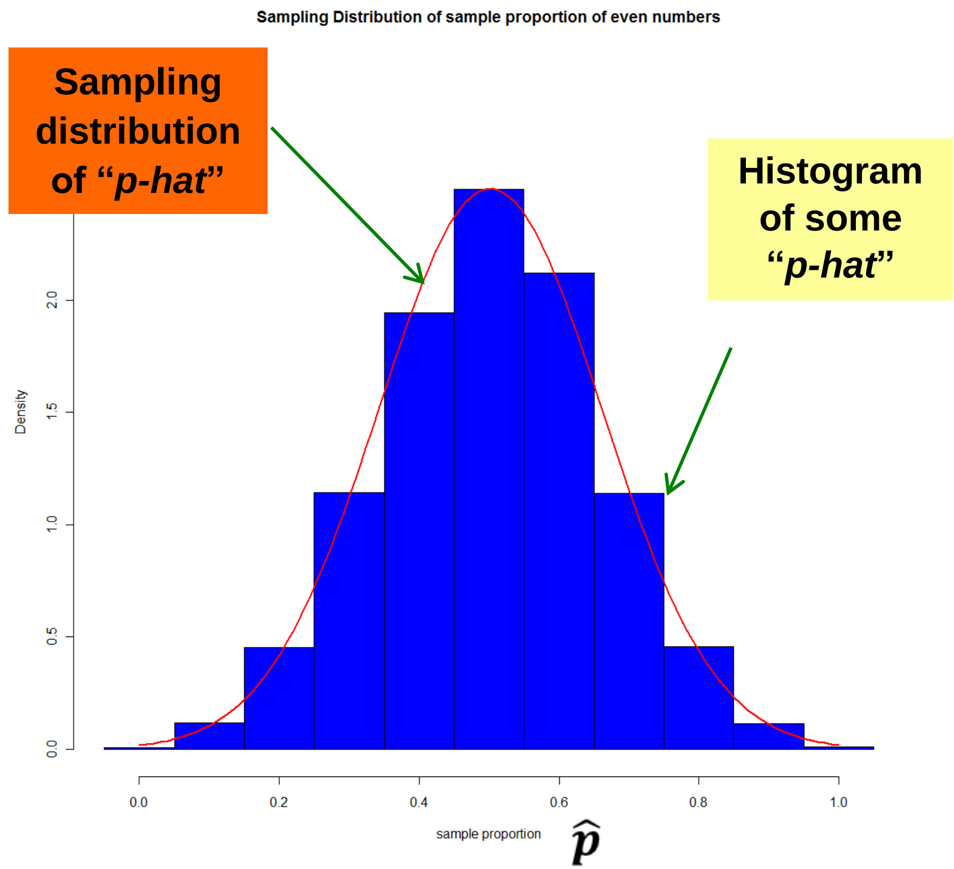
There is some variability in values of a statistic over different samples.

Sampling Distribution of sample proportion of even # of 10 random digits

(1) Select 10 random digits from Table B, and then take the sample proportion of even #.

(2) Repeat this process a lot of times, say 10,000 times.

(3) Make a histogram of these 10,000 sample mean's. The probability distribution looks like a Normal distribution.



The probability distribution of a statistic is called its **sampling distribution**.

Center of p-hat = 0.5018
SD of p-hat ≈ 0.1598
Note: n=10.

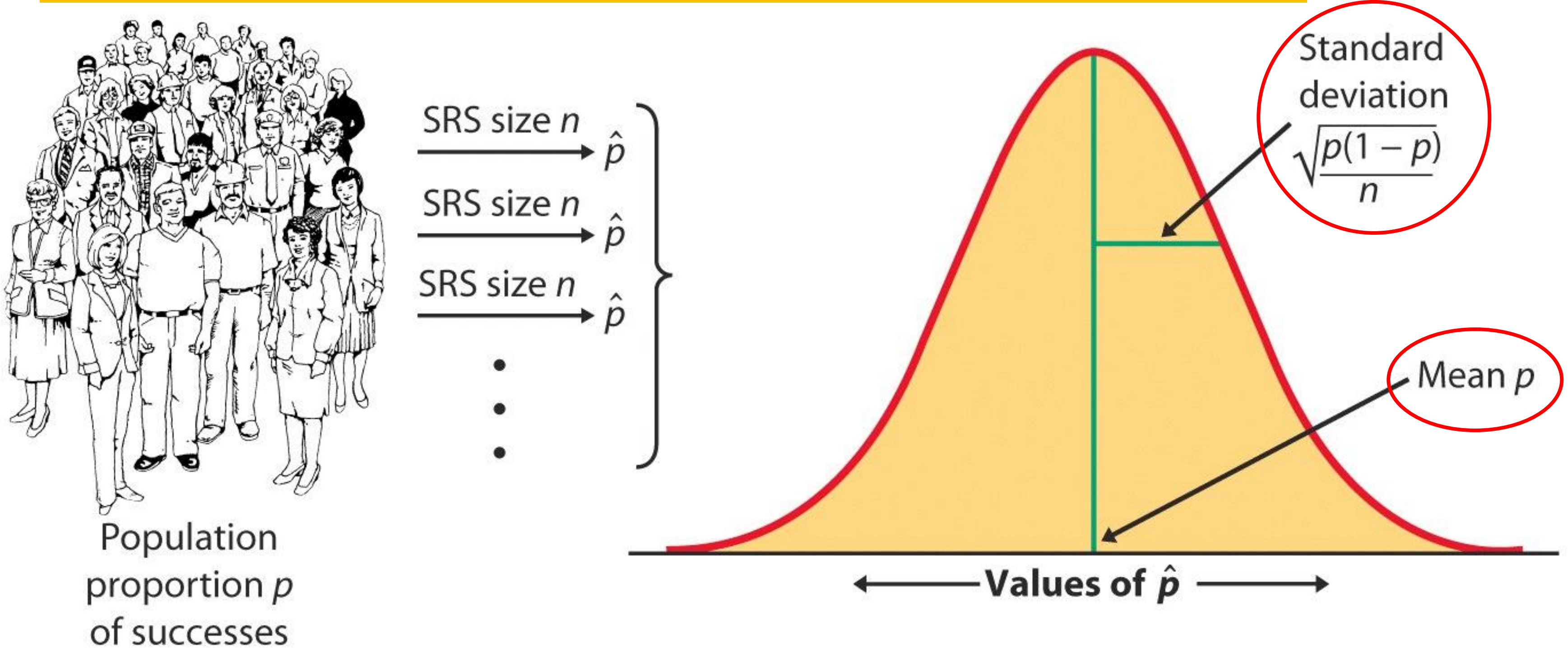
SD of p-hat = $\sqrt{\dfrac{p(1-p)}{n}}$

Sampling distribution of the sample proportion

The sampling distribution of $\hat{p}$ is **never exactly normal**. But as the sample size increases, the sampling distribution of $\hat{p}$ becomes approximately normal.

The normal approximation is most accurate for any fixed *n* when *p* is close to 0.5, and least accurate when *p* is near 0 or near 1.

When does the normality apply: ***np* ≥15 *and n*(1 - *p*) ≥15**



Sampling Distribution of $\hat{p}$

- If data are obtained from a SRS and ***np*≥15 and *n*(1-*p*)≥15**, then the sampling distribution of $\hat{p}$ has the following form:

- **For sample percentage:**
- $\hat{p}$ is approximately normal with mean p and

standard deviation: $\sqrt{\frac{p(1-p)}{n}}$

$$\text{Standardize: Z-score of } \hat{p} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Sampling distribution of a sample Proportion = distribution of $\hat{p}$

$\hat{p}$ follows Approximately N(p, $\sqrt{\frac{p(1-p)}{n}}$)

Standardize: Z-score of $\hat{p}$ = $\dfrac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$;

Reverse: $\hat{p} = p + \sqrt{\frac{p(1-p)}{n}} * Z$

Note: data are obtained from a SRS and **$np \geq 15$ and $n(1-p) \geq 15$.**

Example 1 (a)

Note: data are obtained from a SRS and $np \geq 15$ and $n(1-p) \geq 15$.

Maureen Webster, who is running for mayor in a large city, claims that she is favored by 53% of all eligible voters of that city. Assume that this claim is true. In a random sample of 400 registered voters taken from this city.

Find Population proportion p= _________.

a.) What is the sampling distribution of p-hat?

b) What is the probability of getting a sample proportion less than 49% in which will favor Maureen Webster?

c.) Find the probability of getting a sample proportion in between 50% and 55%.

d) Is it reasonable to assume Normal shape for this sampling distribution? Explain.

(b) Z=(0.49-0.53)/0.02495 = -1.60
Pr(Z<-1.60)
=normalcdf(-E99, -1.6, 0, 1)
= 0.0548

(c) Z=(0.5-0.53)/0.02495 = -1.20;
Z=(0.55-0.53)/0.02495 = 0.80;
Pr(-1.20 <Z<0.80)
=normalcdf(-1.20, 0.80, 0, 1)
=0.673

(d) Yes.
n*p=400*0.53=212, voting for this person.

n*(1-p) = 400*0.47=188, voting against for this person.

Example 1 (b)

Maureen Webster, who is running for mayor in a large city, claims that she is favored by 53% of all eligible voters of that city. Assume that this claim is true. **If instead we choose a random sample of 1,000 registered voters taken from this city.**

a.) What is the sampling distribution of p-hat?

b) What is the probability of getting a sample proportion less than 49% in which will favor Maureen Webster?

(b) Z=(0.49-0.53)/0.0157829 = -2.534388

Pr(Z< -2.534388) = 0.005703126

Example 2:

- A researcher studied the use of prenatal care among low-income American women. She found that only **51 percent** of these women had adequate prenatal care. Let's assume that the population of similar low-income American women, 51 percent had adequate prenatal care. If **200 women** from this population are drawn at random, what is the probability that **less than 45 percent** will have received adequate prenatal care?

$p = 0.51$	$n = 200$	$\hat{p} = 0.45$
$\mu_{\hat{p}} = p = 0.51$	$\sigma_{\hat{p}} = \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{(0.51)(0.49)}{200}} = 0.0353$	$z_{\hat{p}} = \frac{\hat{p} - p}{\sigma_{\hat{p}}} = \frac{0.45 - 0.51}{0.0353} = -1.7$
$P(\hat{p} < 0.45) = P(z_{\hat{p}} < -1.7)$		
$= \mathbf{0.0446}$		

Example 3:

- Researcher estimated that **64 percent** of U.S. adults ages 20-74 were overweight or obese (overweight: BMI 25-29, obese: BMI 30 or greater). Use this estimate as the population proportion for U.S. adults ages 20-74. If **125 subjects** are selected at random from the population, what is the probability that **70 percent or more** would be found to be overweight or obese?

$p = 0.64$	$n = 125$	$\hat{p} = 0.7$
$\mu_{\hat{p}} = p = 0.64$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ $= \sqrt{\frac{(0.64)(0.26)}{125}}$ $= 0.0428$	$z_{\hat{p}} = \frac{\hat{p} - p}{\sigma_{\hat{p}}}$ $= \frac{0.70 - 0.64}{0.0428}$ $= 1.40$
$P(\hat{p} \geq 0.7) = P(z_{\hat{p}} \geq 1.4)$		
$= 1 - P(z_{\hat{p}} < 1.4)$		
$= 1 - 0.9192 = \mathbf{0.0808}$		

More exercise

1. 30% of all autos undergoing an emissions inspection at a city fail in the inspection. Among 200 cars randomly selected in the city, the percentage of cars that fail in the inspection is around_____, with SD______. would it be unusual to have sample percentage 35%?

2. 60% of all residents in a big city are Democrats. Among 400 residents randomly selected in the city, would it be unusual to have sample percentage percentage<58%?

3. In airport luggage screening it is known that 3% of people have questionable objects in their luggage. For the next 1600 people, use normal approximation to find the prob that at least 4% of the people have questionable objects.

4. It is known that 60% of mice inoculated with a serum are protected from a certain disease. If 80 mice are inoculated, find the prob that at least 70% are protected from the disease.

Ans: 1. p=0.3, SD=.0324, $Z_{0.35}$=1.54,

2. p=0.6, sd=0.0245, $Z_{0.58}$=-0.82, ans=0.2061

3. p=0.03, sd=0.00426, $Z_{0.04}$=2.35, ans=0.0094

4. p=0.6, sd=0.0548, $Z_{0.7}$=1.82, ans=0.0344

Sampling Distribution of the Difference between Two Sample Proportions

Central Limit Theorem

Sampling Distribution of Sample Mean = Distribution of $\bar{x}$

$$\mu_{\bar{X}} = \mu$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

$\bar{X}$ follows Approximately $N(\mu, \frac{\sigma}{\sqrt{n}})$

Standardize: Z-score of $\bar{X} = \dfrac{\bar{X} - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)}$

Sampling Distribution of sample Proprtion = Distribution of $\hat{p}$

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

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

$\hat{p}$ follows Approximately $N(p, \sqrt{\frac{p(1-p)}{n}})$

Standardize: Z-score of $\hat{p} = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$;

Central Limit Theorem

Difference Between Two Sample Means

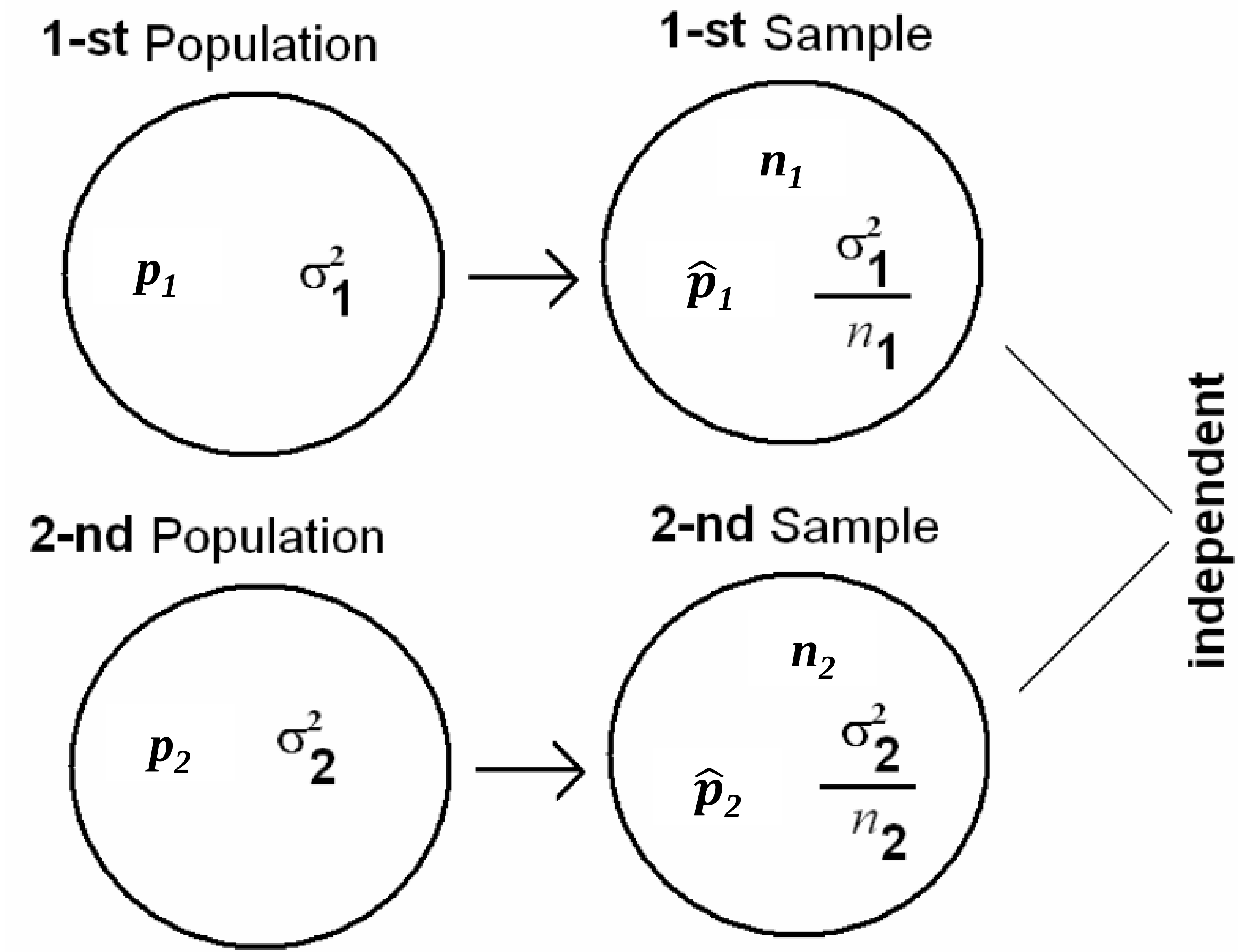
$$\mu_{\overline{X}_1-\overline{X}_2} = \mu_1 - \mu_2$$

$$\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\overline{X}_1 - \overline{X}_2 \sim \mathrm{N}\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

$$\Leftrightarrow$$

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathrm{N}(0,1)$$


1-st Population
p_1
σ_1^2
1-st Sample
n_1
$\hat{p}_1$
$\frac{\sigma_1^2}{n_1}$
2-nd Population
p_2
σ_2^2
2-nd Sample
n_2
$\hat{p}_2$
$\frac{\sigma_2^2}{n_2}$
independent

Theorem

If n_1 and n_2 are large, then the sampling distribution of $(\hat{p}_1 - \hat{p}_2)$ *is approximately normal with mean*

$$\boldsymbol{\mu_{(\hat{p}_1 - \hat{p}_2)} = p_1 - p_2}$$

And Variance

$$\boldsymbol{Var\,(\hat{p}_1 - \hat{p}_2) = \sigma^2_{(\hat{p}_1 - \hat{p}_2)} = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$\boldsymbol{SD\,(\bar{x}_1 - \bar{x}_2) = \sqrt{\sigma^2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

that is:

$$z_{\hat{p}_1-\hat{p}_2} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sigma_{\hat{p}_1-\hat{p}_2}}$$

$$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$z_{\hat{p}_1-\hat{p}_2} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

Example 4:

- Researchers found that among U.S. adults ages 75 or older, **34 percent** had lost all their natural teeth and for U.S. adults ages 65-74, **26 percent** had lost all their natural teeth. Assume these properties are parameters for the United States in those age groups. If a random sample of **200 adults** ages 65-74 and an independent random sample of **250 adults** ages 75 or older are drawn from these populations, find the probability that **the difference in percent of total natural teeth loss is less than 5 percent between the two population.**

Solution

$$z_{\hat{p}_1-\hat{p}_2} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sigma_{\hat{p}_1-\hat{p}_2}}$$

$$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$p_1 = .34$	$n_1 = 250$	$\hat{p}_1 - \hat{p}_2 = 0.05$
$p_2 = .26$	$n_2 = 200$	
$\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2$ $= .34 - 26$ $= 0.08$	$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$ $= \sqrt{\frac{(0.34)(0.66)}{250} + \frac{(0.26)(0.74)}{200}}$ $= 0.0431$	$z_{\hat{p}_1-\hat{p}_2} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sigma_{\hat{p}_1-\hat{p}_2}}$ $= \frac{0.05 - 0.08}{0.0431} = -0.7$
$P((\hat{p}_1 - \hat{p}_2) < 0.05) = P(z_{\hat{p}} < -0.7)$		
$= \mathbf{0.2420}$		