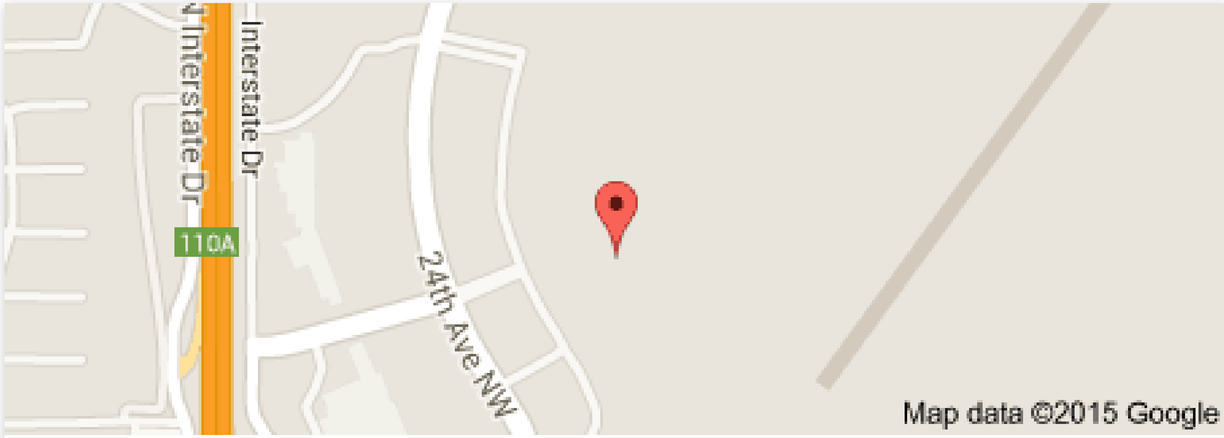


Crest Foods

Directions

Grocery Store

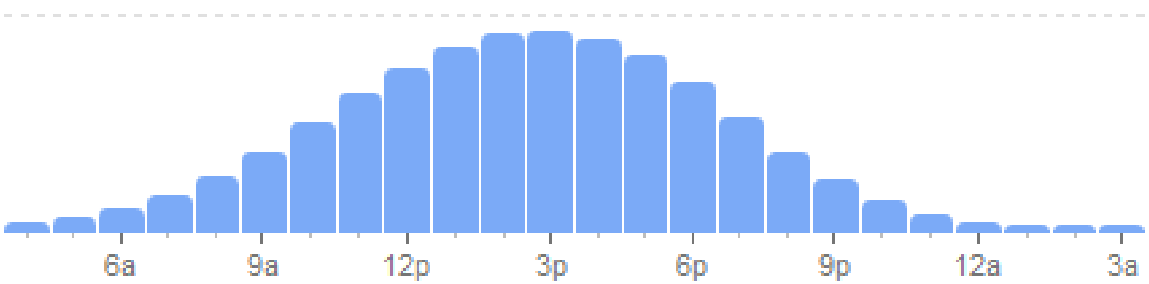
Address: 2550 Mt Williams Dr, Norman, OK 73069

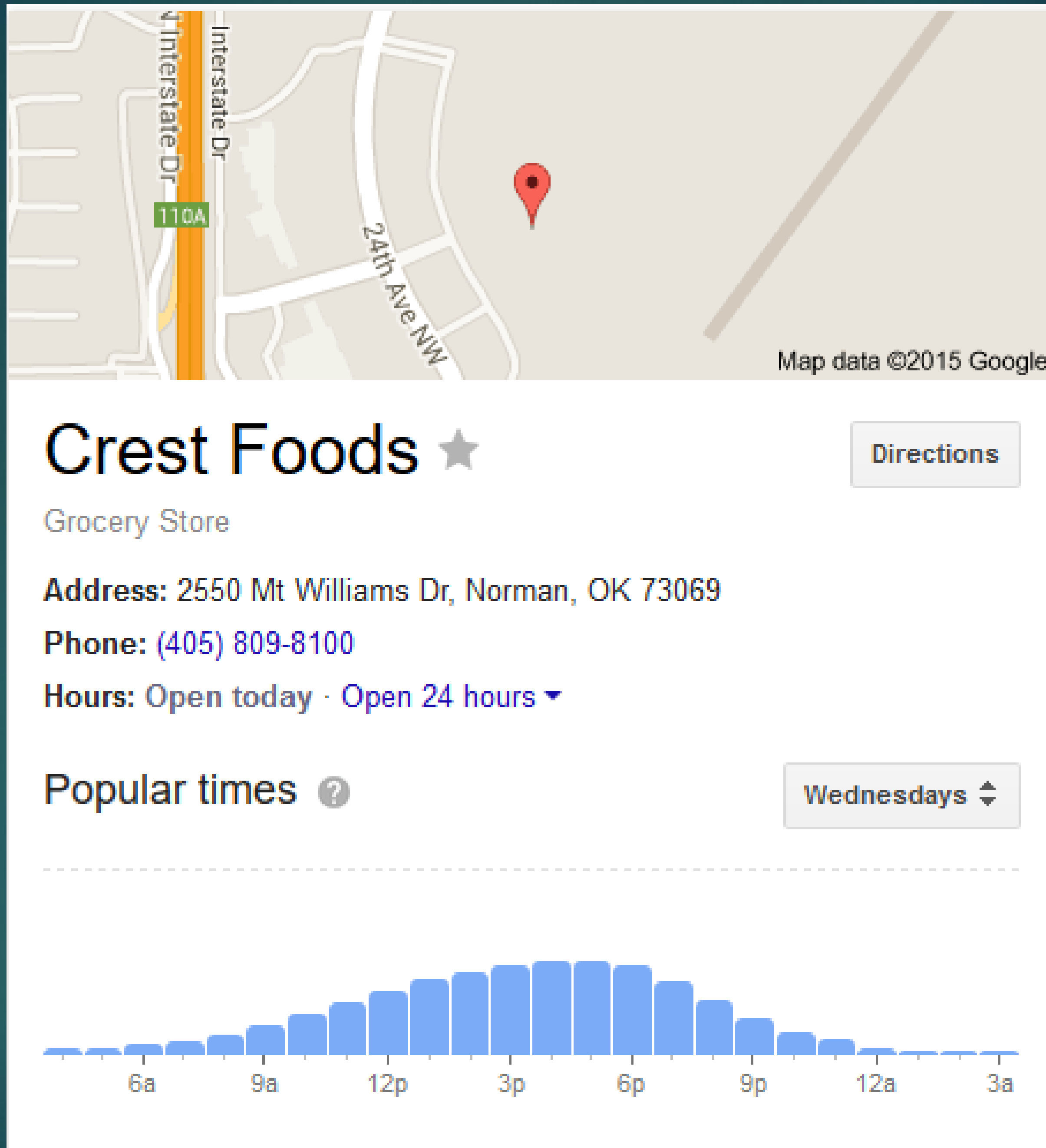
Phone: (405) 809-8100

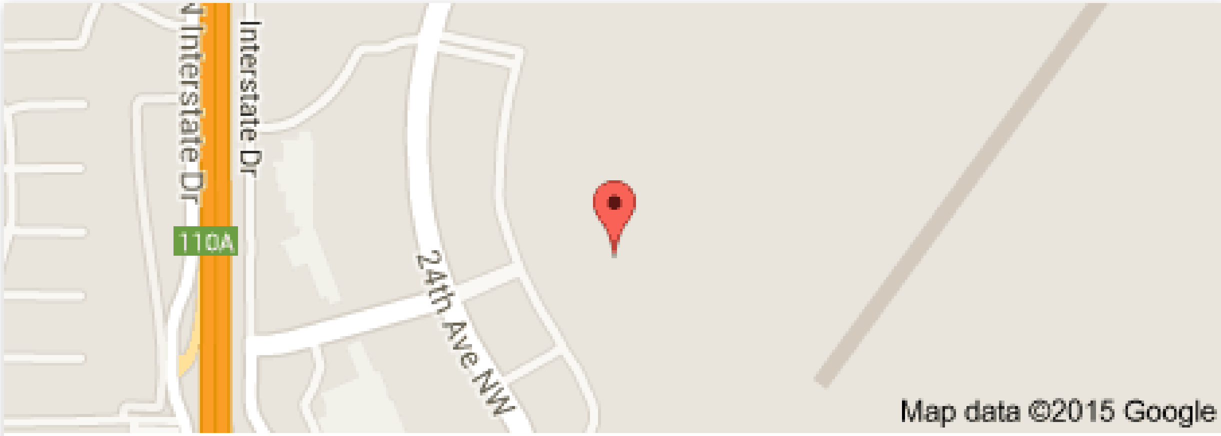
Hours: Open today · Open 24 hours

Popular times

Sundays




Interstate Dr
Interstate Dr
110A
24th Ave NW
Map data ©2015 Google
Crest Foods
Directions
Grocery Store
Address: 2550 Mt Williams Dr, Norman, OK 73069
Phone: (405) 809-8100
Hours: Open today · Open 24 hours
Popular times
Wednesdays
6a
9a
12p
3p
6p
9p
12a
3a



Crest Foods ★

Directions

Grocery Store

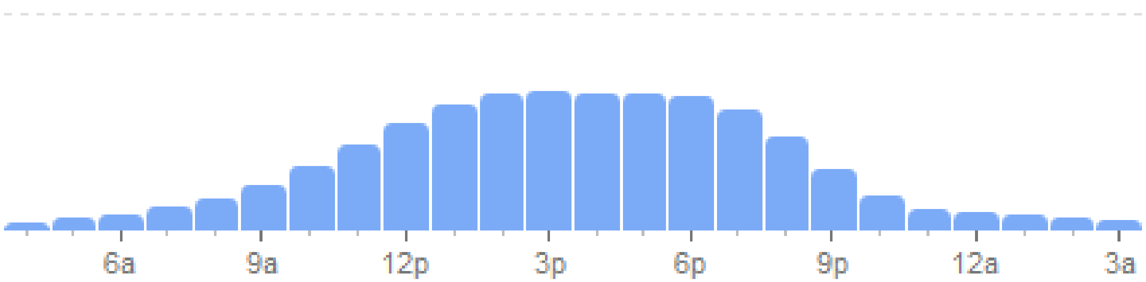
Address: 2550 Mt Williams Dr, Norman, OK 73069

Phone: (405) 809-8100

Hours: Open today · Open 24 hours

Popular times

Fridays

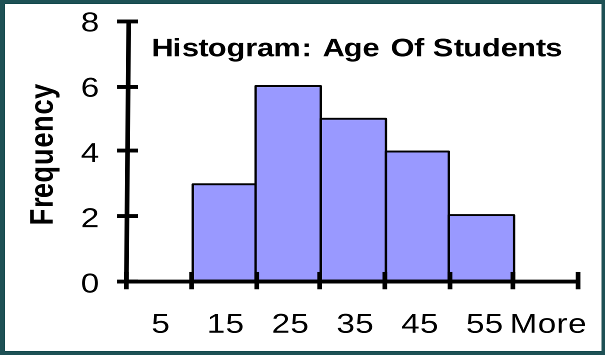


Chapter 7

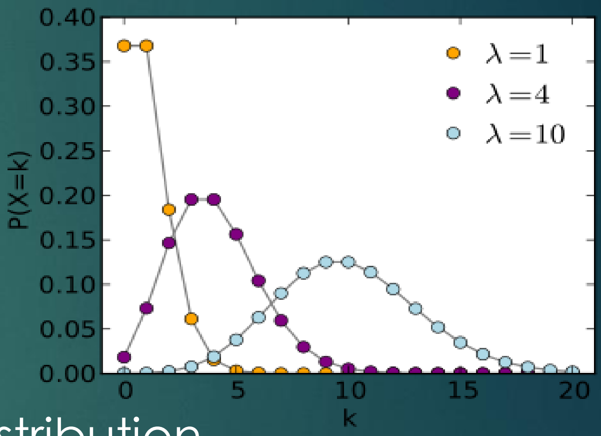
Sampling Distributions

Review: Distributions

- (Relative) Frequency and (Relative) Cumulative distribution of observed data.



- Probability Distribution
  - Discrete; e.g., Binomial, Poisson, and Hypergeometric Probability Distributions
  - Continuous: e.g., (Standardized) Normal Distribution



Now, let's talk about Oddland a bit.

Using sample mean to obtain the population mean.

- We are going to choose a stratified sample of the population in Oddland.
- We plan to ask them a series of questions about their gender, age, and their partying habits.
- Using this sample, we can obtain the mean of, say, age for this sample.

Using sample mean to obtain the population mean.

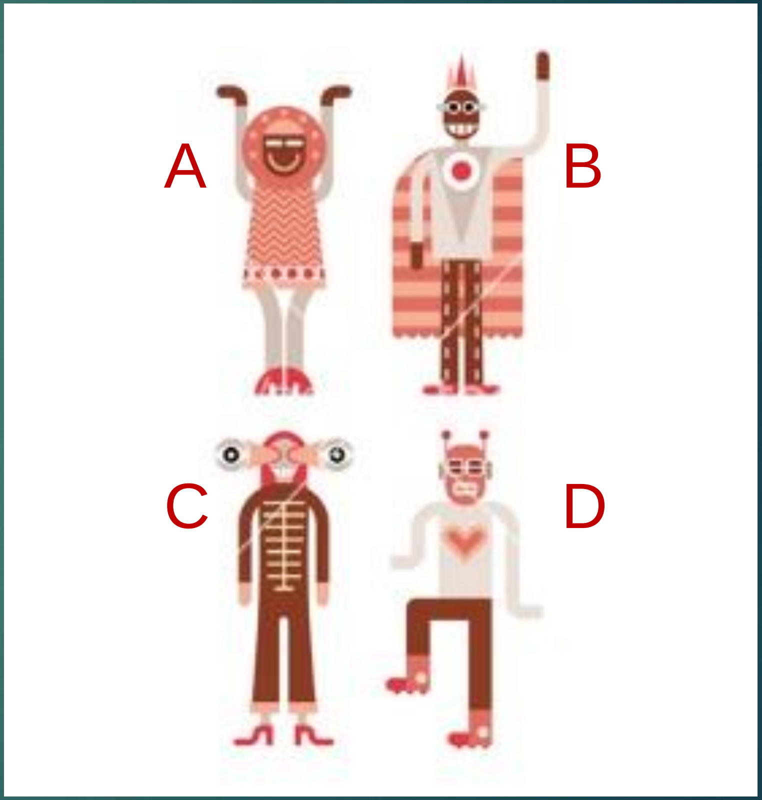
- We are going to choose a stratified sample of the population in Oddland.
- We plan to ask them a series of questions about their gender, age, and their partying habits.
- Using this sample, we can obtain the mean of, say, age for this sample.
- Can we use the computed mean, and say something about the population?
- Here is an important question:

If we choose another stratified sample of individuals, are we able to recover the same mean for the age, as we found in our first sample?

The mean age in Oddland

- **There are four people in Oddland.**
- Thus, population size **N=4**
- Random variable, X,
  is the age of individuals
- Values of X:

  18, **20**, **22**, and **24** (years)



The mean age in Oddland

Summary Measures for the Population Distribution:

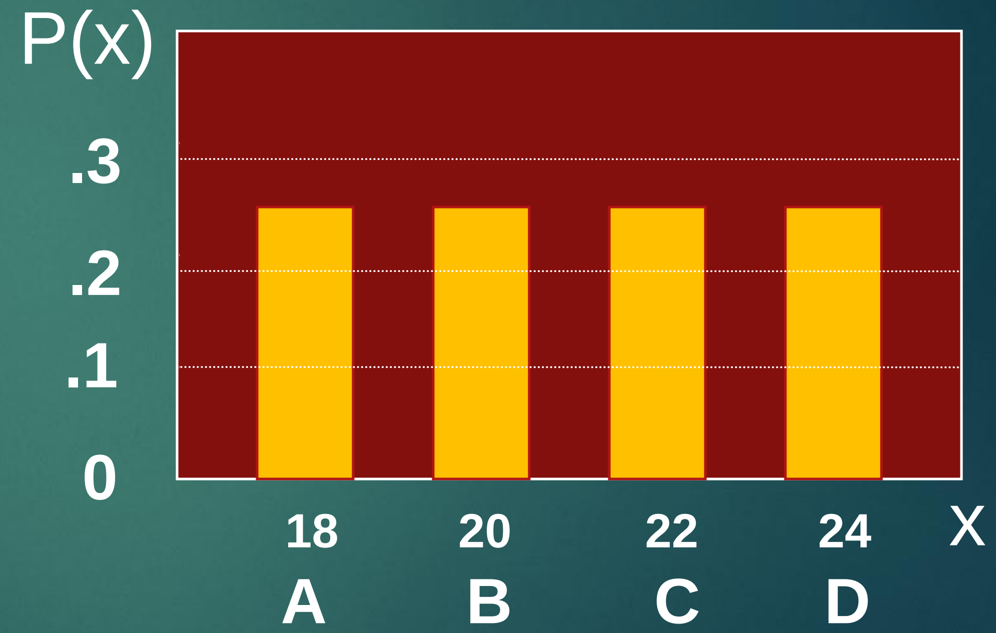
Mean

$$\mu = \frac{\sum X_i}{N}$$

$$= \frac{18+20+22+24}{4} = 21$$

SD

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}} = 2.236$$



Uniform Distribution

The mean age in Oddland

Now consider all possible samples of size **n=2.**

1st Obs	2nd Observation			
	18	**20**	**22**	**24**
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

With replacement, we end up with **16** samples.

The mean age in Oddland

1st Obs	2nd Observation			
	18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

Computing the Mean for Each Sample

1st Obs	2nd Observation			
	18	20	22	24
18	18			
20				
22				
24				

The mean age in Oddland

1st Obs	2nd Observation			
	18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

Computing the Mean for Each Sample

1st Obs	2nd Observation			
	18	20	22	24
18	18	19		
20				
22				
24				

The mean age in Oddland

1st Obs	2nd Observation			
	18	**20**	**22**	**24**
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

Computing the Mean for Each Sample

1st Obs	2nd Observation			
	18	**20**	**22**	**24**
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

16 Sample Means

Sampling Distribution

- From a population of 4 individuals, we obtained 16 different samples of size 2 (with replacement), and the sample means varied greatly.
- We were planning to use a **SAMPLE STATISTIC** (e.g., sample mean) to estimate the **POPULATION PARAMETER** (e.g., population mean). But we ended up with so many different sample means.
- We are just hopeless…

Sampling Distribution

- From a population of 4 individuals, we obtained 16 different samples of size 2 (with replacement), and the sample means varied greatly.
- We were planning to use a **SAMPLE STATISTIC** (e.g., sample mean) to estimate the **POPULATION PARAMETER** (e.g., population mean). But we ended up with so many different sample means.
- We are just hopeless…
- Or, are we?!
- Well, we can make use of the distribution of sample statistics. Let's call it **SAMPLING DISTRIBUTION OF THE MEAN**.

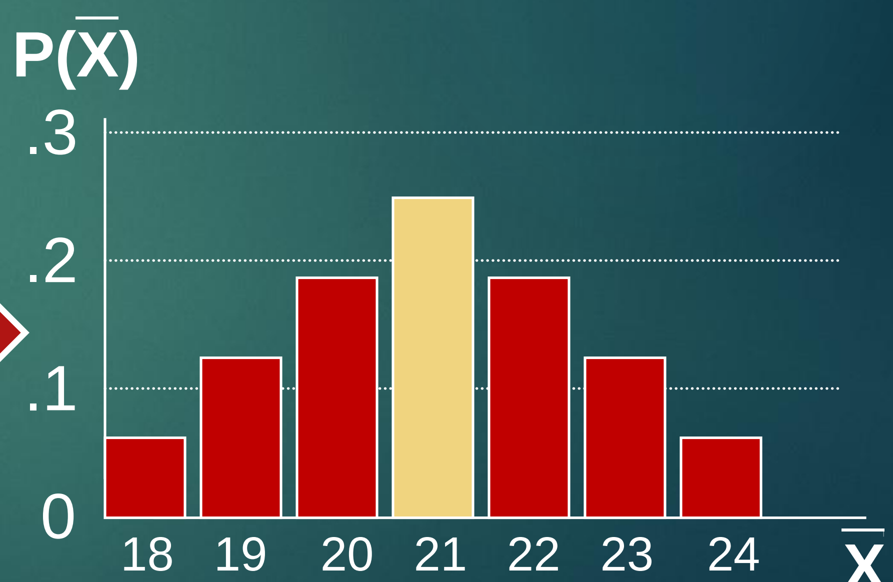
Developing a Sampling Distribution

Sampling Distribution of All Sample Means

16 Sample Means

1st Obs	2nd Observation			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Sample Means Distribution



No longer a uniform distribution

Developing a Sampling Distribution

Summary Measures of this Sampling Distribution:

$$\mu_{\bar{X}} = \frac{18 + 19 + 19 + \cdots + 24}{16} = 21$$

$$\sigma_{\bar{X}} = \sqrt{\frac{(18-21)^2 + (19-21)^2 + \cdots + (24-21)^2}{16}} = 1.58$$

Note: Here we divide by 16 because there are 16 different samples of size 2.

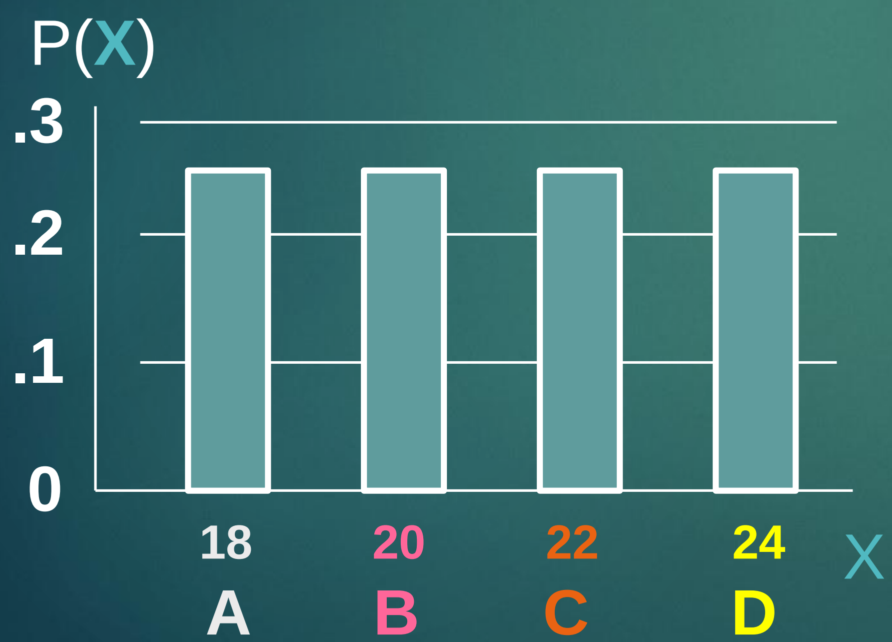
Comparing the Population Distribution to the Sample Means Distribution

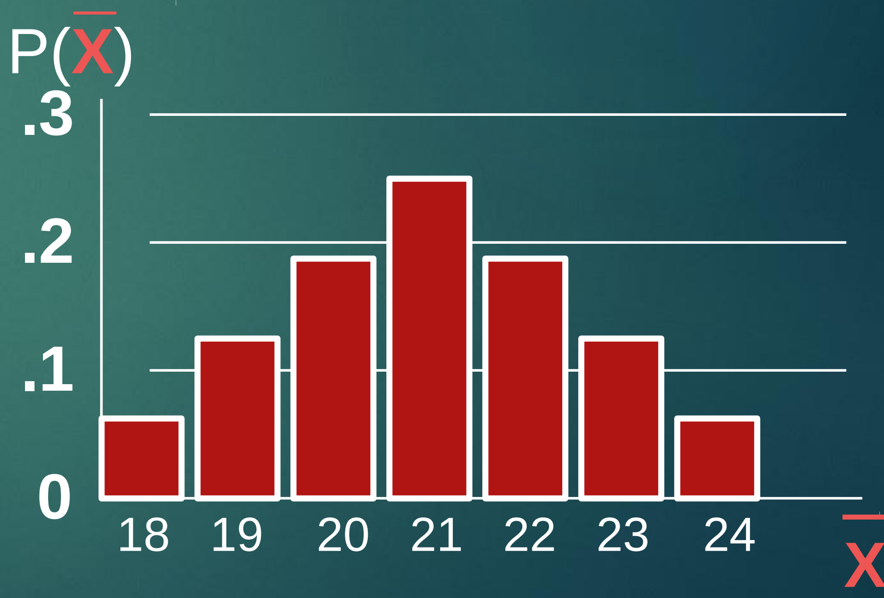
Population
N = 4

$\mu = 21 \quad \sigma = 2.236$

Sample Means Distribution
n = 2

$\mu_{\bar{X}} = 21 \quad \sigma_{\bar{X}} = 1.58$





Sampling Distribution of the Mean: Standard Deviation

- Different samples of the same size from the same population will yield different sample means
- A measure of the variability in the mean from sample to sample is given by the **Standard Error of the Mean**

Sampling Distribution of the Mean when Population is Distributed Normally*

* Yes, even in Oddland the population may be distributed normally!

Sampling Distribution of the Mean When the Population is *Normal*

- If a population is normal with mean **μ** and standard deviation **σ**, the sampling distribution of the mean is also normally distributed with:

$$\mu_{\bar{X}} = \mu \quad \textbf{AND} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

- Since the sample size appears in the denominator, the standard error of the mean decreases as the sample size increases

The Z-value for Sampling Distribution of the Mean:

- Z-value for the sampling distribution of $\bar{X}$:

$$Z = \frac{(\bar{X} - \mu_{\bar{X}})}{\sigma_{\bar{X}}} = \frac{(\bar{X} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

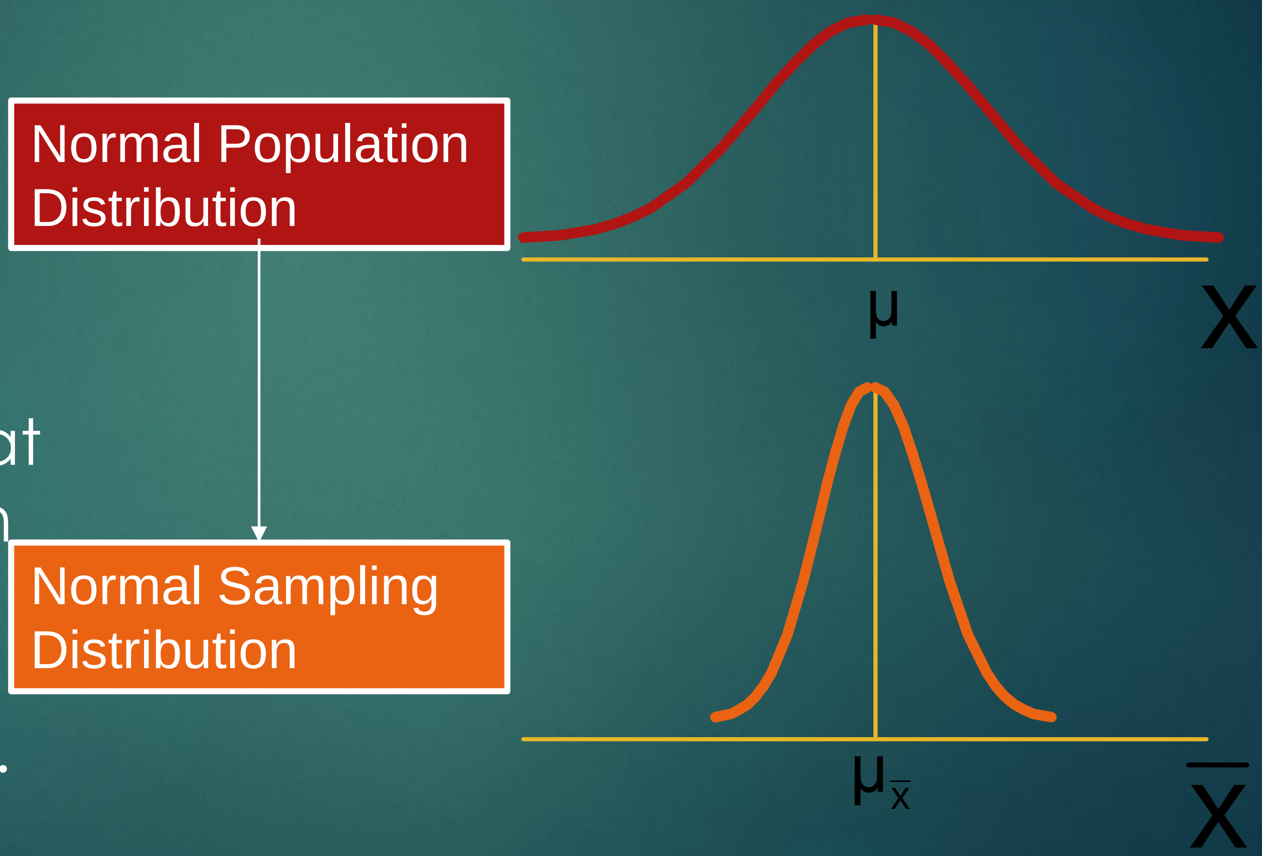
where:
$\bar{X}$ = sample mean
μ = population mean
σ = population standard deviation
n = sample size

- Using the Sample Mean along with the Population Mean and Standard Deviation, we are able to compute the Z-value for each Sample Mean.

Sampling Distribution Properties

The sample mean is an **unbiased estimator** of population mean when population is normally distributed

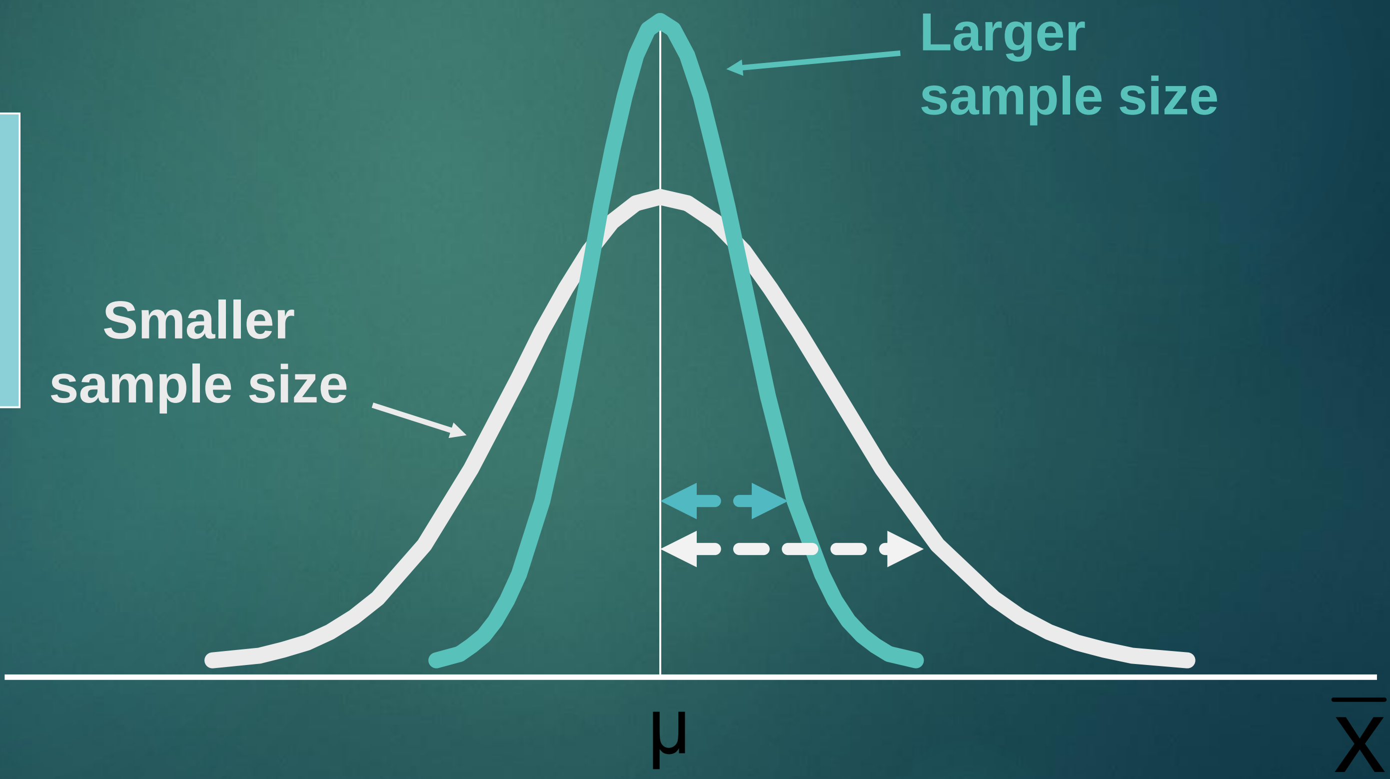
Since the Sample Mean (Statistic) is equal to Population Mean (Parameter), we can safely say that the Sample Mean is an unbiased estimator of Population Mean.



Sampling Distribution Properties

With an increase in sample size, we are able to obtain a better estimation of population mean since the standard deviation of the sample mean distribution decreases.

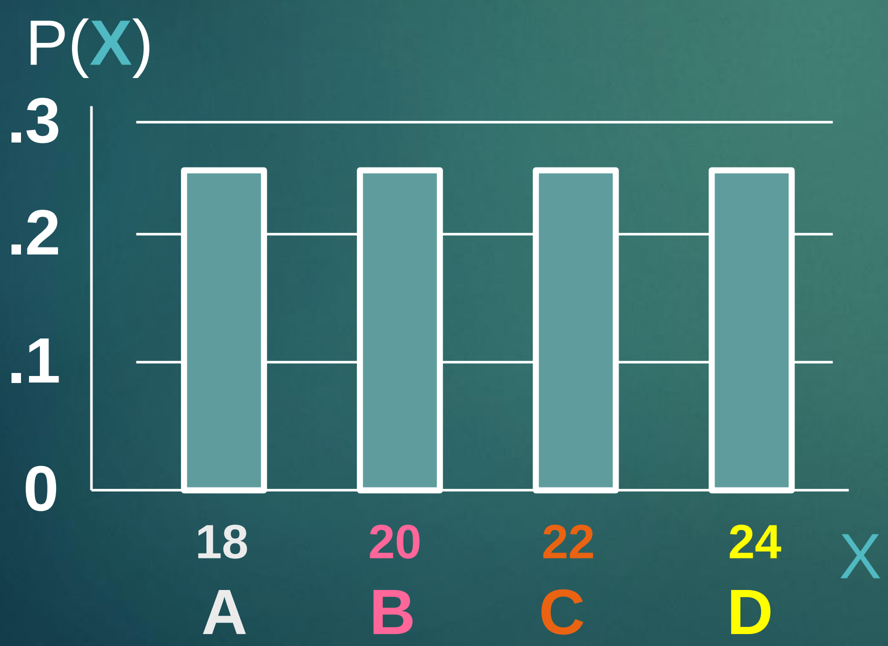
$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$



Back to the Oddland example:

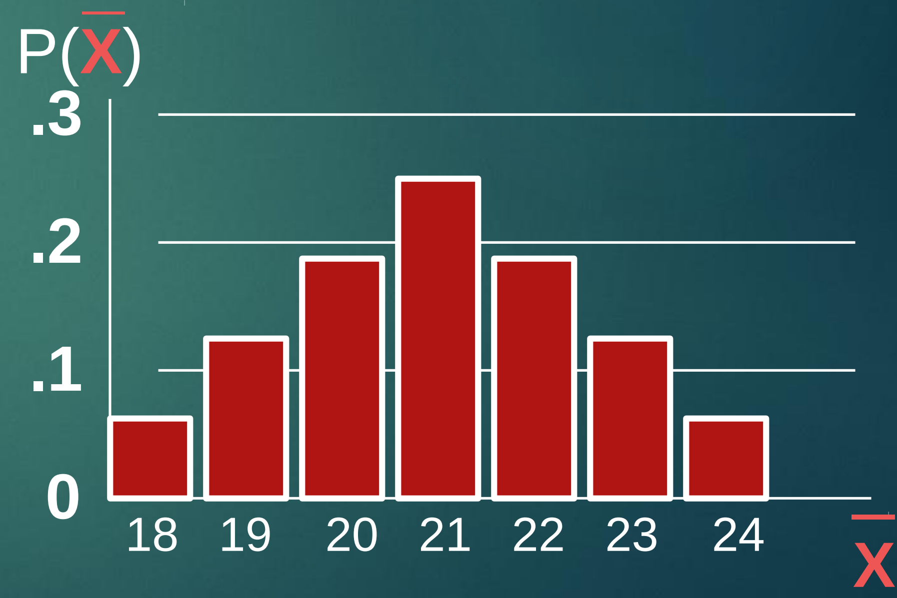
Population
N = 4

$\mu = 21 \qquad \sigma = 2.236$



Sample Means Distribution
n = 2

$\mu_{\bar{X}} = 21 \qquad \sigma_{\bar{X}} = 1.58$



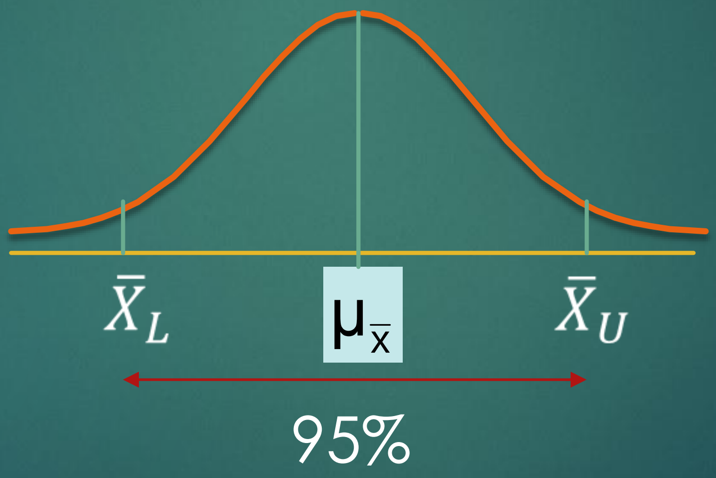
Let's Determine An Interval Including A Fixed Proportion of the Sample Means

- **μ** is the population mean, and **σ** is the population standard deviation.
- We happen to know their values: $\mu = 368$, $\sigma = 15$

Find a symmetrically distributed interval around **μ** that will include 95% of the sample means when sample sizes are equal to 25: i.e., $n = 25$.

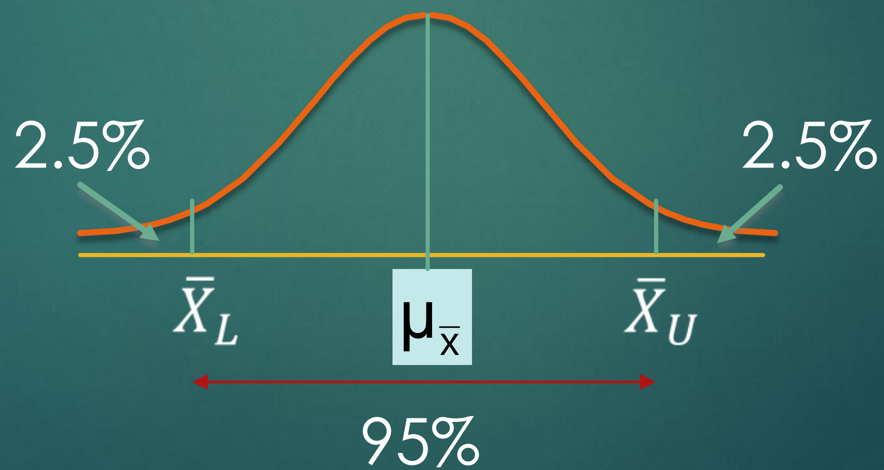
Let's Determine An Interval Including A Fixed Proportion of the Sample Means

- Since the interval contains 95% of the sample means 5% of the sample means will be outside the interval



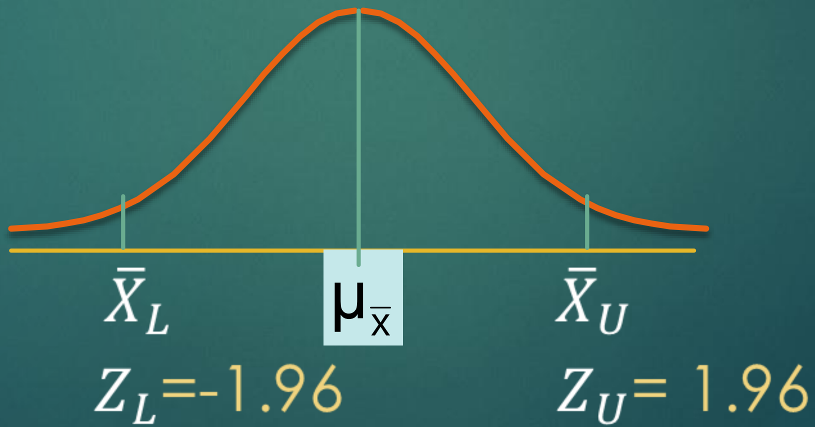
Let's Determine An Interval Including A Fixed Proportion of the Sample Means

- Since the interval contains 95% of the sample means 5% of the sample means will be outside the interval
- Since the interval is symmetric 2.5% will be above the upper limit and 2.5% will be below the lower limit.



Let's Determine An Interval Including A Fixed Proportion of the Sample Means

- Since the interval contains 95% of the sample means 5% of the sample means will be outside the interval
- Since the interval is symmetric 2.5% will be above the upper limit and 2.5% will be below the lower limit.
- Using Excel, the Z score with probability 2.5% (i.e., P=0.0250) is -1.96 (*=NORM.S.INV(0.025)*) and the Z score with probability 97.5% (0.9750) is 1.96 (*=NORM.S.INV(0.975)*). Thus, Z_L=-1.96 and Z_U=1.96



Let's Determine An Interval Including A Fixed Proportion of the Sample Means

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \rightarrow \bar{X} = \mu + Z\frac{\sigma}{\sqrt{n}}$$

Let's Determine An Interval Including A Fixed Proportion of the Sample Means

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \rightarrow \bar{X} = \mu + Z\frac{\sigma}{\sqrt{n}}$$

- Calculating the lower limit of the interval

$$\bar{X}_L = \mu + Z_L \frac{\sigma}{\sqrt{n}} = 368 + (-1.96)\frac{15}{\sqrt{25}} = 362.12$$

- Calculating the upper limit of the interval

$$\bar{X}_U = \mu + Z_U \frac{\sigma}{\sqrt{n}} = 368 + (1.96)\frac{15}{\sqrt{25}} = 373.88$$

- **95% of all sample means of sample size 25 are between 362.12 and 373.88**

Sampling Distribution of the Mean when Population is **not** Distributed Normally*

* Well, it is Oddland after all.

Sampling Distribution of the Mean When the Population is **not** Normal

- We can apply the **Central Limit Theorem**:
  - Even if the population is not normal,
  - …sampling distribution of the means will be **approximately normal** as long as the sample size is *large enough*.

Properties of the sampling distribution:

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

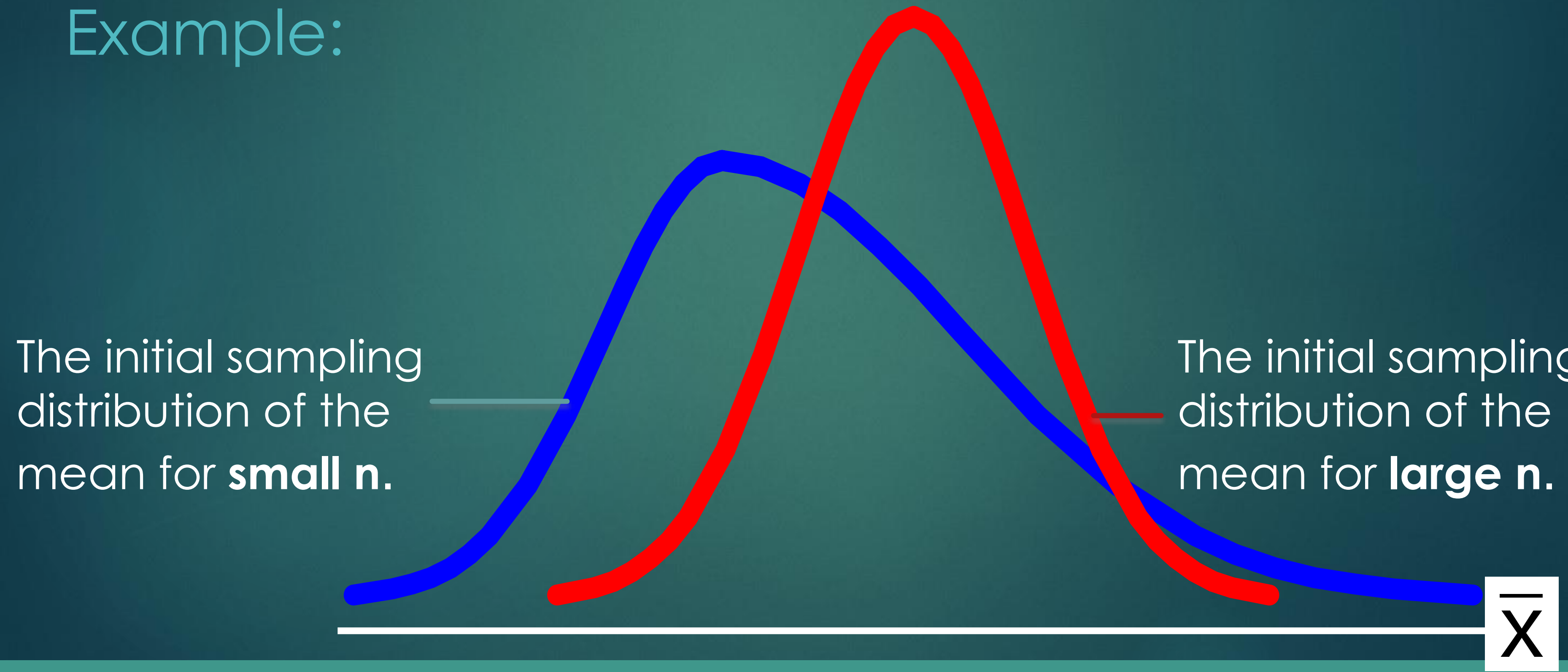
and

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

Central Limit Theorem

Thus, regardless of the population distribution, the sampling distribution of the mean becomes almost normal as we increase the sample size.

Example:



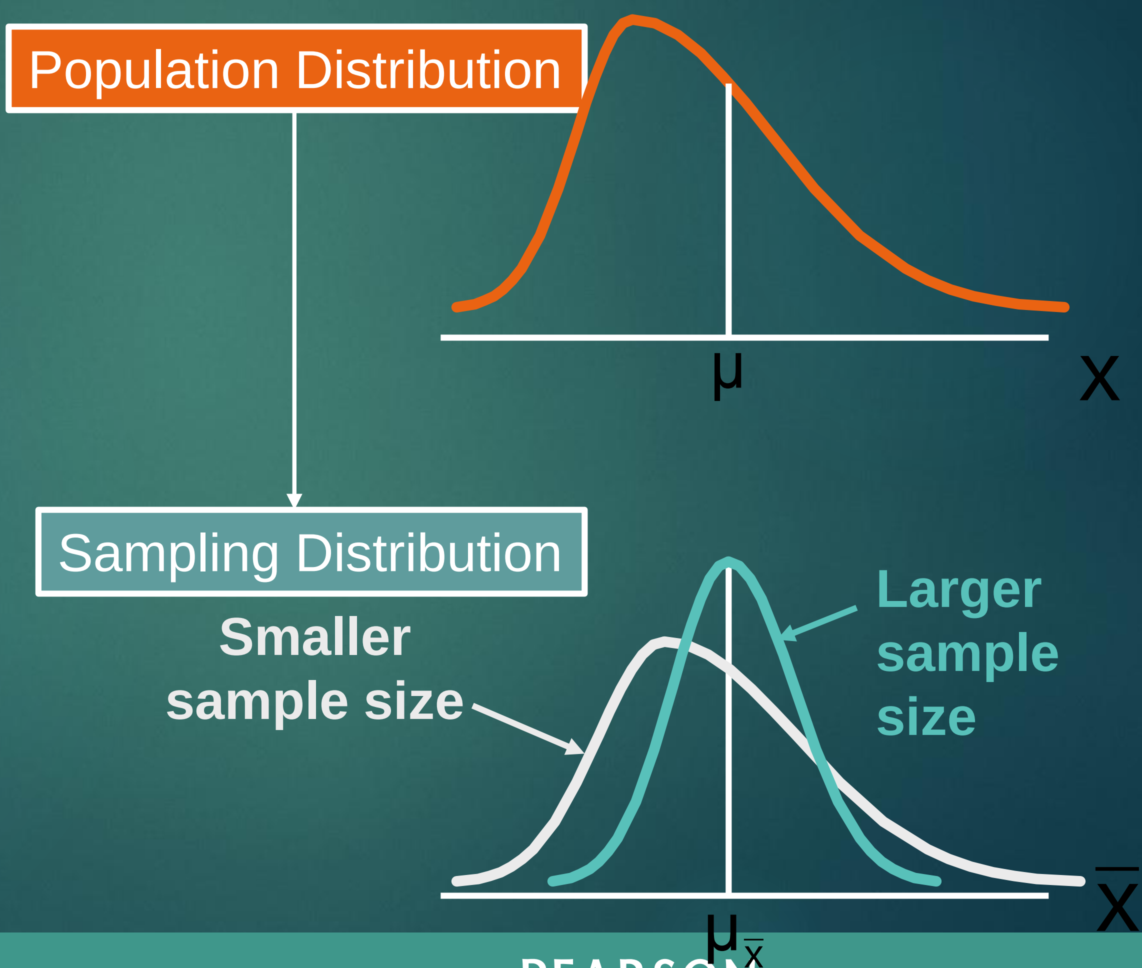
Sampling Distribution of the Mean When the Population is **not** Normal

Central Tendency

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

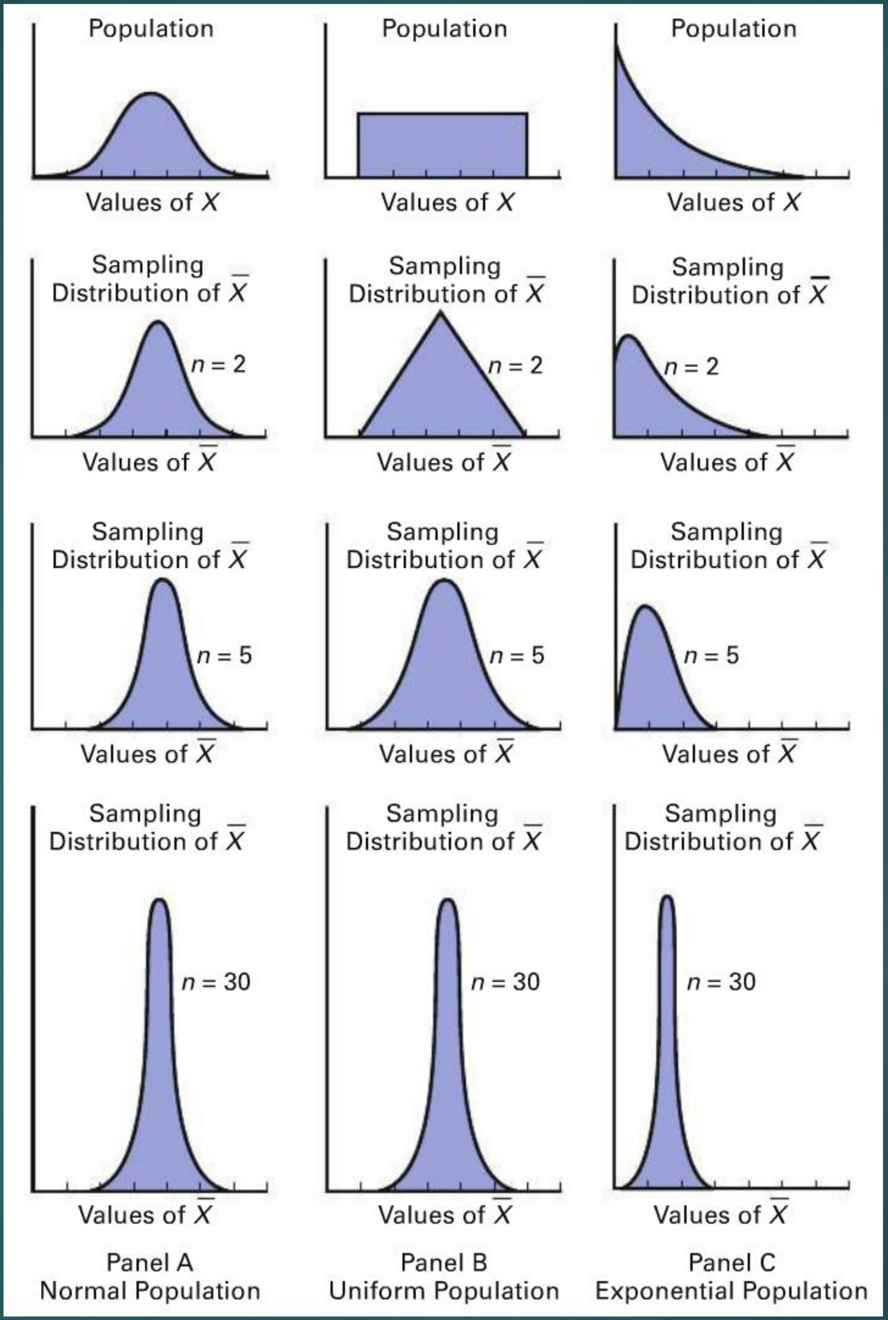
Variation

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



How Large is *Large Enough*?

- For most distributions, **n > 30** will give a sampling distribution that is nearly normal
- For fairly symmetric distributions, **n > 15**
- For a normal population distribution, the sampling distribution of the mean is always normally distributed


Population
Values of X
Population
Values of X
Population
Values of X
Sampling Distribution of X̄
n = 2
Values of X̄
Sampling Distribution of X̄
n = 2
Values of X̄
Sampling Distribution of X̄
n = 2
Values of X̄
Sampling Distribution of X̄
n = 5
Values of X̄
Sampling Distribution of X̄
n = 5
Values of X̄
Sampling Distribution of X̄
n = 5
Values of X̄
Sampling Distribution of X̄
n = 30
Values of X̄
Sampling Distribution of X̄
n = 30
Values of X̄
Sampling Distribution of X̄
n = 30
Values of X̄
Panel A
Normal Population
Panel B
Uniform Population
Panel C
Exponential Population

Example

- Suppose a population has mean $\mu = 8$ and standard deviation $\sigma = 3$. Suppose a random sample of size $n = 36$ is selected.

Example

- Suppose a population has mean $\mu = 8$ and standard deviation $\sigma = 3$. Suppose a random sample of size $n = 36$ is selected.
- What is the probability that the sample mean is between 7.8 and 8.2?

Example

- Suppose a population has mean $\mu = 8$ and standard deviation $\sigma = 3$. Suppose a random sample of size $n = 36$ is selected.
- What is the probability that the sample mean is between 7.8 and 8.2?
- **n>30** → Sampling Distribution of the Mean is normal

Example

- Suppose a population has mean μ = 8 and standard deviation σ = 3. Suppose a random sample of size n = 36 is selected.
- What is the probability that the sample mean is between 7.8 and 8.2?
- **n>30** → Sampling Distribution of the Mean is normal
- The mean of the sampling distribution of the mean is equal to the population mean: $\mu_{\bar{X}} = 8$
- Its standard deviation is: $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = 0.5$

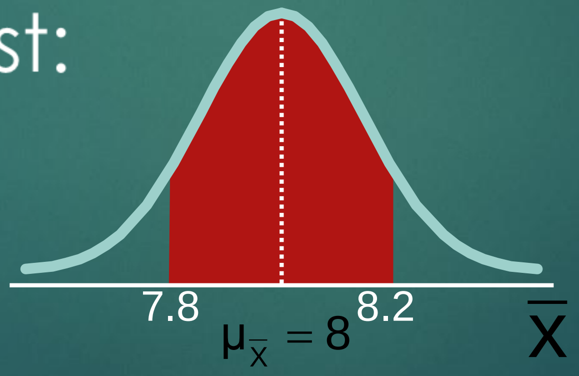
Example

- Population Parameters: $\mu = 8$ and $\sigma = 3$.
- Sample size $n = 36$.
- The mean of the sampling distribution of the mean: $\mu_{\bar{X}} = 8$
- The standard deviation of the sampling distribution of the mean: $\sigma_{\bar{X}} = 0.5$
- What is the probability that the sample mean is between **7.8** and **8.2**?

$$P(7.8 < \bar{X} < 8.2) = ?$$

Example

- $$P(7.8 < \bar{X} < 8.2) = ?$$
- Since $\bar{X}$ is normally distributed, finding this probability is quite easy for us:
- **Step 1**: Draw the distribution to identify the area of interest:



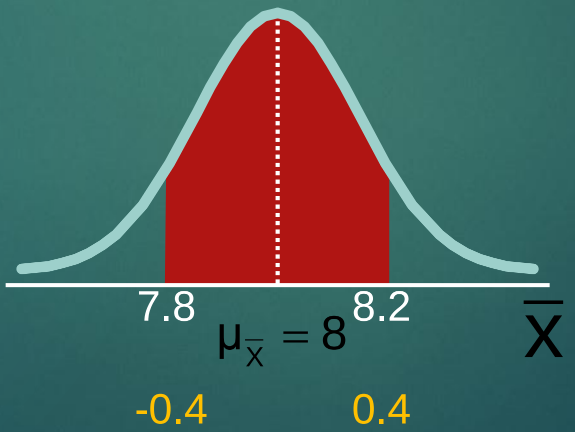
Example

$$P(7.8 < \bar{X} < 8.2) = ?$$

- **Step 2:** Translate $\bar{X}$ to Z

$$Z_L = \frac{7.8 - 8}{0.5} = -0.4 \; and \; Z_U = \frac{8.2 - 8}{0.5} = 0.4$$

$$P(-0.4 < Z < 0.4) = ?$$

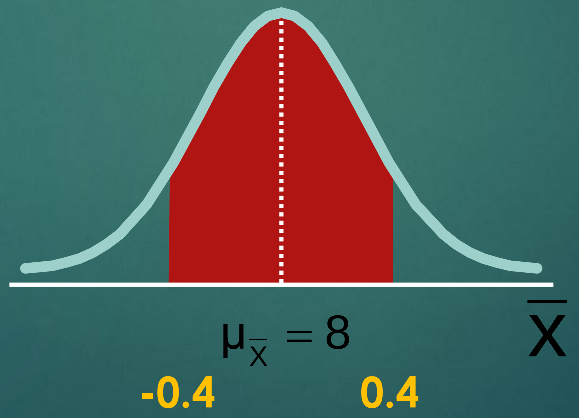


Example

$$P(7.8 < \bar{X} < 8.2) = ?$$

- **Step 2:** Compute the probabilities using the Standardized Normal Distribution Table

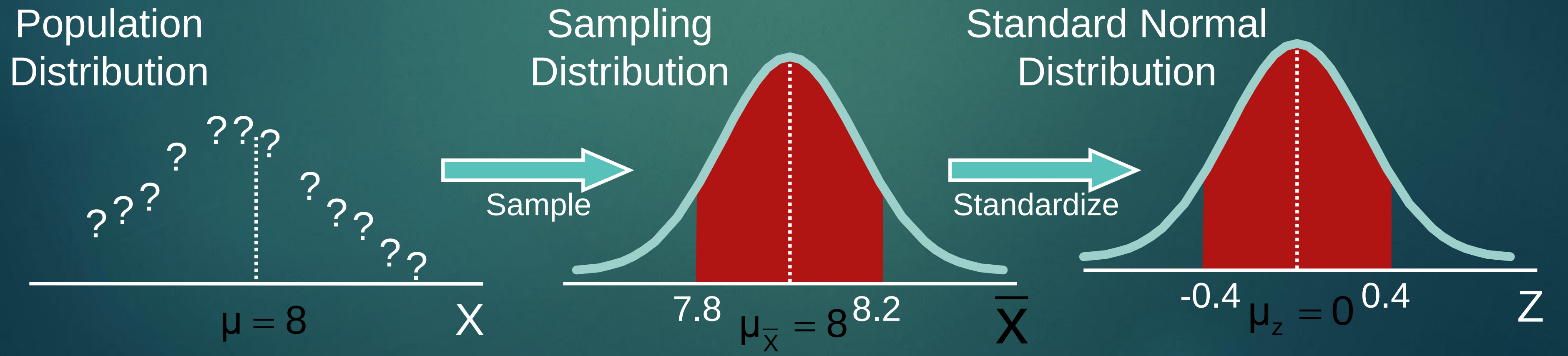
$$P(-0.4 < Z < 0.4) = P(Z < 0.4) - P(Z < -0.4)$$
$$= 0.6554 - 0.3446$$
$$= \mathbf{0.3108}$$



Example

What is the probability that the sample mean is between **7.8** and **8.2**?

$$P(7.8 < \bar{X} < 8.2) = 31\%$$



Sampling Distribution of Proportions

Population Proportions

- π : the proportion of the population having some characteristic
  - Examples:
    - Market shares: The proportion of customers that purchase the products produced by a given firm compared to the share of all the other firms that produce the same product.
      - Population Proportion (Parameter) → The market share in the US
      - Sample Proportion (Statistic) → The market share in a random sample

Population Proportions

- π : the proportion of the population having some characteristic

- **Sample proportion (p)** provides an estimate of π:

$$p = \frac{X}{n} = \frac{\text{number of items in the sample having the characteristic of interest}}{\text{sample size}}$$

- **$0 \leq p \leq 1$**

Population Proportions

- π : the proportion of the population having some characteristic

- **Sample proportion (p)** provides an estimate of π:

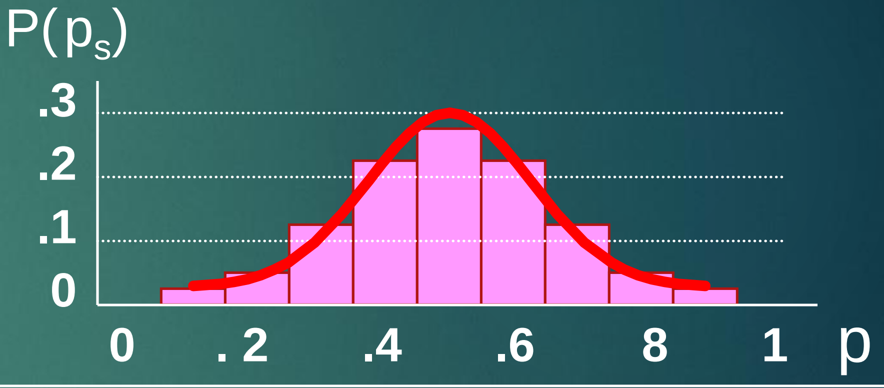
$$p = \frac{X}{n} = \frac{\text{number of items in the sample having the characteristic of interest}}{\text{sample size}}$$

- **$0 \leq p \leq 1$**

- Assuming sampling with replacement, **p** is approximately distributed as a **normal distribution** when n is large such that **$n\pi \geq 5$** AND **$n(1-\pi) \geq 5$**

Sampling Distribution of Proportion

- Approximated by a normal distribution if:
  - $$n\pi \geq 5$$
    and
    $$n(1-\pi) \geq 5$$



Sampling Distribution of Proportion

$$\mu_p = \pi$$

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

(where π = population proportion)

Z-Value for Proportions

Standardize p to a Z value with the formula:

$$Z = \frac{p - \pi}{\sigma_p} = \frac{p - \pi}{\sqrt{\dfrac{\pi(1 - \pi)}{n}}}$$

Example

- If the true proportion of voters who support Proposition A is $\pi = 0.4$, what is the probability that a sample of size 200 yields a sample proportion between 0.40 and 0.45?

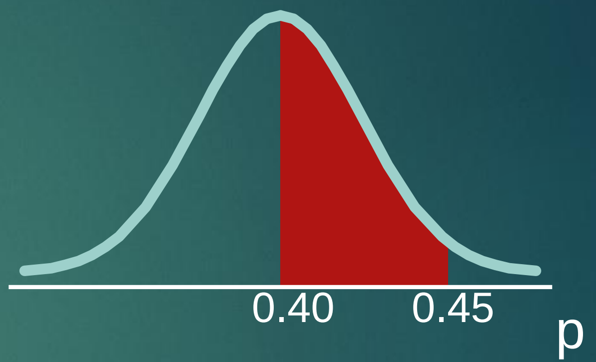
- i.e.: **if $\pi = 0.4$ and $n = 200$, what is**

  $P(0.40 \le p \le 0.45)$?

Example

$$\mu_p = \pi$$

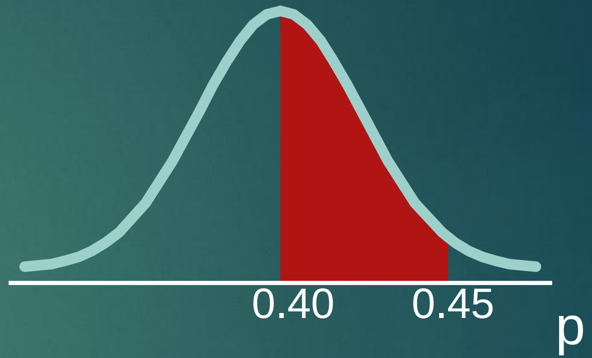
$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$



Example

$$\mu_p = \pi$$

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$



$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.4(1-0.4)}{200}} = 0.03464$$

Example

$$\mu_p = \pi$$

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.4(1-0.4)}{200}} = 0.03464$$

$$P(0.40 \le p \le 0.45) = P\left(\frac{0.40-0.40}{0.03464} \le Z \le \frac{0.45-0.40}{0.03464}\right)$$

$$= P(0 \le Z \le 1.44)$$

Example

- if $\pi = 0.4$ and $n = 200$, what is
  $P(0.40 \le p \le 0.45)$?

Utilize the cumulative normal table:

$P(0 \le Z \le 1.44) = 0.9251 - 0.5000 = \boxed{0.4251}$