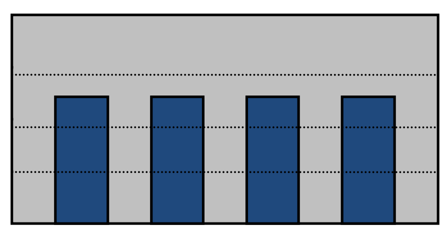
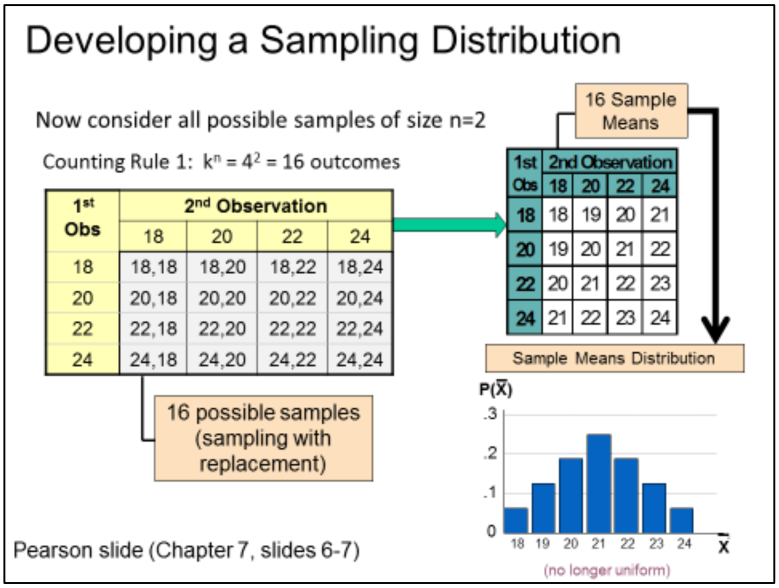
STAT 206: Chapter 7 (Sampling Distributions)

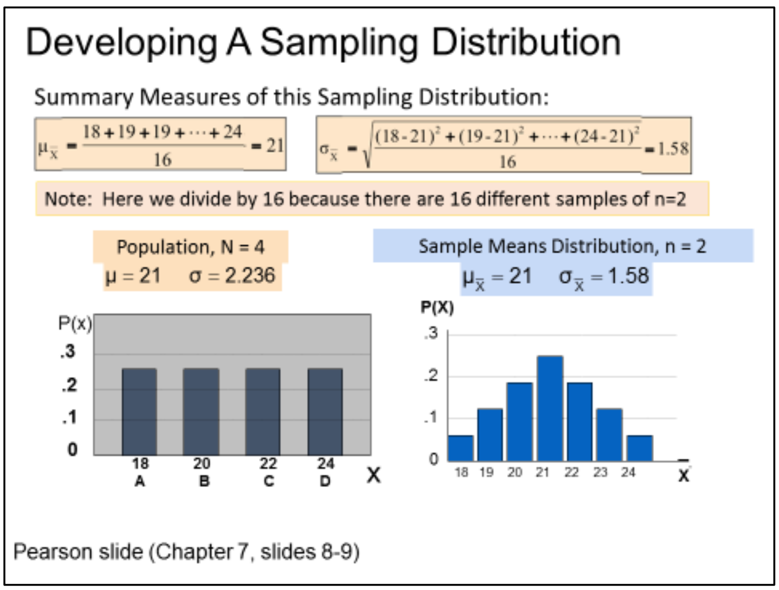
Ideas in Chapter 7:

- Concept of the sampling distribution
- To compute probabilities related to the sample mean and the sample proportion
- The importance of the Central Limit Theorem

- 7.1 Sampling Distributions
  - **SAMPLING DISTRIBUTION** is a distribution of all of the possible values of a **sample statistic** for a given sample size selected from a population
  - **EXAMPLE**: Cereal plant Operations Manager (OM) monitors the amount of cereal in each box. Main plant fills thousands of boxes of cereal during each shift. Speed of process produces variability. Average weight must be 368 grams of cereal. To maintain quality control, does OM take just ONE sample?
  - **EXAMPLE**: Suppose you sample 50 students from USC regarding their mean GPA. If you obtained many different samples of size 50, you will compute a different mean for each sample.
  - We are interested in the distribution of all potential means for a particular sample size (n is the same for each sample)
- **Developing a Sampling Distribution**
  - Assume there is a population …
    - Population size N=4
    - Random variable, X, is age of individuals
    - Values of X: 18, 20, 22, 24 (years)
    - $\mu =$
    - $\sigma =$



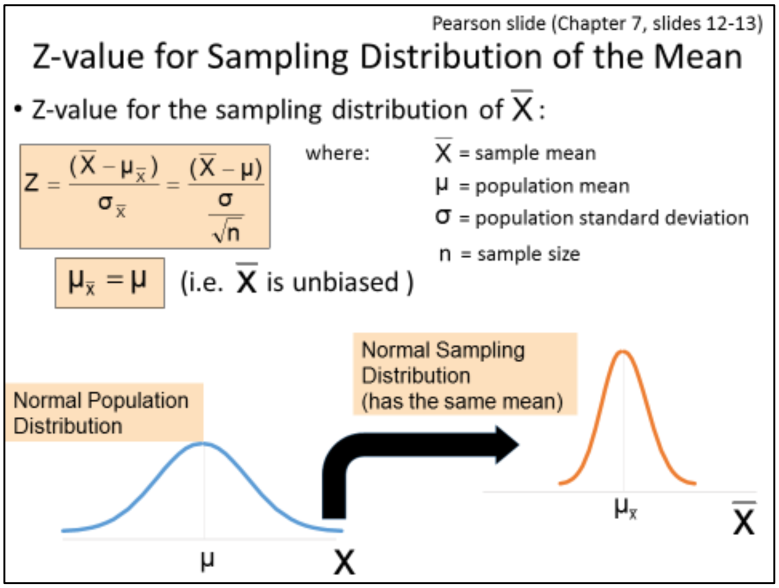




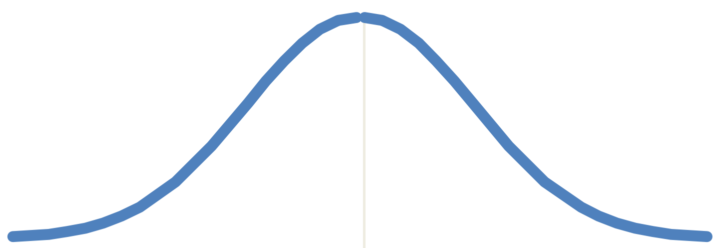
- **7.2 Sampling Distributions of the Mean**
  - **SAMPLING DISTRIBUTION**

  - Different samples of the SAME SIZE from the SAME POPULATION

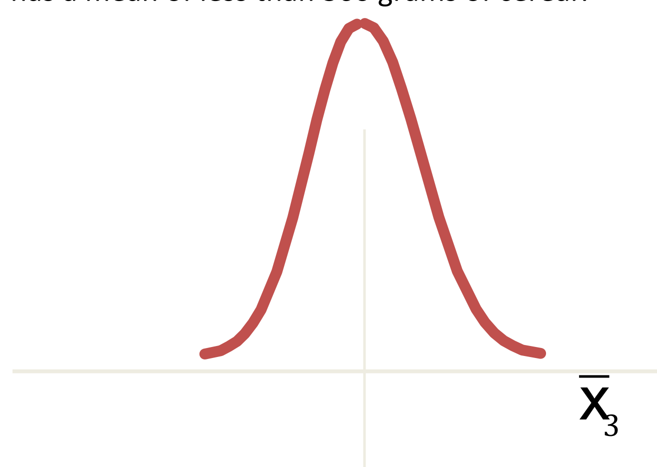
- If a **population is normal** with mean **μ** and standard deviation **σ**, the sampling distribution of $\bar{X}$ is also normally distributed with mean = $\mu_{\bar{X}} = \mu$ and
standard deviation (standard error of the mean) = $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

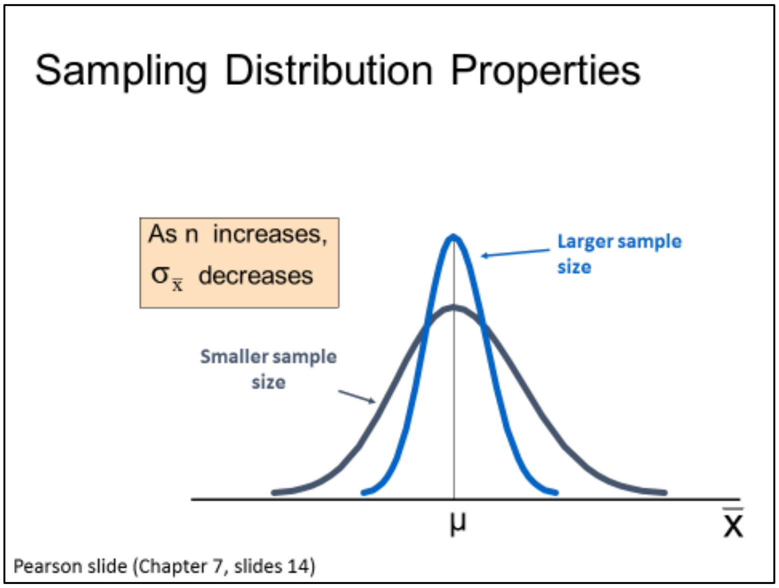


- **EXAMPLE**: Cereal plant Operations Manager (OM) monitors the amount of cereal in each box. Amount of cereal in boxes (X) is approximately normal with
mean = 368 grams and standard deviation = 15.
What is the probability that a randomly chosen **box** contains less than 360 grams of cereal?



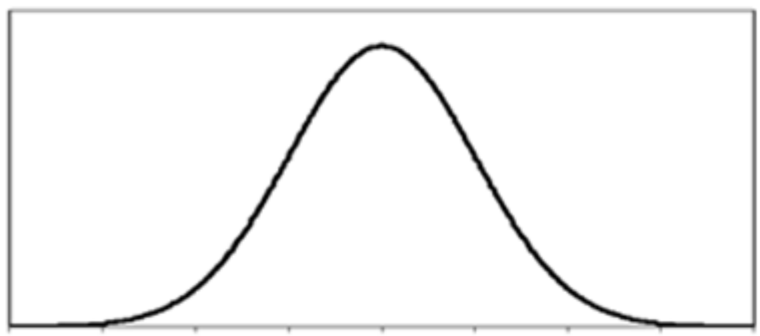
- **EXAMPLE**: Cereal plant Operations Manager (OM) monitors the amount of cereal in each box. Amount of cereal in boxes (X) is approximately normal with
mean = 368 grams and standard deviation = 15.
What is the probability that a **random sample of size 25** has a mean of less than 360 grams of cereal?



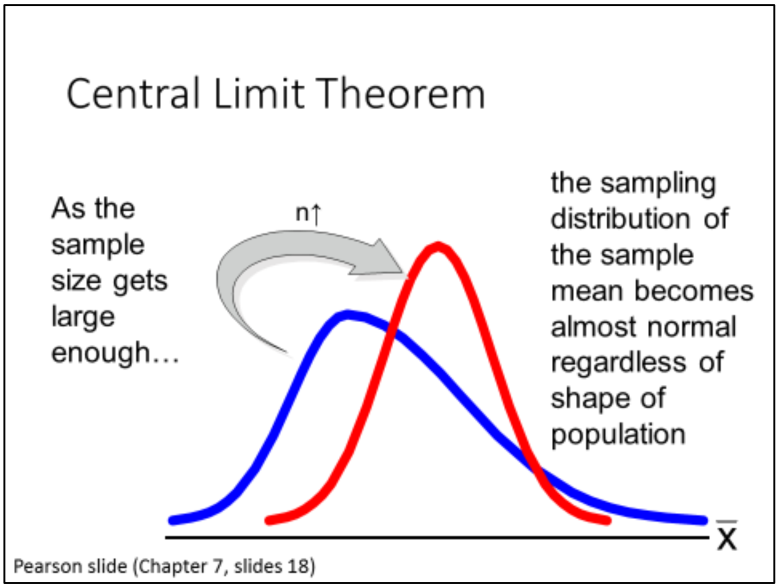


- **Interval Including A Fixed Proportion of the Sample Means?**
  - For the distribution with $\mu_{\bar{X}} = \mu = 368$ and $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{15}{\sqrt{25}} = \dfrac{15}{5} = 3$

    find interval (symmetrically) around μ that will include 95% of the sample means.

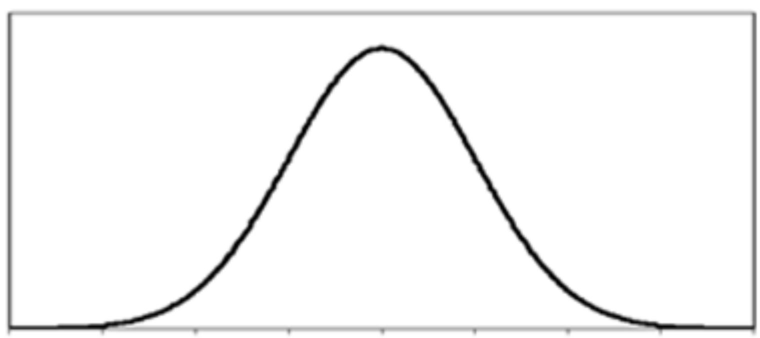


- **But what if our population is NOT normal?**
  - **CENTRAL LIMIT THEOREM!**
  - Even if the population is not normal,
  - …sample means from the population will be ***approximately normal*** as long as the sample size is large enough
  - **Properties:**
    - mean = $\mu_{\bar{X}} = \mu$ and
    - standard deviation (standard error of the mean) = $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$



- **How large is "large enough"?**

- **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?



- **Proportions**
  - $\boldsymbol{p}$ = the proportion of the population (parameter) having some characteristic (that is the proportion of "successes")
  - $\widehat{\boldsymbol{p}}$ = sample proportion provides an estimate of the parameter, p
  - $\widehat{\boldsymbol{p}} = \frac{\#\text{"}\boldsymbol{successes}\text{"}}{\boldsymbol{sample\ size}}$
  - $0 \leq \widehat{\boldsymbol{p}} \leq \mathbf{1}$
  - $\widehat{\boldsymbol{p}}$ is approximately normally distributed when n is large
    (assuming sampling without replacement from an "infinite" population of sampling with replacement from a finite population)
  - n is assumed to be "large" if $\boldsymbol{np} \geq \mathbf{5}\ and\ \boldsymbol{n}(\mathbf{1} - \boldsymbol{p}) \geq \mathbf{5}$
  - Mean of the $\widehat{\boldsymbol{p}}$ values $\boldsymbol{\mu}_{\widehat{\boldsymbol{p}}} = \boldsymbol{p}$ and $\boldsymbol{\sigma}_{\widehat{\boldsymbol{p}}} = \sqrt{\frac{\boldsymbol{p}(\mathbf{1}-\boldsymbol{p})}{\boldsymbol{n}}}$
    where $\boldsymbol{p}$ is the population proportion

- **Z-value for Proportions**
- To standardize a value of $\hat{p}$ calculated from a sample:

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

- **EXAMPLE**: If the true proportion of voters who support a particular candidate is p = 0.40, what is the probability that a sample of size 200 yields a sample proportion ($\hat{p}$) between 0.40 and 0.45?