Chapter 7 Reference: AP Statistics
Sampling Distributions

Sampling Distributions: General Information

- Sampling distribution of a statistic is the distribution of values taken by the statistic in all possible samples of the same size from the same population.

- Sampling distributions are the foundation for the study of statistical inference (what we will be doing second semester) which uses sample data to draw conclusions about the population from which the data come. Statistical inference will help answer the question, "How often would this method give a correct answer if I used it very many times?"

- We are interested in sampling/sampling distributions only when the population is large enough to make taking a census impractical/impossible (which is most of the time)

- Population $\rightarrow$ parameter, fixed, does not change (and often unknown in reality), examples include μ, p, σ (know what these symbols represent)

- Sample $\rightarrow$ statistic, almost always changes, is not fixed, varies from sample to sample, value of statistic varies in repeated random sampling, sampling variability, examples include $\bar{x}$, $\hat{p}$, s (know what these symbols represent)

- Sampling distributions (via histogram or density curve) are often described by SOCS (Shape, Outliers, Center, Spread)

- Bias (of a statistic vs. sampling methods (like convenience, voluntary, etc.) which was learned in Chapter 4): When mean of sampling distribution is not equal to mean of parameter's true value.

- Unbiased Statistic/Unbiased Estimator: A statistic used to estimate a parameter is unbiased if the mean of its sampling distribution is equal to the true value of the parameter being estimated. The statistic is called an unbiased estimator of the parameter. For example, $\bar{x}$ could be an unbiased estimator of μ or $\hat{p}$ could be an unbiased estimator of p

- Variability of a Statistic: Spread of the sampling distribution; larger samples (say n = 1000 vs. n = 50) give smaller spread (smaller SD); as long as population is at least ten times as large as the sample, spread of the sampling distribution is approximately the same for any population size.

- Be sure to look at target representations of bias and variability in the text, page 434.

OVER

Sampling Distributions of a Sample Proportion	***Sampling Distributions of Sample Means***
Proportions are just another way of looking at count data (like binomial random variables); i.e., you have two math classes ***OR*** 2/5 of your classes are math classes; so, much in this chapter will be similar to binomial chapter; $\hat{p} = \frac{successes}{sample.size} = \frac{X}{n}$	
SRS of size *n* from a large population. Then: the mean of the sampling distribution of $\hat{p}$ is exactly *p ($\hat{p}$ is an unbiased estimator for p);* the standard deviation of the sampling distribution of $\hat{p}$ is $\sqrt{\frac{p(1-p)}{n}}$ * * Note: Use the standard deviation formula only when the population is at least ten times as large as the sample	SRS of size *n* from a large population with mean μ and standard deviation σ. Then: the mean of the sampling distribution of $\bar{x}$ is exactly μ ($\bar{x}$ is an unbiased estimator for μ); the standard deviation of the sampling distribution of $\bar{x}$ is $\frac{\sigma}{\sqrt{n}}$ * (observe that the SD of the distribution decreases for larger samples; the SD decreases at the rate of $\sqrt{n}$ so you must take a sample four times as large to cut the SD of $\bar{x}$ in half) * Note: Use the SD formula only when population is at least ten times as large as the sample Important: All of the above are true no matter what the population distribution looks like (i.e., Normal, skewed, etc.)
Use the Normal approximation to the sampling distribution of $\hat{p}$ for values of *n* and *p* that satisfy $np \geq 10$ and $n(1-p) \geq 10$ (same as binomial approximation!)	SRS of size *n* from a Normal population with mean μ and standard deviation σ, then the sampling distribution mean $\bar{x}$ (unbiased estimator of μ) is also Normally distributed; If shape of population is not Normal or is unknown, consider the Central Limit Theorem (CLT): Central Limit Theorem: SRS size *n* from any population with mean μ and finite SD σ. When *n* is large (general rule of thumb, n ≥ 30), the sampling distribution of the sample mean $\bar{x}$ is approximately Normal. Three possible cases: (1) Population is Normal, then shape of sampling distribution is Normal, regardless of sample size; (2) Any population shape, small *n*, then shape of sampling distribution is similar to shape of population; (3) Any population shape, large *n*, then shape of sampling distribution is approximately Normal (CLT)