Summary (Unit 3B - Sampling Distributions)

We have finally reached the end our discussion of probability with our discussion of sampling distributions, which can be viewed in two ways. On the one hand Sampling Distributions can be viewed as a special case of Random Variables since we discussed two special random variables: the sample mean (x-bar) and the sample proportion (p-hat). On the other hand, Sampling Distributions can be viewed as the bridge that takes us from probability to statistical inference.

As mentioned in the introduction, this last concept in probability is the bridge between the probability section and inference. It focuses on the relationship between sample values (**statistics**) and population values (**parameters**). Statistics vary from sample to sample due to **sampling variability**, and therefore can be regarded as **random variables** whose distribution we call the **sampling distribution**.

In our discussion of sampling distributions, we focused on two statistics, the **sample proportion**, p-hat and the **sample mean**, x-bar. Our goal was to explore the sampling distribution of these two statistics relative to their respective population parameters, p and μ (mu), and we found in **both** cases that under certain conditions the **sampling distribution is approximately normal**. This result is known as the **Central Limit Theorem.** As we'll see in the next section, the Central Limit Theorem is the foundation for statistical inference.

> **Outside Reading:** Little Handbook – Behavior of Sample Means (≈ 3000 words)

Sampling Distributions

A **parameter** is a number that describes the population, and a **statistic** is a number that describes the sample.

- Parameters are fixed, and in practice, usually unknown.
- Statistics change from sample to sample due to sampling variability.
- The behavior of the possible values the statistic can take in repeated samples is called the **sampling distribution** of that statistic.
- The following table summarizes the important information about the two sampling distributions we covered. Both of these results follow from the **central limit theorem** which basically states that as the sample size increases, the distribution of the average from a sample of size n becomes increasingly normally distributed.

			Sampling Distribution		
Variable	**Parameter**	**Statistic**	**Center**	**Spread**	**Shape**
Categorical (example: left-handed or not)	p = population proportion	$\hat{p}$ = sample proportion	p	$\sqrt{\frac{p(1-p)}{n}}$	Normal if np ≥ 10 and n(1 - p) ≥ 10
Quantitative (example: age)	μ = population mean, σ = population standard deviation	$\bar{x}$ = sample mean	μ	$\frac{\sigma}{\sqrt{n}}$	Normal if n > 30 (always normal if population is normal)