Chapter 5: Sampling Distributions for Means

**Population Distribution**

The population distribution of a variable is the distribution of its values for all members of the population.

- This is also the distribution of the variable when we choose one individual from the population at random.

**The Distribution of a Statistic**

- A statistic from a random sample or randomized experiment is a random variable.
- The probability distribution of the statistic is its sampling distribution.

- If our random variable follows a normal distribution, we can use the normal tables to calculate probabilities...this doesn’t happen in real life very often

**So...we use the Central Limit Theorem**

Draw a SRS of size $n$ from any population with mean $\mu$ and finite standard deviation $\sigma$. When $n$ is large, the sampling distribution of the sample mean $\bar{X}$ is approximately normal:

$$\bar{X} \text{ is approximately } \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$$

In other words, regardless of the original distribution of $X$, when you have a large sample size ($n$), then the distribution of $\bar{X}$ is approximately normally distributed.

**Mean and Standard Deviation of a Sample Mean**

Let $\bar{X}$ be the mean of an SRS of size $n$ from a population having mean $\mu$ and standard deviation $\sigma$. The mean and standard deviation of $\bar{X}$ are

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

$$Z = \frac{\bar{X} - \mu_\bar{X}}{\sigma_{\bar{X}}}$$

The formula we use for $Z$ is:
**Example.** (Moore and McCabe 4th edition)

Scores of students on the ACT college entrance examination in 2001 had mean $\mu = 21.0$ and standard deviation $\sigma = 4.7$. The distribution of scores is only roughly normal. A SRS of 50 students who took the test is selected. What is the approximate probability that the sample mean of these students is 23 or higher?

**Example.**

A soda company used a filling machine to fill their bottles of soda. The company claims that the bottles contain 380 ml of soda. However, the contents of a bottle $X$ are known to vary according to a normal distribution with mean 380 ml and standard deviation 2 ml. Suppose a SRS of 10 bottles is selected from the plant. What is the probability that the mean content of the 10 bottles is between 379 and 382 ml?

Is it appropriate to use the normal distribution here? Why or why not?

Yes, because the original distribution for $X$ is normal, the $\bar{X}$ sample mean distribution will also be Normal. If the distribution for $X$ had not been Normal, then as long as we have a large (>30) sample size for a SRS, then the sample mean $\bar{X}$ distribution will be Normal according to the Central Limit Theorem.
Additional problems:

1. The average number of years graduating seniors take to complete their degree is normal with mean 4.5 and standard deviation 0.75. John, who took 6 years to complete his degree, believes the average is wrong. He takes a random sample of 25 graduating seniors and finds they took an average of 4.75 years to complete their degree. What would the probability be that the sample mean is greater than 4.75 given that the population mean is actually 4.5?

2. In 2000, according to the U.S. Census Bureau, the number of people in a household had a mean of 2.6 and a standard deviation of 1.5.

   a. What is the probability that a randomly selected household has more than 5 people?

   b. Suppose Ted who lives in a community that had large families thinks the Census Bureau is wrong. He selects a SRS of 50 households he finds the sample mean to be 3.5. What is the probability that the sample mean will be more than 3.5 given the population mean is truly 2.6 and the standard deviation is 1.5?
3. The time that a technician requires to perform preventative maintenance on an air-conditioning unit is governed by an exponential distribution with a mean of 1 hour and a standard deviation of 1 hour. Your company operates 70 of these units. What is the probability that their average maintenance time exceeds 50 minutes?

4. A laboratory weighs filters from a coal mine to measure the amount of dust in the mine atmosphere. Repeated measurements of the weight of dust on the same filter vary normally with standard deviation of 0.08 mg because the weighing is not perfectly precise. The dust on a particular filter actually weighs 123 mg. Repeated weighings will then have the Normal distribution with mean 123 mg and standard deviation 0.08 mg. What is the probability that the laboratory reports an average of 3 weighings as 124 mg or higher for this filter?
5. The level of nitrogen oxides (NOX) in the exhaust of a particular car model varies with mean 0.9 grams per mile and a standard deviation 0.15 g/mi. A company has 125 cars of this model in its fleet. What is the level L such that the probability that the sample mean is greater than L is only 0.01?

6. The NCAA requires Division I athletes to score at least 820 on the combined math and verbal parts of the SAT exam to compete in their first college year. In 2002, the scores of the 1.3 million students taking the SATs were approximately normal and mean 1020 and standard deviation 207. What percent of all students had scores less than 820?