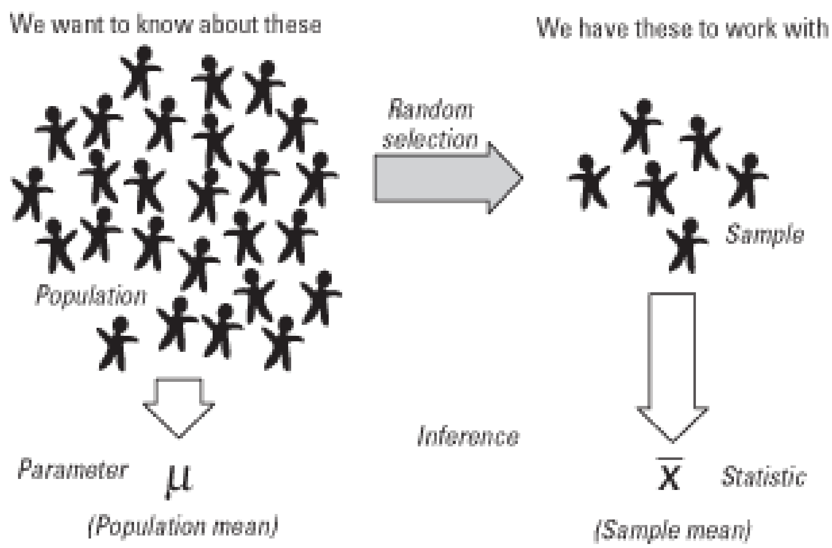
**Math 247: Sample Means of Random Samples** (Section 9.1)

What do we do if we have a research question that pertains to a **quantitative** variable? Give some examples of quantitative research questions:

Our goal in this section is to figure out how reliable a sample mean is as an estimator of a population mean.

Once we've determined this, we can answer quantitative research questions!




**Review: statistics vs. parameters.**

Parameters are values that pertain to the __________________________________________________.

Statistics are values that pertain to a ________________________________________.

We use ____________________________________ to estimate _____________________________________.

So, we have a **sample (**with **KNOWN** values**)** that provides us ***clues*** about the what's **true of the population (**which is **UNKNOWN).**

To understand how sample means "behave" we have to first look at situations where we know everything about the population (we have ________________________ data), then examine what happens when we draw random samples, find the mean of those samples, and see how well those means represent the true population mean.

**Example:** Imagine you had the census data on how many hours adults sleep per night. What do you think the shape of the distribution of sleep hours would be; i.e., left skewed, right skewed, uniform, or symmetric?

Sketch your idea here:

Using the Pop 1 Sleep Hours data in the Rossman/Chance Applet for Sampling Distribution of the mean, check the shape of the population distribution of Sleep Hours. The shape is approximately _________________________

The mean is ________________________ and the standard deviation is ______________________________

Now take samples of size n = 3.

What do you notice about the mean of **each** individual sample vs. the mean of the population?

What do you notice about the **mean** of the sample means?

What value measures the spread of the distribution?

What do we call this number when it pertains to a Sampling Distribution?

Now describe shape you see for the **sampling distribution** of sample means for each of the following and note what happens to the spread of the sampling distribution.

| | n = 3 | n = 10 | n = 25 |
|---|---|---|---|
| Shape? | | | |
| Mean? | | | |
| Spread? | | | |

**Example:** Imagine all the pennies currently in circulation (the pennies that people and banks and stores and such actually have).

If you knew all the ages of these pennies (had this HUGE census data set), what would the shape of the distribution of ages be; i.e., left skewed, right skewed, or symmetric?

Sketch your idea here:

Using the Penny Age data in the Rossman/Chance Applet for Sampling Distribution of the mean, check the shape

of the population distribution of Penny Ages. The shape is ____________________________________________

Describe the shape you see for the sampling distribution of sample means for each of the following:

| | n = 3 | n = 10 | n = 25 |
|---|---|---|---|
| Shape? | | | |

If the underlying population is skewed, and the sample size is small, the sampling distribution of means is NOT Normal!!!

But. as the sample size increases, the shape of the Sampling Distribution of the Means becomes approximately Normal, with n = 25 (some books say n = 30) being a rule-of-thumb cut off for the distribution to be "Normal enough" to use z-values for making inferences.

**Summary:**

The sample mean is an unbiased estimator of the population mean:

The shape of the Sampling Distribution of the sample mean is approximately normal, provided

Either (a) the underlying **population** is normally distributed

Or (b) the **sample** size is large (a sample size of 25 or more)

As the sample size increases, the standard error (spread of the sample means) gets smaller.

**The Central Limit Theorem** (Section 9.2)

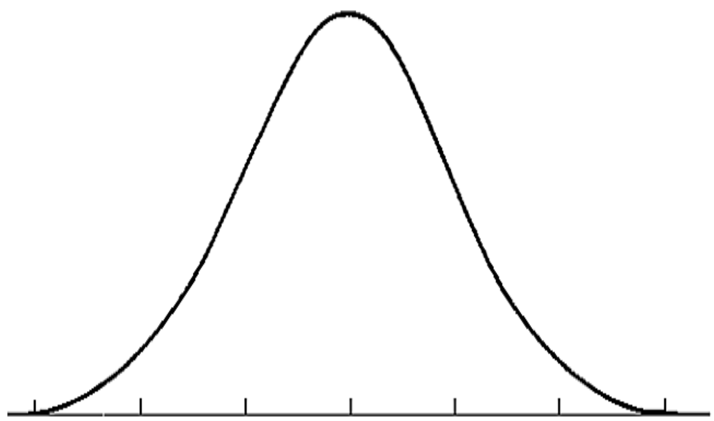
**Conditions** that must be met **for Central Limit Theorem to apply** (page 416):

1. **Random and Independent.** The sample must be random and the observations in the sample must be independent from one another.
2. **Large Sample.** Either the population distribution is Normal OR the sample size is large ($n \geq 25$)
3. **Large Population.** The population must be at least 10 times the size of the sample.

**The Central Limit Theorem**. Provided the conditions above are met (large sample or population distribution is normal) then *the distribution of sample means will be approximately normal*.

The mean of the sample means is the same as the population mean. (The sample mean is an "unbiased estimator" of the population mean.)

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



The standard error is the standard deviation of the all the sample means.

Notation and formula: SE = $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

**Example: Air Pollution and Children's Health.** One of the measurements used to determine the health of a person's lungs is the amount of air a person can exhale under force in one second. This is called the "forced expiratory volume in one second" and is abbreviated $FEV_1$.

Previous studies have established that the mean $FEV_1$ for 10-year-old boys is 2.1 liters and that the population standard deviation is 0.3 liters. A random sample of 100 10-year-old boys who live in a community with high levels of ozone pollution is found to have a mean $FEV_1$ of 1.95 liters. What is the probability that the sample of boys in the polluted of areas would have a mean of 1.95 liters if the true population average for boys in polluted areas was actually 2.1 liters?

Before we compute, are the conditions for the Central Limit Theorem met in this example? Explain your answer.

Use the sampling distribution for the mean to find the probability.

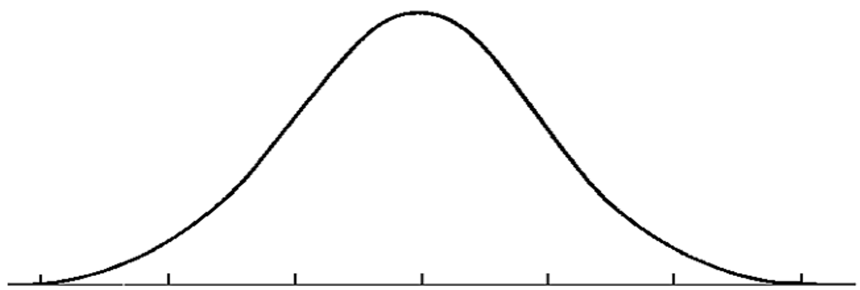
Parking Lot!

Find the mean and standard error for the sampling distribution.  Use proper notation!

Fill in the mean and standard error marks on the distribution.

Plot the <u>sample mean</u> and shade in the area that corresponds to the probability of getting an observation this low or lower if the average FEV1 of kids in polluted areas is really the same as the average for all boys.

Sketch the z-axis under the $\overline{x}$ - axis



Find z for the sample mean.

$$z = \frac{\overline{x} - \mu}{SE}$$

Based on the z-value alone, what would the probability be of getting a sample with a mean of 1.95 liters, if the true mean was 2.1 liters?

Discuss.

If this were a hypothesis test, what would we call this probability?