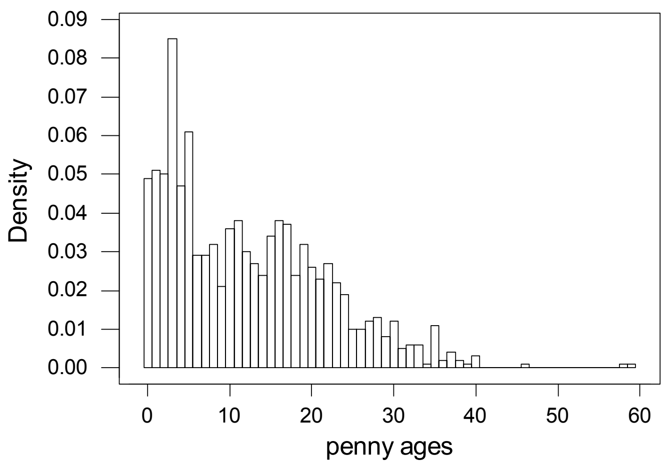
# Stat 321 - Day 32
## Sampling Distributions of Sample Means

You have begun to study the probability (sampling) distributions of statistics, which are functions of the random variables in a random sample. You have calculated the exact sampling distribution of several statistics with very small sample sizes, and you have used simulation to approximate sampling distributions. Today you will focus on the sample mean $\overline{X}$ as the statistic of interest, and you will use simulation to explore what happens to its sampling distribution as the sample size increases and with different shapes of population distributions.

Another way to think of a *random sample* is by selecting values from a population so that each value in the population has the same chance of being chosen and all possible samples of that size have the same chance of being chosen. A *parameter* is then a numerical characteristic of the population, while a *statistic* is a numerical characteristic of the sample.

**Example: Penny Ages**
Consider a population of 1000 penny ages, whose distribution is pictured below:




(a) Comment on the shape of this distribution.

(b) Open the Minitab worksheet `pennies.mtw` and calculate the mean and standard deviation of these 1000 penny ages (`MTB> describe c1`).

mean $\mu$ = standard deviation $\sigma$ =

(c) Take 1000 samples of size $n$=2 each from this population:

```
MTB> sample 1000 c1 c2;
SUBC> replace.
MTB> sample 1000 c1 c3;
SUBC> replace.
```

Calculate the 1000 sample means:

```
MTB> rmean c2-c3 c10
MTB> name c10 'sample means'
```

Examine a histogram of the distribution of these 1000 sample means, and calculate their mean and standard deviation:

```
MTB> histogram c10
MTB> describe c10
```

Comment on the shape and record these values:

shape:

mean of sample means:

standard deviation of sample means:

(d) Repeat (b) with samples of size $n$=5. [*Hints*: Use `c2-c6` to store the 1000 samples, then put the sample means into `c10`.] Comment on the shape and record these values:

shape:

mean of sample means:

standard deviation of sample means:

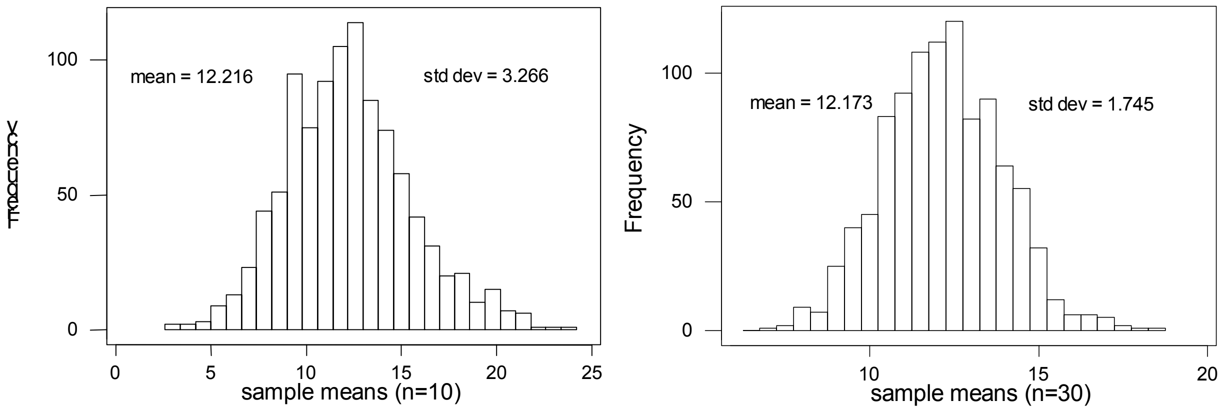
(e) How has the shape of the (approximate) sampling distribution changed as the sample size increased? How has the mean of the sample means changed? How about the standard deviation?

shape:

mean:

standard deviation:

The following histograms reveal the (approximate) sampling distributions of the sample mean $\overline{X}$ for sample sizes of $n$=10 and $n$=30:




(f) What do these simulation reveal about the *shape* of the sampling distribution of $\bar{X}$ as the sample size increases?

(g) What do these simulation reveal about the *mean* of the sampling distribution of $\bar{X}$ as the sample size increases?

(h) What do these simulation reveal about the *standard deviation* of the sampling distribution of $\bar{X}$ as the sample size increases?

**Example: Random Lunch Times (cont.)**
Re-consider Mary, with her lunch arrival time following a uniform distribution between 0 and 60 minutes after noon (with mean $\mu$=30 and standard deviation $\sigma$=17.32).

(i) Use simulation to investigate the distribution of her sample mean arrival time in a sample of size $n$=2. [*Hint*: Use the `uniform` subcommand to the `random` command.] Then repeat for sample sizes of $n$=5 and $n$=20. Record your findings in the table below.

| sample size | shape of samp. dist. | mean of sample means | std. dev. of sample means |
|---|---|---|---|
| 2 | | | |
| 5 | | | |
| 20 | | | |

**Example: Birth Weights (cont.)**
Recall that birthweights of babies born in the U.S. follow a normal distribution with mean 3250 ounces and standard deviation 550 ounces.

(j) Use simulation (`normal` subcommand to `random` command) to investigate the distribution of sample mean birthweights in a sample of size $n$=2. Then repeat for sample sizes of $n$=5 and $n$=20. Record your findings in the table below.

| sample size | shape of samp. dist. | mean of sample means | std. dev. of sample means |
|---|---|---|---|
| 2 | | | |
| 5 | | | |
| 20 | | | |

**Central Limit Theorem:**
Your simulation analyses should convince you that the following results (known as the *Central Limit Theorem*, abbreviated CLT) about the sampling distribution of a sample mean $\overline{X}$ are reasonable:

- The mean of the sampling distribution of $\overline{X}$ equals the population mean μ, regardless of the sample size or the population distribution.
- The standard deviation of the sampling distribution of $\overline{X}$ equals the population standard deviation σ divided by the square root of the sample size, regardless of the population distribution.
- The shape of the sampling distribution of $\overline{X}$ is approximately normal for large sample sizes, regardless of the population distribution, and it is normal for any sample size when the population distribution is normal.

(k) Prove the first of these assertions using properties of expected value. [*Hint*: Remember that

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}.]$$

(l) Prove the second of these assertions using properties of variance.