**Theorem 1** (The Central Limit Theorem). Let $X_1, X_2, \ldots, X_n$ be iid with $E(X_i) = \mu$ and $V(X_i) = \sigma^2$. If $n$ is sufficiently large then $\bar{X}$ has an approximately normal distribution with mean $\mu_{\bar{X}} = \mu$ and variance $V(\bar{X}) = \frac{\sigma^2}{n}$.

**Exercise 1** (Exercise 52). The lifetime of a certain type of battery is normally distributed with a mean value of 10 hours and a standard deviation of 1 hour. There are four batteries in a package. 95% of all packages will have a total lifetime less than what value?

**Solution 1.** We are interested in the total lifetime of the batteries in the package, $T$. The random variable $T$ is normally distributed with mean $\mu_T = N\mu = 4(10) = 40$ and standard deviation $\sigma_T = \sigma\sqrt{n} = 2$.

We are looking for the 95th percentile of $T$. Start with the 95th percentile of a standard normal random variable, $1.645$, and un-standardize. The 95th percentile of $T$ is

$$1.645\sigma_T + \mu_T = 1.645(2) + 40 = 43.29$$

**Exercise 2.** Suppose that M&Ms are produced so that the average weight of an M&M is 2.2 g with a standard deviation of 0.4 g.

a) Use the central limit theorem to estimate the probability that the total weight of a package of 36 M&Ms is less than 75 g.

b) Estimate the probability that the total weight of a package of 100 M&Ms exceeds 225 g.

c) Find the value $c$ such that about 98% of all packages of 100 M&Ms will have a total weight between $220 - c$ and $220 + c$.

**Solution 2.** We are interested again in the sample total weight $T$ for a package of M&Ms. By the Central Limit Theorem for large enough $n$ (more than 30) $T$ will be normally distributed with mean $\mu_T = n\mu = n(2.2)$ and standard deviation $\sigma_T = \sigma\sqrt{n} = 0.4\sqrt{n}$.

a) The sample size of 36 is large enough for us to apply the CLT. For this sample size we have $\mu_T = 36\mu = 79.2$ and $\sigma_T = 6\sigma = 2.4$.

$$P(T \leq 75) \approx \Phi \left( \frac{75 - \mu_T}{\sigma_T} \right) = \Phi \left( \frac{75 - 79.2}{2.4} \right) = \Phi (-1.75) = 0.0401$$

b) Again we can apply the CLT, this time with $\mu_T = 100\mu = 220$ and $\sigma_T = 10\sigma = 4$.

$$P(T > 225) = 1 - P(T \leq 225) \approx 1 - \Phi \left( \frac{225 - \mu_T}{\sigma_T} \right) = 1 - \Phi (1.25) = 1 - 0.8944 = 0.1056$$

c) We have the same situation as in part (b). Symmetry of the distribution for $T$ means that we $220 - c$ will be the first percentile for $T$. We find the first percentile of the standard normal distribution: $-2.33$. Standardizing $220 - c$, we find that

$$\frac{(220 - c) - 220}{4} = -2.33.$$ 

Solving for $c$ gives

$$c = 4(2.33) = 9.32.$$
In real world applications of statistics we usually wish to use data from a sample to make inferences about population parameters. For example, the sample mean (for a random sample) is a reasonable estimate for the population mean, so we say that $\overline{X}$ is a **point estimator** for the parameter $\mu$. A point estimator $\hat{\Theta}$ for a parameter $\Theta$ is **unbiased** if $E(\hat{\Theta}) = \Theta$. Proposition 1 of section 5.4 tells us that $\overline{X}$ is an unbiased estimator for $\mu$.

The **standard error** of an estimator $\hat{\Theta}$ is its standard deviation $\sigma_{\hat{\Theta}}$. The standard error is a measurement of the accuracy of the estimator. Unfortunately the calculation of $\sigma_{\hat{\Theta}}$ often involves unknown parameters. The solution is to use an estimator for the unknown parameters. This gives the **estimated standard error** of the estimator $\hat{\Theta}$.

For example, the central limit theorem tells us that for large $n$ the standard error for $\overline{X}$ is $\frac{s}{\sqrt{n}}$ (where $s$ is the population standard deviation). An unbiased estimator for $\sigma$ is the sample standard deviation $S$ given by

$$S^2 = \frac{\sum (X_i - \overline{X})^2}{n-1}.$$  

The estimated standard error of $\overline{X}$ is then $\frac{S}{\sqrt{n}}$.

**Exercise 3.** Tests of the shear strength of 10 random spot welds yield the following data (psi):

<table>
<thead>
<tr>
<th>392</th>
<th>376</th>
<th>401</th>
<th>367</th>
<th>389</th>
<th>362</th>
<th>409</th>
<th>415</th>
<th>358</th>
<th>375</th>
</tr>
</thead>
</table>

Assume that we know shear strength to be normally distributed.

a) Estimate the average shear strength of a spot weld.

b) Calculate the estimated standard error of your estimate in part a.

**Solution 3.**

a) Our best estimate for the average shear strength for all spot welds is the sample average $\overline{x}$.

$$\overline{x} = \frac{392 + 376 + \cdots + 375}{10} = 384.4$$

b) The estimated standard error will be $\hat{\sigma}_x = \frac{s}{\sqrt{10}}$. Use the shortcut formula $s^2 = \frac{\sum x^2 - n(\overline{x})^2}{n-1}$.

$$s^2 = \frac{392^2 + 376^2 + \cdots + 275^2 - 10(384.4)^2}{9} \approx 395.1556$$

$$\hat{\sigma}_x = \frac{s}{\sqrt{10}} \approx 6.286$$