8. Precision, Accuracy, Standard Error & the Central Limit Theorem

It is important to not only understand the assumptions and inferences we can make from statistical tests (we will start to cover this next week), but also some of the pits we can fall into in when we collect data and start basic calculations. In this lecture we will cover how precision, accuracy, bias, and error are related, the relationship to standard error and why sample size is important to consider in statistics.

8.1 Terminology

**Precision** – a measure of how close measured/estimated values are to each other

**Accuracy** – a measure of how close an estimator is expected to be to the true value of a parameter

**Bias** – how far the average statistic lies from the parameter it is estimating

**Error** – the difference between an observed value (or calculated) value and its true (or expected) value

How are these terms related (Fill in the table below)?
8.2 Standard Error (of the mean)

**Standard error** – standard deviation of a statistic

**Standard error of the mean** – reflects the overall distribution of the means you would get from repeatedly resampling

Formula for *standard error of the mean*:

- The standard error of the mean (SE\(_x\)) is the standard deviation of a sample-based estimate of the population mean. If you repeated your experiment over and over again, you would each time get a slightly different value. How much this value varies is a measure of precision of your mean estimate. The beauty of the SE\(_x\) is that you don’t actually have to repeat your experiment over and over again to get this estimate. Let’s see if we can confirm this empirically ...

- R has a random number generator `rnorm()` that generates a string of random numbers with a mean of 0 and a standard deviation of 1. Modify this command to get a different standard deviation (multiply by a number) and a different mean (add a number). Repeatedly run the code below (each time you get a slightly different random s1 dataset) and fill out the table below.

```
s1=rnorm(10)*8+20
s1
mean(s1)
sd(s1)
se1=sd(s1)/sqrt(10)
#repeat for s2 through s6
sd(list of x values)
mean(list of SE values)
```

<table>
<thead>
<tr>
<th>Sample</th>
<th>x</th>
<th>n</th>
<th>SD</th>
<th>SE=SD/√n</th>
</tr>
</thead>
<tbody>
<tr>
<td>s1</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s2</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s3</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s4</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s5</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s6</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- The standard deviation of the means (column x) should be approximately the same as any SE\(_x\) estimate, and very close to the mean of multiple SE\(_x\) estimates.

8.3 Central Limit Theorem

The Central Limit Theorem states that:

Why is the CLT an important theorem for statistics?

**CHALLENGE:**

1. How is standard error related to precision?
2. What is the difference between precision and accuracy?
3. Think about your data. What types of errors can occur? How may these errors affect the results of your analysis?