We can use a simpler process for calculating mean and standard deviation if we use a simpler type of data.

In Chapter 15, we defined an event as a particular outcome:

Rolling a fair die, the event is rolling a 3 or a 5.

We defined success as rolling a 3 or 5, and failure as rolling a 1, 2, 4, or 6. We calculate the probability of success as 1/3 (33%) and probability of failure as 2/3 (67%).

Using this information, we can construct a random experiment in which we roll a fair die 100 times. To calculate the mean and the standard deviation, we use the sample size (n), the probability of success (p), and the probability of failure (q):

$$\mu = n \cdot p$$
$$\sigma = \sqrt{n \cdot p \cdot q}$$

We can expand this to other types of events – surveys, for example. We want to estimate how many people will vote for a Democrat in the next election. We poll n people and calculate the percent of people who will vote for a Democrat – this is our probability of success (p). Probability of failure (they don't vote for a Democrat) is 1-p or q.

We calculate the mean and standard deviation for our sample, and use this as an estimate for the actual population – this is **statistical inference**.

Naturally, there is error between the sample and the population – sampling error. We express the error as a percent of the standard deviation:

$\frac{\sigma}{n}$ – this is called the **standard error**.

A confidence interval is simply a range of values for the mean that we are confident the actual population will fall into. This depends on our **confidence level**:

- If we want to be 68% confident, then we are talking about ±1 standard deviation from our calculated mean.
- If we want to be 95% confident, then we are talking about ±2 standard deviations from our calculated mean.
- If we want to be 99.7% confident, then we are talking about ±3 standard deviations from our calculated mean.

The wider the interval, the more likely the population value will fall within our confidence interval.