Confidence Intervals

Introduction to Confidence Intervals

Recall the way we are thinking about data and the inferential process:

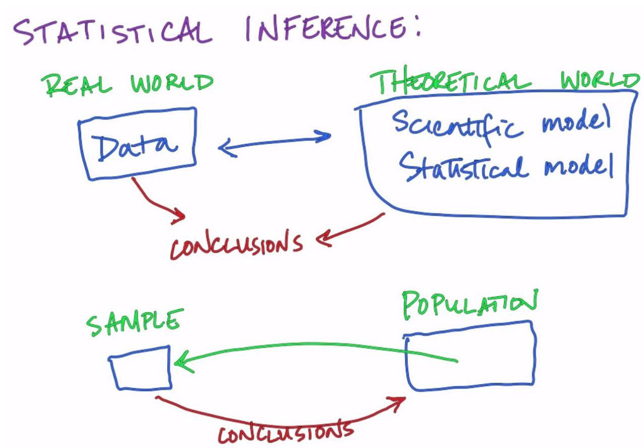

Figure 1: Overview of the inferential process

In the real world, we have observed data such as age estimates of skeletons, or counts of people with HPV infections, or the sequence of heads and tails we get when repeatedly flipping a coin. We want to use the imperfect and incomplete information from our real-world data to understand the theoretical world which we cannot directly observe. In the theoretical world, we might have a scientific model that describes the nature of the variables, and a statistical model that describes the nature of the variation or the probabilities associated with the variables and statistics that we observe in the data.

We use the data from the real world in conjunction with the models we devise in the theoretical world in order to make conclusions. Sometimes we think of the theoretical world as a population from which we observed a sample to get our data. The goal is to make inferences about the population based on that sample.

Example 1

Consider the example of flipping a fair coin 10 times. We know that a coin is equally likely to come up heads or tails and so we expect to get 5 heads in 10 flips. The probability model for the number of heads from flipping a coin 10 times is the $Binomial(10, 0.5)$ model. The binomial model is our statistical model in the theoretical world.

Of course, there is variability in the number of heads we might get in our ten flips. We do not expect to get exactly 5 heads each time. We can see from the $Binomial(10, 0.5)$ distribution in Figure 2 that 6 heads is a reasonably likely value. Although it is possible to get 0, 1, 9, or 10 heads, these extreme values are quite unlikely.

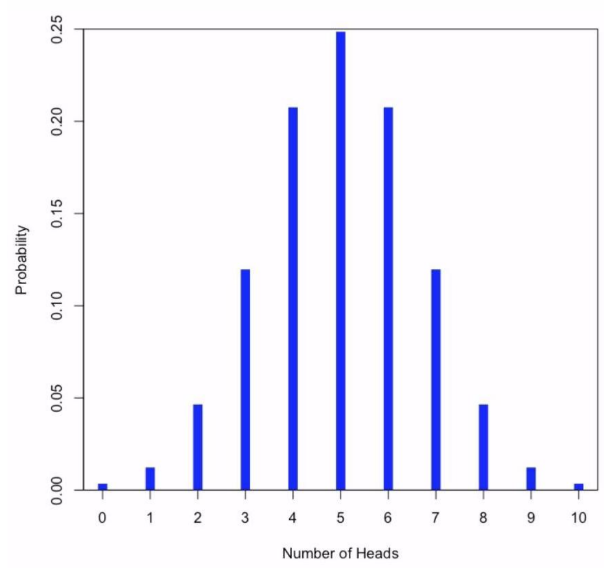

Figure 2: The $Binomial(10, 0.5)$ model for number of heads

We can also think about this in terms of the proportion of heads, Figure 3. We expect to get 50% heads or a proportion of 0.5. Observing 60% heads in 10 flips is reasonably likely. 0%, 10%, 90% or 100% heads, is quite unlikely however. Where do we draw the line for the range of possible values for the proportion of heads that is what we would expect to get most of the time?

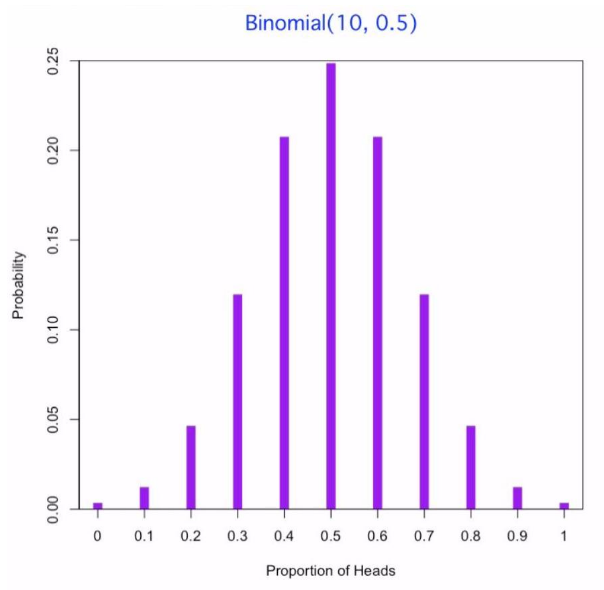

Figure 3: The $Binomial(10, 0.5)$ model for the proportion of heads

Example 2

Consider the experiment of 10 flips of a beer cap which is red on one side and silver on the other. When we carried out this experiment, the beer cap came up red 4 times. In this case, we do not know the true probability of the the beer cap coming up red. If a beer cap is equally likely to come up red or silver on each flip, then the number of reds in ten flips, would have the same sampling distribution as that for the number heads when flipping a coin ten times. We see above in Figure 2 that the value 4 has a reasonably high probability of occurring in a $Binomial(10, 0.5)$ experiment.

However, the $Binomial(10, 0.5)$ is not the only plausible model. For example, if there is only a 20% chance of getting a red cap on a flip, the theoretical model for this experiment would be the $Binomial(10, 0.2)$ model. Based on this theoretical model, observing 40% reds still seems reasonable, shown in Figure 4.

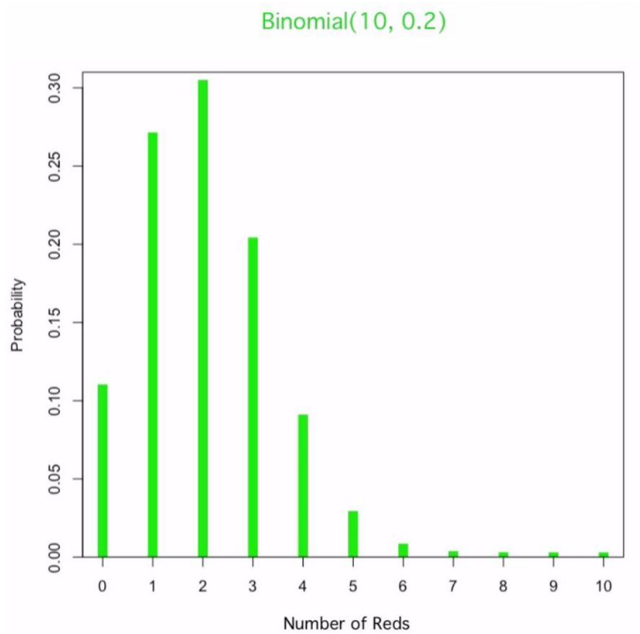

Figure 4: The $Binomial(10, 0.2)$ model for the number of heads

Example 3

Next, consider repeating the beer cap experiment 1000 times. We observed 576 reds in 1000 flips. Again, the statistical model for this situation would be binomial but we do not know the probability of getting red on any one flip. Based on these real-world data, what can we say about the proportion of times the bottle cap comes up red? Intuitively, we can use the data we observed to estimate the proportion to be 57.6%. This statistic seems like a reasonable estimate for a parameter in a theoretical world.

It would be good to be able to report more than a point estimate such as 57.6%. Most likely the true proportion is not exactly 57.6%. Thus, it would be helpful to know a range of plausible values for the theoretical world parameter. Based on these data, could it be as low as 40%? Or as high as 60%?

It may not seem like estimating the proportion of times a flipped beer cap comes up red is a pressing scientific problem but this general scenario is extremely common. Governments and political parties and the press regularly use polls to estimate the proportion of people who are going to vote for a particular candidate, or who support a particular issue. These organizations survey a sample of people to make inferences about a population. The data from a survey is the observed data analogous to our beer cap flips and how many came up red. The parameter a survey might be trying to estimate is the proportion of people who support a political candidate in a population. In our example, this theoretical parameter corresponds to the true probability of a beer cap coming up red when flipped.

Example 4

Let's introduce one more example that we'll be using in the upcoming lectures. A face lift is a common cosmetic surgical procedure. One of the goals of patients undergoing this surgery is to look younger. A few years ago some surgeons came to the Statistical Consulting Service at the University of Toronto to ask for our help in quantifying just how much younger a patient might look after performing this surgery. They wanted to be able to tell their potential patients that on average people look x number of years younger with this procedure.

The surgeons took photos of 60 patients just before their surgery and again one year later. Then, for each photo, they asked 10 people to estimate the age of the person in the photo, and they averaged the 10 estimates together for a measure of the perceived age. How much younger a person looks post surgery could then be measured as the difference in the perceived age for their before versus after photo, adjusted for the year that passed between when the photos were taken.

Based on the sample of 60 people we would like to make inferences about a theoretical world which includes the population of all facial plastic surgery patients. We can estimate the mean perceived age change but we would also like to know how precise our estimate of the mean is. If the surgery makes people look 10 years younger on average based on our data, do we think the theoretical world mean might be as high as 15, or as low as 5? In addition to the estimate of the theoretical world mean we need a range of plausible values for it.

The topic of the next few lectures is **confidence intervals** which are methods for giving a range of plausible values for our theoretical parameters, given the information we have in our data from the real world. In order to use inferential procedures we need our data to either come from a randomized experiment, or from a random sample from some population. Another assumption of inferential procedures in this course is that our data are independent and there is no relationships between any of the observations. This is the most important condition that we need for inferential procedures. If we do not have independence we need much more complicated methods.