The Process of Statistical Tests

Introduction to Statistical Tests

We have already seen a number of situations where we looked at data and saw something interesting, like a difference between two groups, and we wondered whether or not that difference is significant. In the next few sections, we will be developing the methodology for carrying out statistical tests in order to answer that question.

Determining whether a difference is statistically significant involves answering the question: "Could a difference observed just be due to chance or is it so unlikely that it must must say something else about the nature of the data, perhaps contradicting an assumption that we have made about it?"

The combination of our real world data with our theoretical world models allows us to make conclusions based on our data. How strong or how broad these conclusions are depends on how our data were collected. These conclusions could be about parameters in our models, or when our data were collected as a sample from a population, the conclusions could be about the population based on our sample. We are now going to start learning about using statistical tests in order to answer specific questions about the values of our population or theoretical world of parameters.

Example 1
Previously we compared life expectancies in two regions of the world by looking at their side-by-side box plots. The box plots show that life expectancies in almost all countries in East Asia & Pacific region are higher than life expectancies in almost every country in Sub-Saharan Africa. Could the differences between these regions be due to the natural variability of the data?

To use statistical tests to answer this, we first devise a theoretical model that assumes no difference in life expectancies between these regions. We then use our knowledge of probability and sampling distributions to see how likely it is to observe even more extreme differences between the two regions. We wish to investigate whether our real world data supports what we are assuming or hypothesizing about the theoretical world, which is that life expectancies are the same in these two regions, or whether our data are providing some evidence that our assumptions or hypotheses are false.

Example 2
In the probability lecture, Jeff flipped a beer cap that was red on one side and silver on the other ten times, and he got four reds. If we assume that a beer cap, like a coin, is equally likely to come up on either side, would four reds be surprising? Probably not.

In fact, we can't say a lot when Jeff only flipped the cap ten times, so Jeff flipped the beer cap a total of 1,000 times and he got 576 reds. That's pretty close to 50-50, but if it was perfect, he would have gotten 500 reds. So assuming this beer cap is equally likely to come up red or silver, is 576 so unlikely that we think it can't be 50-50 red or silver. If it is unlikely, we would see what we call a *statistically significant* result.

Example 3
A recent University of Toronto study had raters examine photos of 60 patients before and after having a facial cosmetic surgery. The raters gave their guess at the age of the person in the photo, and the perceived age was calculated as the average of the raters' guesses. For each patient we have the difference in their perceived age from before to after the surgery.

We are interested in whether the perceived age changes from before to after surgery. There is a probability distribution of possible perceived age changes. If we model it in our theoretical world as being centered at zero, in other words we expect no perceived age change, how likely is the mean difference that the surgeons got?

Example 4
In a book published in 1868, the German doctor Wunderlich reported on the analysis of over 1 million temperature readings from 24,000 patients (a very large data set especially in 1868). Ignoring fluctuations due to time of day, Wunderlich reported that 37 ° Celsius is the mean temperature of healthy adults.

A 1992 article from the journal of the American Medical Association reported on a study to critically appraise Wunderlich's result, which stood as the standard for more than 100 years. The researchers in this article took measurements on 148 healthy men and women and they found an average mean body temperature of 36.8 ° Celsius. Did Wunderlich's result need to be updated? In other words, if Wunderlich was right and his result still held when this modern study was carried out, how likely is it to get differences as large or larger than 0.2 of a degree?

Wunderlich's axiom is our model for the theoretical world. Our real world data do not completely agree and we need to determine whether these data have a reasonable chance of occurring or whether they are giving us evidence that our theoretical world model might not be correct.

In this chapter, we will be studying the methodology of statistical tests. They will give us a structure to investigate the types of questions that we have illustrated in these examples. We will be able to conclude if observed differences are due to chance or perhaps they contradict an assumption we have made about our theoretical model.