From Interval Estimation (Ch 7) to Hypothesis Testing (Ch 8)

Confidence Intervals are a way of using sample data to produce an interval that is likely to include a population parameter value. For example, if we have a random sample of 49 SFU undergrads and survey them about the amount of family financial support in the past twleve months, then we might want to know what the average family financial support was among all SFU undergrads. Suppose the sample data from the sampled students averages $1000 and has an SD of $500, then we would construct an interval estimate, a 95% Confidence Interval for example, of $\$1000 \pm 2 * 500/\sqrt{49}$ or (857,1143). The meaning of this interval is that there is a 95% probability that an interval computed in this way would include the true population mean, and we say that we are 95% Confident that the mean support is in (857,1143). Often this is all we need to do.

However, suppose someone says that last year, the support was $1225 on average, based on an expensive census survey of all SFU undergraduates. We may now have a slightly different question to ask of our population this year – Has the average support declined? Note that our 95% confidence interval does not include $1225, so already we may find it unlikely that the population mean is the same as last year. But it would be nice to know how unlikely – is it credible that the population mean has not declined - would we say that the "hypothesis" of "no change" has been disproved?

Hypothesis testing is a way to evaluating the credibility of a hypothesis. Usually, it is the credibility of the hypothesis of "no change" or "no difference" in a population parameter, that we are testing with our sample data. It is traditional, if not quite obvious, that the way we would test if the average support was lower this year than last is to compute the probability that would tell us, if the current population mean still were $1225, then how likely is it that the sample mean this year (in our sample of 49 cases) would be $1000 or less? This probability is called a P-value. When the P-value is small, we reject the "no change" hypothesis. In the example described, the P-value is the probability that a $N(1225,500/\sqrt{49})$ random variable is less than 1000. We compute $z = \dfrac{1000-1225}{500/\sqrt{49}} =$ -3.15 and from Table 3 this probability is 0.0008. This is a very small probability. And yet, our random sample of size 50 did produce this mean of $1000. What can we conclude? If 1225 is the true population mean, then our sample is a very rare one. But isn't it more sensible to believe, instead of the occurrence of a rare event, that the assumption of "pop mean =1225" is wrong? In fact, is it not true that the small probability of .0008 makes the assumed hypothesis of pop mean=1225 incredible? Our rational conclusion therefore is to reject the hypothesis that "pop mean = 1225" and instead believe that this year's support must be less than last year's.

This discussion has demonstrated the logic of hypothesis testing. In summary,

The CI gave us an interval that was very likely to include the true population mean. The hypothesis test (HT) told us whether or not a proposed population mean was credible.