Random Samples

A random sample has a particular meaning in sampling and analysis theory. One does not automatically get a “random sample." Biologists tend to call about anything a “randomsample" (one word) when, in fact, a random sample is one such that each of the samples of size n has an equal probability of being chosen. This probability is

$$1 \Big/ \binom{N}{n} .$$

Like a valid conclusion, the random sample is defined much by the *method* by which it was selected, as opposed to some other criterion.

Often, we see where samples were collected only along a road or trail; however, this poor practice prevents one from making a valid inference about the population away from the road. Desert tortoises have been surveyed only on “good areas" where they are common and the sample size can be large. Again, this is poor practice (flawed methodology) because inference to other areas cannot be made (the inductive inference from the sample to the population of interest is flawed). Likewise, populations studied “close to camp" represent poor practice; the only inference must be with respect to populations “close to camp" and these have little meaning.

An *Estimator* is a formula or equation that allows an estimate to be computed, given the sample data. The estimator for a simple binomial proportion is $\hat{p} = y/n$, where the data here are the number of successes (y) out of n trials. If $y = 88$ and $n = 100$, then the *estimate* is 0.88.

A gold fish bowl nearly full of marbles is often used in FW663 to motivate concepts. The bowl contains 631 marbles, of which 236 are clear, while the others are opaque. Thus, the parameter (p) is 236/631 = 0.3740095. The *estimator* of this parameter is equation that allows an estimate to be computed, given the sample data. The estimator here is the simple binomial proportion $\hat{p} = y/n$. If we drew a random sample of $n = 100$ and observed $y = 31$ clear marbles, the *estimate* would be 0.31.

The notion of repeated sampling and achieved confidence interval coverage is a binomial process (i.e., the parameter is either covered or not covered by a particular interval).