SAMPLES TO POPULATIONS

Measurements taken on the sampling units in a sample are the only information that is observed directly in statistics. All other statements or conclusions are made using inference.

Since the statistical population is the set of all sampling units of interest, the value of a population parameter at the time when a sample is collected is considered 'fixed'. For example, image if we could know the number of sleep hours for all first year students attending the University of Toronto. Then the population parameter representing the mean number of sleep hours would be fixed because someone else collecting the same data from the same statistical population at the same time would get the same population parameter value. In contrast, descriptive statistics are not fixed because two people collecting their own samples would select di􀁽erent sampling units by chance, and end up with different mean values.

The word that connects the descriptive statistics of a sample to the population parameter of a statistical population is called 'estimation'. Essentially, the sample descriptive statistics provide an estimate of the population parameter. The action of estimating is straightforward in the sense that the descriptive statistic simply becomes the population parameter estimate. However, the causal nature of this connection between the sample and the statistical population will turn out to be quite important for inference.

For the purposes of making inferences, the descriptive statistics of most interest are the mean (m) and standard deviation (s), which are matched to the population parameters μ and σ respectively.

### Sampling distributions

In a nutshell, a sampling distribution is the distribution of some descriptive statistic that would emerge if one were to repeatedly draw samples from the statistical population.

We do not repeat studies in practice, but the study of sampling distributions is critical in statistics because they allows us to make statements about things we have not seen yet. Sampling distributions are the machinery that gives us statistical inference.

## Shape independence

The first key characteristic of sampling distributions is that as long as the sample size is sufficiently large, the shape of a sampling distribution is not influenced by the shape of the statistical population. Even if the statistical population is strongly asymmetrical, or has multiple peaks, the sampling distribution is a smooth bell-shaped distribution. The reason is that taking the mean of multiple sampling units has the effect of averaging over asymmetries in the statistical population creating a symmetrical sampling distribution.

Shape independence in the sampling distribution only holds if the sample size is sufficiently large. How large is large enough? If your statistical population is already symmetrical and roughly bell-shaped, then sample sizes of n=5 are enough to ensure that the shape of the sampling distribution is independent of the statistical population. However, if your statistical population is asymmetric or has multiple peaks, then sample sizes of around n=30 are needed to ensure that the shape of the sampling distribution is independent of the statistical population.

The second key characteristic of sampling distributions is that the variance of the distribution depends on sample size. As the size of a sample increases, the variance of the sampling distribution decreases. The reason is that the larger the sample size, the more accurate the estimate of the mean and the less it varies among repeated samples.

This lesson is about some of the formal properties of sampling distributions. By the end, you will be able to:

* Explain the term standard error.
* Calculate the estimate of the standard error of a sampling distribution using information in a sample.
* Identify the conditions when it is appropriate to assume that a sampling distribution is a Normal distribution.
* Explain when a Student's t-distribution should be used in the context of estimating a sampling distribution from a sample.