sample interval – for instance, a regular plantation, or a row crop. If you have no valid basis for stratifying (see above), then a systematic sample of the whole area is an area-unbiased way of collecting a representative sample; but note that there is a high chance of missing or under-sampling rare classes. Note that systematic sampling can be combined with stratification, in place of randomisation, within each stratum.

3. **Model-driven**. This is a relatively new approach, which offers statistical and time efficiency advantages, if it can be applied. It relies on having an independent and unbiased (but not necessarily very accurate) proxy variable which is proportional to the true value of the variable you wish to sample, over the whole area. Often this comes from a satellite data product. Then you can place quite a small and deliberate (ie non-random) sample to cover the range of the proxy variable. For instance, if you are setting out to sample tree biomass and you know that biomass is proportional to tree cover and have a map of tree cover, you can ensure that you have about 30 samples in total, spread between low and high tree cover locations. The population estimate comes from the complete coverage by the proxy variable, calibrated by the placed samples.

Practical Task: Sampling design

1. Work out how many random samples (n) you would need to place for your estimate to be within ± 10% of the true mean, with 95% confidence, for a variable with a standard deviation across your sample area equal to 25% of the mean, and assuming that the data are normally distributed? The equation is

   $n = (z_{0.5}\,\sigma/CI)^2$

   $z_{0.5}$ is the 'z-statistic' for a probability of 0.025 in each of the two tails of a normal distribution, and is equal to 1.96. CI stands for 'Confidence Interval' .

   $n = \left(1.96 \times \frac{0.25}{0.10}\right)^2 = 24.01$ But $n \in \mathbb{Z}^+ \rightarrow n = 24$

   • Strictly speaking n is actually 25 as 24 does falls just short of garuntering the specified confidence interval

2. If the same area has two rather different subclasses, making up 60% and 40% of the total area respectively, and the within-class standard deviation is 6% and 4% respectively, and you wanted an overall accuracy equal to the specification above, how many samples would you need to place in each subclass, and thus how many overall?

   $n = \left(1.96 \times \frac{0.04}{0.10}\right)^2 \times 0.4 + \left(1.96 \times \frac{0.06}{0.10}\right)^2 \times 0.60 = 1.997$ But $n \subset \mathbb{Z}' \Rightarrow n = 2$

3. Work out the shortest path to visit each sample location in the following examples. If you walk at 3 km an hour, and it takes an extra 5 minutes per location to navigate to a random location but 1 minute to navigate to a point on a systematic grid, how long is the travel time in each instance?