



Chapter 2: Populations and Sampling

Progress Questions

1. This question referred to the assignment at the end of chapter 1. Potential flaws in the method were touched upon in the discussion of that assignment there. Perhaps the greatest flaw is the inability to be able to use the information gained as a predictive measure as the two census points are precisely that: they only tell you what the traffic flow was at that location on those two occasions. It is a sort of 'snapshot' research which, while interesting in itself, has little other value. Measuring traffic flow between 9am and 10am on Monday will tell you nothing about traffic flow between 10am and 11am, or even between 9am and 10am on Tuesday. There is a great deal of this type of 'snapshot' research in the social sciences, such as opinion polling, and while they may say something very useful about the moment, as it were, the information gained has a short shelf life. You would need to measure the rate of traffic flow over an extended time period to be able to determine trends that are useful tentative predictors. Another potential flaw is the high risk of miscounting. If traffic flow is high, for example, vehicles may be passing as such a rate as to make it difficult to maintain an accurate count. The use of several observers, each with responsibility for counting particular types of vehicle (perhaps, also, travelling in a particular direction), would go a long way to overcome that.
2. It is possible to confuse the two terms: *population* and *sample*. The *population* is the total number of individual subjects whose characteristics match those you have specified and who are potential members of your sample. There is the notion of a *universal population* that consists of all subjects. The research population is a sub-set of that in which the members conform to your specification for the research. All households in the UK may form the *universal population*. However, the researcher may only be interested in those households that fall within Council Tax Band D. This sub-set of households forms the *research population*. The key feature here is that the *research population* contains *all* households in Council Tax Band D.
3. Continuing with the theme in 2, the number of households in Council Tax Band D will almost certainly be too large to survey all of them. It would be necessary to survey a sample of these households. As explained in the chapter, the difficulty lies in who should be sampled. The principle behind *probability sampling* is that that decision should be left to chance – the act of randomness. The most important principle of this approach is that every member of the *research population* has an equal chance of being selected as the sample is drawn up. This is slightly confusing as we also talk about the probability of being selected, and this needs to be explained to prevent the confusion. There is a distinction between the probability of an individual being selected and that of the individual having a particular characteristic. For example, if there are 52 playing cards, the probability of any individual

card being drawn is $1/52$, if this is an unbiased draw. However, because there are four suits (clubs, hearts, diamonds and spades) the likelihood of it being a club is $1/4$. So, while any one individual has the same chance as any other of being drawn, we might expect the sample drawn from the population to reflect the characteristics of the population. This is the key feature of probability sampling.

Returning once again to our Council Tax Band D population, there are some difficulties that need to be addressed. For example, the numbers and geographical spread of households conforming to this criterion is still too huge and diverse to handle easily. Under these circumstances, ***multistage cluster sampling*** would be the most appropriate method of drawing the sample. For example, the population might be grouped into geographical regions, then after sampling, the sample might be grouped into rural, urban, metropolitan groupings and sampled again. Within each sampled group, you may group into wards and sample the wards. Perhaps, then, the sample would be sufficiently manageable in size. At each stage of the stratification process, random sampling has been adopted to select the households to go forward into the next layer of stratification. At each stage, population and sample members have an equal chance of being selected, but we hope that they are representative of the particular characteristics we are investigating because of the relative proportion of each within the population as a whole.

If the survey were going to look at Council Tax Band D households within a particular local authority, then ***stratified random sampling*** would be a good method. The number of households is certainly considerably smaller and each would be identifiable very quickly and easily from council records. The population could then be ***stratified*** on the basis of, say, wards and then Band D households within each ward sampled. In this way, a representative sample of Band D households is the likely outcome.