

# NCEA Level 2 Mathematics

## 20. Sampling

Last time, we mainly looked at the broad picture: what we need to think about, in general, when we try to answer a statistical question. This time we will begin to think about some of the practical issues we need to overcome.

As we've already discussed, it is usually impractical to measure an entire population. Our goal is therefore to measure a smaller sample and then extrapolate our findings. This process, known as *statistical inference*, requires us to have a good method for choosing our sample so that it is representative.

We will look at several examples of bad methods of sampling to begin with.

### The examples

1. I asked all my friends whether they own a car, and none of them do.
2. A survey of high-school students samples all the Y13 students at a particular school, and concludes that only 7% of students use illegal drugs.
3. In 1936, a US presidential election poll posted questionnaires to ten million people selected from telephone books and club membership lists, and got 2.4 million responses. Based on these, they predicted a decisive victory for one candidate (57% of the popular vote). In reality, the other candidate won by a landslide (62%).\*
4. A psychiatrist finds that practically everyone is neurotic.
5. A drug trial is performed; the patients were analysed according to the treatment they actually took, rather than the treatment they were assigned at the randomisation stage of the trial.

### The problems

1. Some people aren't my friends. In addition, most of my friends live in urban areas with frequent public transport, and tend to be more affluent.
2. This doesn't include students who drop out of school, or are homeschooled. It also only measures students at a particular school, which might be more or less affluent than average and thus drug use by its students might be more or less probable.
3. Despite the large sample size, the sampling method used tended to screen out the poor (who didn't belong to clubs, or have telephones) who were more likely to vote for the other candidate.
4. The psychiatrist's patients are far from a sample of the population.
5. This might seem reasonable (if 30% of participants drop out, they didn't receive the benefit of the treatment and so shouldn't be part of the 'participated in treatment' group during analysis). However, the problem is that the question that should be being answered is 'is this treatment effective?' rather than 'out of the people who chose to take our tablets, is the treatment effective?'. After all, if people don't end up taking the medication after being given it, this is philosophically and medically the same as if the medication was ineffective.†

READING: <http://www.goodmath.org/blog/2016/09/29/polls-and-sampling-errors-in-the-presidential-debate-results/>

SURVEY DESIGN: <https://www.youtube.com/watch?v=G0ZZJXw4MTA>

### Questions

---

\*Freedman et al., *Statistics*. Section 19.2.

†See Ben Goldacre, *Bad Pharma*, pages 200-1.