

Recruiting a Representative Sample

Gabriella Rundblad
King's College London

Introduction

Why do we need a representative sample? If we are interested in drawing conclusions about our informants only and not anyone else, then we do not need a representative sample. In fact, our informants would not be a sample – they would simply be our informants. This would, however, mean that we cannot extend our conclusions and say, for example, that since we found that our student informants are opposed to student fees we can assume that most students are opposed to fees.

The necessity for a representative sample stems from two issues: 1) no one can test an entire population because even the smallest population would take too long to recruit and test, and 2) most researchers seek general conclusions that apply to a population and not just a few individuals. Therefore, recruiting suitable and representative informants is not a question of getting people to take part - on the contrary, the informants need to be selected very carefully.

In addition, the study, the intended role of the informants in it and its possible effect on the informants must be addressed, but that is the topic of my ethics paper.

Identifying a Population and Deciding Sample Criteria

The first task at hand is to determine which population needs investigating in accordance with the research question/hypothesis chosen for the study. Next, the sample criteria – i.e. the criteria we need to follow in order to ensure that our sample will be representative of the population we wish to investigate – needs to be decided. If we are doing a gender study, then we need to recruit representative men and women; if a study of gender in university education, we need representative male and female students from representative universities; if a study of gender in university education in multiethnic London, we need representative male and female students from representative universities with representative ethnic backgrounds; and so on.

Our sample criteria can involve all the usual social, geographical and linguistic factors: age, gender/biological sex, class, ethnicity, nationality, birthplace, current place of residence, education (of the informant and/or informant's parent), literacy, native language(s), IQ, etc. The list is long and the choice crucial – whether that choice involves selecting one or two informants for an interview or selecting anything from 20 to several hundred participants for a questionnaire study.

Imagine that we are intending to carry out a questionnaire study investigating gender and ethnicity where we have decided to focus in particular on black

females students in London. Clearly we do not need to recruit any males for this study. But what about the women we need. One way to ensure a representative sample is to try to find out how many black female students attend London universities and whether they are e.g. Black (British) Caribbean or Black (British) African. Many universities have this type of statistical information available (e.g. KCL annually publishes its current student profile (www.kcl.ac.uk/about/structure/admin/acareg/sda/stats/prof.html)) and there are also specialised market research organisations (e.g. MORI's Social Research Institute (www.ipsos-mori.com/sri/)) that provide useful information. If we are able to identify the various other factors for our target population, we can recruit a representative sample from it.

Controlling for Other Factors

While universities and statistical databases provide a lot of information about gender, ethnicity, disability, age and domicile, they do not always provide information on e.g. how many Black African females with no disability between the age 21-29 from England there are. In short, they might list their statistics using one factor at a time. This unfortunately means that it might be difficult identify the various other factors and thus to choose the right informants.

One way of addressing this issue is to look at the majorities for the various factors – e.g. if the majority of Black students are of African background, one could (perhaps incorrectly) assume that more African than Caribbean black females should be recruited and if the majority of students are 20 years of younger, perhaps only black females aged 20 or younger should be asked to participate in the study. Thus by mimicking various aspects of the general (student) population, one could hope to mimic one's target population.

Another approach would be to not identify the other factors, but instead to control for them. If we are investigating differences and similarities in opinions between female and male students, we need to be sure that any differences or similarities are not due to one of the other factors. We can control for age by recruiting informants of the same age, from the same part of the world, with the same background, speaking the same native language(s), etc. This approach allows us to assume that because all non-targeted factors have the same values, these factors are probably not the reason for any differences and similarities and instead they are due to the target factor gender.

How Many Informants?

After we have decided what we are studying and who we need to recruit to do our study, we need to decide how many informants we need. Clearly, there is no one answer. For a qualitative study using interviews, we need very few informants, but for a questionnaire study, an intervention study or an experiment of some sort we need many more.

But how many do we need? The answer depends primarily on the number of factors (i.e. social, geographical and/or linguistic factors) our research question/hypothesis includes.

If we are doing a study of gender in university education in multiethnic London and we want to extend our results to the general population, we need many more informants than if we are investigating black female students attending one London university. The reason for this is that for the first we need men from all sorts of backgrounds and all sorts of universities and women from all sorts of backgrounds and all sorts of universities, but for the second study we “only” need black female students from one university.

Imagine that we want to know whether there is a difference in opinion about student fees depending on whether students are working class or middle class students and/or whether they are male or female. In short we have two factors: class and gender. We hypothesise that female working class students will complain most about fees whereas male middle class students will not (I will not dwell on why we may or may not want to hypothesise this as this is just an example!) If we were to go and ask 400 students (i.e. 100 male middle class, 100 female middle class, 100 male working class, and 100 female working class) we might find that 95% of female working class students are opposed to student fees compared to 25% male middle class students with our male working class students at 65% and female middle class students at 45%. If we were to do an inferential statistics test to see if we have statistically significant class differences and gender differences, we would find that we indeed do.

We could ask whether we would have got the same level of certainty (or power of the test (see below)) if we had used fewer informants. The answer is yes and no: if we had had as few as 15 male middle class, 15 female middle class, 15 male working class, and 15 female working class, we would still have found class differences and gender differences. But if we had only had 13 students per group we would no longer have found a gender difference. In short, we need at least a total of 60 informants.

If we compare our first potential study, with another potential one where we are investigating black female students. Here we would not need to compare class or gender because there are no factors at all – we would compare black female students who oppose fees with black female students who do not. If we included as few as 16 informants in our study, and if 12 of them opposed fees and 4 did not, we would find a statistically significant difference. In short, when there are no factors involved and we are comparing something as simple as a yes/no opposition to some “obvious” issue, we can “get away with” a sample as small as 16. But maybe we need to ask if an investigation with more or less given answers is interesting and worthwhile?!

A-Priori Testing

The easiest way to find out how many informants are needed while maintaining a high power of the statistical test used is by carrying out an a-priori test. One of the most commonly used tools available for free is G*Power (www.psych.uni-duesseldorf.de/aap/projects/gpower/). The software is provided together with a manual on how to use it.

Let us look at our example with yes/no questions where we would use a chi-square test to analyse our data once collected. Here G*Power would recommend we use 52 informants – this is based on a desire for high power (i.e. 0.95), a degree of freedom of 1 (i.e. we have two groups: yes vs no sayers) and a large effect size (i.e. detect large differences). As mentioned earlier, the larger the difference the fewer informants are needed. Our example above had some enormous differences and thus we found that 16 people would be enough. What G*Power calculates are “normal” large differences – thus 52 informants. If we feel that 52 informants are beyond our capability, we can compromise with G*Power – if we say that 40 informants is what we expect to be able to recruit and ask G*Power to calculate the power of our test we find a reduction in power from 0.95 to 0.92 – which is acceptable although 0.95 and 52 informants is better.

If we look at our more complex example involving one or more factors in a questionnaire study, we find that G*Power does not have a special selection for non-parametric tests, so we use the t-test (means) option. Set for a large effect size, G*Power recommends 84 informants (or 70 if our hypothesis is one-tailed). If we compromise with 60 informants, the power of the test is reduced to 0.91 – which once again is acceptable as long as the differences we expect indeed are large because we will not be able to detect medium or small ones.

Getting Access to the Sample

On getting access to potential informants and other ethical considerations, please see my paper Ethics is Essential.

Summary

Choosing who to recruit as an informant and why is naturally very crucial to any study, and we have addressed some of the ways in which suitable informants can be identified. Nevertheless, the question most students seem to worry most about is how many informants are necessary. We can sum up our answer as: it depends on how many factors we are interested in (more factors => more informants), how certain we wish to be (more informants => better statistics), what question we ask (greater differences => fewer informants) and what “replies” we allow (yes/no => fewer informants). But above all, who you recruit is more important than how many – representativity is key to qualitative and quantitative studies alike.