What sampling problems might you have conducting an opinion poll?

The first know example of an opinion poll dates back to 1824 by the publication Harrisburg Pennsylvanian, which aimed to predict the outcome of the forthcoming Presidential election between John Adams and Andrew Jackson. However it was conducted in a particularly crude and unscientific manor. In modern times opinion polls are far more complex in their search for the best i.e. most accurate results. Accuracy in any opinion poll, no matter what the subject is the cornerstone. However having an opinion poll that is completely accurate is very problematic. Failing to take certain factors into account can lead to a catastrophically inaccurate poll. The most prominent example is this would be the Literary Digest's 1936 opinion poll for the results of that year's Presidential election. The Literary Digest conducted the first real nationwide opinion polls in 1916, sending out millions of postcards and counting the returns. In this case they correctly anticipated Woodrow Wilson's victory. Their opinion polls were again correct for the next four elections. In 1936, the 2.3 million replies that they received predicted that Alf Landon would win the Presidential election. However Franklin Roosevelt won by a Landslide. So what was different about this election compared to the previous five, which the Literary Digest had all predicted correctly? Well the problems with this opinion poll can be highlighted by looking at the opinion poll conducted by George Gallop in the same year. Whereas the Literary Digest poll had hoped to be accurate simply by the fact that is used huge a huge sample to create the poll, Gallop realised that it was not the scale of the poll that was important; it was that it was done in a scientific manor that counted.

The main reason for the failure of the LD was that it used telephone numbers details and car ownership details to find addresses to send the postcards to. In 1936 only affluent people could afford a car or a telephone; this was also during the Great Depression! The main issue for the voting public was economical. Therefore gaining opinion from a wealthy sample, when the main issue in an election is economics will never give a fair representation, and hence the LD poll proved wildly inaccurate as it excluded those of lower socioeconomic status.

So it is clear that for a poll to be accurate, it must be representative of the population as a whole. It does not matter if you poll 500 people or five million people, as long as it is representative. This is taken into account in every poll now undertaken. It is understood that it must be as representative as possible; however the problem is now actually achieving a representative sample, and there are many sampling problems to take into account that may jeopardise the accuracy of the poll.

There are two main scientific ways of conducting an opinion poll and achieving a reasonably representative sample. The first used (and slightly less accurate) is the quota method as used by Gallop in 1936. This method attempts to be representative by choosing respondents individually to match the quota of the population as a whole. So for example if you were doing an opinion poll of a country where 48% of the population was female, then this would have to be reflected in your sample. If 34% were of African descent then 34% of your sample would have to be of African descent. As well as sex and

race the sample would also have to have the correct quota in terms of age, geographical location, socioeconomic status and so on. A perfect sample will perfectly represent the country as a whole.

The other and most commonly used method of sampling for opinion polls is that of probability sampling. This works by surveying people completely randomly. Although using the word 'random' makes it sound haphazard and resoundingly unscientific, if a poll is completely random, it will become scientific and accurate, as the laws of statistics state that if it is completely random it should be quite representative. If for example, 20% of the population lives in a rural area, then in a completely random poll, the likelihood that a person in a rural area is called is one in five i.e. 20%.

There are several reasons why the probability method is a more desirable choice than the quota method. The main reason is that with the quota method the interviewers use their own discretion in choosing the respondents. This introduces a possible source of bias on the part of the interviewer. It is also very difficult to create quotas that are exactly accurate. For example 0.05 percent of the population might be quarter Greek, quarter Italian and half British in their racial make-up. Will the person conducting the poll go as far to make sure that 0.05 percent of his sample are quarter Greek, quarter Italian and half British? There are so many possible mixes in every respect that the quota method will only be able to take into account the more obvious factors, such as race, age etc. that I mentioned earlier. The probability method in theory includes every possible permutation as it is completely random, and although the sample eventually taken will only be 1500-2000, these 1500-2000 could potentially be anyone in the population. Incidentally the reason why I used the figures of 1500-2000 is because statistically this is believed to give an accurate representation of the population as whole to a degree of roughly five percent. After this the accuracy level does not tend to be reduced significantly unless a much larger sample is used.

Generally probability surveys are carried out over the telephone. The companies conducting the surveys go to great lengths to ensure that the calls are as random as possible and every household in the country has an equal chance of being called and being included in the sample. They do this by a procedure known as Random Digit Dialling. The method is very complicated, but in essence the statisticians obtain lists of every area code and number combination in the country, and from these take random digits that create a possible phone number, which are then called. With this method even new and unlisted phone numbers have an equal chance of appearing in the sample. Similar methods are often attempted using the internet, with e-mail. However the fact that once a certain percentage of the population actually have the internet immediately renders the samples very unproportional to the population as a whole.

It may appear that the probability method of sampling gives perfect results for opinion polls. However there are many problems that the probability method faces like any other, and some are more obvious than others.

The most obvious, yet possibly the most important problem with sampling is the fact that it is up to the participant in the first place whether they wish to participate in the poll. Only certain types of people choose to reply to polls. This means there is something different about those people who choose to participate than those who choose not to participate. Take an example using the probability method over the phone using RDD. Although the numbers that are dialled are completely random, the person who is being asked to take part in the poll may well say they do not wish to take part. They could decline for several reasons. They could for example be a businessman who has just arrived home from work. Someone who has not been at work all day may be less tired and feel more inclined to take part in the poll. This could lead to you having a higher percentage of people in your poll who are unemployed than people who have a full time job than is reflected in the population as a whole. Or someone who is politically-interested may be more inclined to take part in the poll. Although it may appear on the surface to be better to have people who actually take an interest in, and some knowledge of politics on the poll, if you have a higher percentage of them in the poll than there is in the population as a whole, then the poll becomes to be less representative of the population as a whole and there for less accurate. These are admittedly rather crude examples. Just because someone has been at work does not necessarily mean they will be more tired than someone that has not, but I think it demonstrates the point that I am trying to illustrate.

Secondly public opinion is always subject to change. Take the example of an American election. A week before the election a perfectly carried out opinion poll may produce very accurate results of what the public were intending to do at that point. If five days before the election, a shocking scandal came out about one of the candidates, it would completely invalidate the results as the sample were not aware of this at the time the opinion poll was conducted, but were when they voted.

Another problem that a completely random sample does not allow for is any bias in the way that the question is presented. For example a question such as 'Would you consider voting for Mr X considering the mess he made in his last term in office' has very negative connotations, and is leading the respondent. As would be something along the lines of 'Do you think Mr X would make a good Prime Minister considering the excellent job he did as home secretary'? All adjectives must be as neutral as possible, and leading the respondent in any way could cause inaccuracies.

There are also many things that a respondent may do to invalidate results. An example would be to possibly put forward a more extreme position than they actually hold in order to make it appear that people are more passionate about their side of the argument.

In conclusion, you can see that there are many problems in sampling when conducting an opinion poll. Through methods such as RDD it is now possible to get a far more representative sample than was previously possible. However this is only half the problem solved. Even a perfectly representative sample (which in itself is impossible) would not yield completely accurate results because of the factors in mentioned above. A completely accurate sample is not possible.