University of Toronto

The Statistics Behind Political Polling:

How population samples are collected and corrected for bias

Aamir Ali

STAD92

Professor Kenneth Butler

5 December 2022

Contents

Introduction

Political Polling, also known as opinion polling, is the process of ascertaining how a population perceives relevant political issues at any given point in time. Like any other scientific research study, it is an intricate and complex process that requires many hurdles to be overcome for rigorous implementation. This paper is interested in exploring the intersection sociology and statistics in the conductance of polling research by asking the question: how is political polling conducted, and how do researchers accurately sample “non-standard” populations?

This question will be answered by exploring two texts relevant to the topic. The first text is the book *Sampling* by Steven K. Thompson, a textbook on statistical sampling methods, and the second book is *The Oxford Handbook of Polling and Survey Methods* (subsequently referred to as *Oxford*), edited by Lonna Rae Atkeson and R. Michael Alvarez. These books were selected since they are authoritative sources on the mathematical and practical sides of polling respectively. Since *Sampling* covers more methods than are just used in polling, this paper will examine the chapters on simple random sampling, stratified sampling, and cluster and systematic sampling. The math behind these techniques will be examined under the lens of chapters from the Survey Design and Data Collection sections of *Oxford*.

The books will be analyzed in 4 sections. First, the statistical sampling methods will be explained in a purely theoretical context to lay the mathematical foundation for polling. Next, various scenarios from *Oxford* will be explored to contextualize the sampling methods into the real world, after which the challenges of political polling, specifically with regards to abnormal populations, will be analyzed. Finally, the paper will look at the discussed theories and scenarios under a critical lens to understand the shortcomings in conventional polling.

Section 1: Sampling methods

The foundation of any poll must be a mathematically sound theory that allows researchers to make meaningful conclusions from the data they collect. It is both wasteful and potentially harmful to the public to conduct research based on faulty premises or foundations as it will lead to flawed conclusions. Political polling is dependent on sampling as it would be extremely costly and time-intensive to poll the opinion of every single individual in a large country like the US on a given issue, even assuming that they all are willing and able to answer the questions. As such, population samples are taken using various methods to ensure they are representative of the population as possible.

The first, and most basic, method of population sampling is simple random sampling, or SRS. As the name suggests, this method involves simply picking members of a population at random to create the sample. This means that every individual and significant feature of the population has a proportionally equal chance of getting selected. For example, if 2 people need to be sampled from a population of 10 individuals where 5 are male and 5 are female, using SRS gives everyone a 10% change of being selected and therefore each gender a 50% chance of getting selected. While this sounds like the ideal sampling methodology, experimentally it is equally likely for every one of the 45 combinations of 2 individuals from a population of 10 to be sampled, but only 25/45, or ~56%, of the possible samples will contain both genders. With this small population and sample, we have just over a 50-50 chance of getting a useful sample, once population sizes increase and samples grow proportionally smaller, the probability of picking a representative sample grows smaller. Selecting 10 individuals at random from a population of 100 individuals yields a sample of 5 males and 5 females with a probability of $\frac{(50c5)^2}{100c10} = 0.259$, or 26%. While the above examples are specifically contrived to highlight the shortcomings of SRS, if nothing is known about a population, collecting random samples, preferably more than one, is the only way to learn enough about it so that better-informed studies can be done in the future.

If we do know the distinct groups present in a population, we can section-off the population to collect a stratified random sample. If we have prior knowledge that a given population contains 400 doctors, 400 carpenters, and 20 programmers, we can use SRS to obtain 40 doctors and carpenters each and 20 programmers to build a sample of 100 individuals that represents the population. If we also made sure to stratify each profession between genders, we now have a sample of 50 males and 50 females, 20 of whom are doctors and carpenters with 10 programmers each. If the subject of the study was salaries for example, this sample would be extremely versatile as we could compare wages for each gender within a given profession, between overall wages professions, and so on. Statistically speaking, the key to a stratified random sample is making sure each stratum is independent from the other so it can be considered its own separate population. This will allow us to sum up stratum variances to get an accurate population variance. For the strata themselves to be analytically useful, we want to reduce variance within them as well, meaning we create as many strata as there are variations of the subject of study to get the ideal stratified random sample. Consequently, a successful stratified random sample will have a mean variance between all possible samples that is lower than the mean variance of simple random samples from the same population. For political polling, stratified sampling is commonly used when separating populations by geography. Regions tend to be separated by city or similar boundaries and population data can be used to determine the proportional size of a given sample. The subject of interest can then be analyzed for populations both within and between the strata.

The final sampling methods covered in this paper are cluster and systematic sampling. Cluster and Systematic sampling are sampling methods developed with the practicality of carrying out a study in mind. When creating strata for a stratified random sample, it may be impractical to go out and measure the population for various reasons. Cluster sampling therefore seeks to do the opposite and create groups of maximum variance so that each measurement unit can have as much diversity as possible. For example, a study on race might stratify a region amongst neighborhoods that are predominantly homogenous and compare between them, but cluster on one extremely diverse neighborhood so that a large representative sample can be taken from just that one region area. Systematic sampling is similar but instead of relying on geographic proximity, such as a cluster of postal codes, it samples every nth member of the region or strata of interest to get an even distribution of members. The randomness in both methods comes from determining the starting point for the measurement. In cluster sampling, the postcode in the region of interest is randomly chosen, while in systematic sampling the starting house in the neighborhood is the random variable. Both these methods drift further from simple random sampling and therefore become less statistically rigorous, but what they trade in rigor they gain in practicality.

Section 2: Polling scenarios

To understand how stratified random sampling was used in political polling, we can examine the practice of random digit dialing, or RDD, in the US. Before the prevalence of mobile phone technology, people would communicate over long distances using a landline telephone located at their residence. These landlines were designated numbers that corresponded to their geographic location, and all assigned phone numbers were compiled into publicly accessible databases known as phonebooks. This was a boon for pollsters as every household had a unique geographically determined identifier associated with it, allowing them to use area codes to create strata around cities of relevance. Researchers could then dial random numbers within these area codes (hence the name) and access anyone in the country with a public phone number from one location. Assuming the person on the phone was willing to give the researcher their time, phone polling was one of the most effective means of accurately sampling populations for both political and other purposes. Assuming cost for long-distance phone calls became an issue, phonebooks also lend themselves to systematic sampling as a randomly chosen phonebook could be used to get a random number in the local area code of interest from which every nth number is dialed to create the sample.

Mobile phone technology essentially made this technique obsolete as phone numbers were no longer geographically assigned, and there was no guarantee of an individual's location based on their phone in the same way that a landline is inherently at a fixed location. As this revolution happened concurrently with the rise of the internet, polls had to evolve into the digital age. Unfortunately for pollsters, as people started having to deal with the firehose of information that is the internet, requests for one's time to respond to random questionnaires were low, if at all present on people's to-do lists, causing nonresponse rates to skyrocket in traditional polls. To combat this, panel polling was established, an example of cluster sampling. Interested individuals could sign up to be part of a survey panel where they would have to list identifying information for some reward. When someone wishes to conduct a study, they ask the panel moderators for a sample of their desired strata or cluster, and the moderator randomly chooses members of the panel that fit the criteria for the researcher to administer the survey on. While this does tackle the issue of identifying information and response rates, the survey must account for being biased towards a certain demographic of people who are internet-savvy and have a desire to be surveyed that did not stem from being asked by a researcher first.

While the above examples give simplistic examples of survey method implementation in the real world, arguably the most complex and intricate form of politically polling is the exit poll. Online surveys and RDD were examples of polls to gauge sentiment of a population *before* an election. For reasons that will be discussed in the next section, while this data is useful for organizations such as political parties trying to create a platform, it is far from gospel. Exit polls in contrast sample the people who have already voted and are leaving polling stations to ascertain the *actual* outcome of an election. In the USA, which is the primary focus of the *Oxford* book, most elections occur concurrently meaning that the entire country goes in to vote for most major elections on the same day. To effectively sample the population of a country the size of a continent, polling organizations first stratify the country into by county. Within these strata, ridings are weighted based on how close their results are predicted to be and then have a weighted random sample drawn for which location researchers will be sent to. At the voting centers, researchers will employ systematic sampling to survey every nth individual leaving the center. Many researchers may be sent to densely populated riding to collect more data in contrast to a low population riding only having a few researchers collecting smaller population samples, or nobody asking them at all. Exit polls usually ask more than just whom an individual voted for as this information will be self-evident when the votes are counted. Instead, they are also interested in why people made the decisions they made, making them an important calibration tool for future opinion polls. Exit polls in many countries also allow media organizations to turn elections into spectacles. People want to know the results of an election as soon as possible and exit poll data can be fed in real time to news broadcasters who can then present the story of an election as it happens. In the USA, pollsters are even able to verify their data with the already counted votes to confirm the accuracy of their systematic samples, meaning many news organizations receive data precise enough to call the results of an election before all the votes have even been tabulated.

Section 3: Polling challenges

The previous section discussed how the sampling methods of interest are implemented in polling research under ideal conditions, which are far from the norm. Pollsters must grapple with three main issues that reduce study quality and sometimes force methodologies to stray from random sampling. The first issue is that studies cost money. Whether discussing the need to make thousands of long-distance calls to conduct an RDD survey, or simply having to include cash incentives for online panelists to get responses, good political polls are expensive to conduct. Stratified random samples, generally considered the gold-standard of survey samples, are primarily restricted by the cost of getting surveys out to people in geographically distant regions. For example, if there is a survey trying to get a sample of 1000 people from a US state, their strata will likely be overwhelmingly comprised of people from large metropolitan areas, with few people from the rural areas. If we compound the fact that rural populations are generally geographically spread out, they become even more financially difficult to get good coverage of.

Assuming the cost per sample is known for the k strata in a potential study, the following linear combination can be created: $c = c_0 + c_1 n_1 + \ c_2 n_2 + \cdots c_k n_k$. In this equation, the total cost, c, is the sum of c_0, fixed overhead costs such as employee wages, etc. plus each of the costs per individual, c_i, for sampling n_i individuals in stratum i. This equation can be optimised to minimize c, which will modify the strata sizes to be larger for cheaper strata and smaller for costlier strata.

Consequently, polls tend to have an urban bias and treat smaller but equally diverse communities as either insignificant or one large monolithic "blob". Urban populations that are clustered together will have similar priorities and issues, such as pollution and real-estate affordability. In contrast, geographically distant communities, regardless of their size, generally won't have as much in common. For example, a small farming town may find that water rights are a contentious issue as their livelihoods depend on it, while another community built around resource mining might prioritise commodity prices and workers' rights. Combining the samples from these two communities or reducing or eliminating them altogether might make a study more financially feasible, but its sample will be inherently less representative of the population.

Compounding the cost issue is that of nonresponse rates. Even if surveys are mailed to a thousand people, many of them will either forget to respond, make a procedural error that makes their response inadmissible, or simply choose not to respond. The nonresponse rate for mail-in and online surveys was determined to be approximately 91%, meaning that under 10% of surveys are responded to. Part of the reason for this is that many people are predisposed to preferring specific survey modes, and so survey that used both online and mail-in surveys, at increased cost, saw a doubling in response rates to around 19%. This means that a mixed mode survey that wants 1000 respondents needs to survey 5000 people on average to meet their sample size. While this has the obvious impact of further ballooning costs, getting an unexpectedly small sample will exacerbate the urban bias. For example, the study with a desired sample size of 1000 might be covering a region with 40% of its population living in a major city and the other 60% spread out over 30 different communities that are geographically distant. Even with an unaltered stratified sample that receives the desired respondents, 400 people will come from one city and the 600 rural residents will be divided up into 20 members from each community. While this ideal breakdown might give just enough respondents for a decent analysis of the population, any cost-driven modifications to the stratum and above-average nonresponse will immediately reduce the quality of the study. For a rough cost estimate of how much such a study would cost, we can look at conducting a hypothetical study in the US state of Texas. To be as charitable as possible, we can assume that administering an online survey has negligible cost. Therefore, our first cost is that of mailing the surveys to random addresses within our strata. Intrastate shipping via USPS costs $9.90 per envelope, but if we want to send 5000 envelopes, let's assume we can get a bulk 50% discount of $4.95. Next, we need to add some sort of monetary compensation for the best possible chance of a response, and we set this to $5 as well. This means that from the outset, we are spending $9.95 per person. Assuming an average response rate of 20%, we mail surveys to 5000 people, bringing our costs to $49,750. Once the surveys are mailed and responses start to come in, we need actual researchers who tabulate the results and analyse them, and they will need salaries as well. The average poll worker earned $12.50 in Texas in 2022, and assuming we need around 3 individuals to manually sort through the results, digitize and clean up the data, and it takes them a week to do so working 40-hour weeks, that is $\$12.50 \times 40 \times 3 = \$1500$. Even this hypothetical best-case budget costs over $50,000; larger studies will easily have their costs balloon to significantly higher amounts.

To offset the cost of statistically rigorous studies, researchers have found an alternative method of gathering good data, which is to increase the *volume* of studies produced on a topic. Researchers take a hypothetically perfect study with a 100% response rate and set it as the benchmark for a study that will predict the population's opinions perfectly every time. They can set exit polls and election results to be this benchmark and compare pre-existing survey methodologies to this perfect result to create confidence intervals of accuracy. If, for example, an expensive study that uses stratified random sampling can correctly predict the opinions of the population with a 95% accuracy rate, we can say that it has a total survey error, TSE of $\alpha = 0.05$. If we then look at a set of less rigorous but cheaper studies that use cluster sampling, for example, which have a hypothetical TSE of $\alpha = 0.10$, in a direct comparison there seems to be no outright winner between the two. Either we lose scientific rigor or a large amount of money. However, the lack of a geographical penalty with cluster sampling can create a scenario where these studies are significantly cheaper. Therefore, if we can cost such a study at less than one-third the price of a stratified sampling study, for example if we target a diverse neighbourhood in a densely populated region, we might be able to aggregate the results of the three studies and get a *stronger* result than just the one stratified study. The math for determining the TSE of at least 2 out of 3 studies works out like so: $0.1^3 \times 0.1^2 \times 0.9 \times 3 = 0.028$. Aggregating the three less-good studies gives a better TSE of $\alpha = 0.028$ compared to just the one stratified study with $\alpha = 0.05$. This is also visible in the real world, where there are many poll aggregators that media outlets will rely on over just one poll. Aggregating studies also allows researchers to offset sampling biases arising from things like access to populations or lack of opinion diversity in an online panel. A study with a left-wing bias and one with a right-wing bias due to the sampled population may not be very reliable in isolation but combining their results may temper the biases and provide a more balanced picture of the population.

Section 4: Shortcomings in conventional sampling

Not all political polling is about determining peoples' views on an issue for the purposes of determining the result of an election. In many cases, polls are used to understand the problems people are dealing with to come up with solutions or divert government funds towards pre-existing solutions. Conventional random polling methods can come up short in such cases as many issues are those that people may be unwilling to speak about, or impact very small or hard-to-access populations. To achieve a stratified, clustered, or systematic random sample, one needs to have some prerequisite information about a population to establish pre-sampling strata, clusters, etc. If a stratum is being created based on race (as an example), this is generally a feature that is observable by just looking at a person, and even if not the case, is information people are generally comfortable sharing in free western society. Even when polling for traits that are not immediately visible, many traits such as income are ones that people are willing to share if given a reasonable expectation of privacy.

This is not always the case. For example, if a survey is attempting to study the cases of sexually transmitted diseases in the sex work industry of a country where it is either taboo or outright illegal, people will be reluctant or outright unwilling to admit their occupation on a traditional survey. Here, macro-level random sampling needs to be left behind and instead a more customised approach must be implemented to gather data. In this example, researchers may themselves have to go out into areas where sex workers ply their trade and establish a relationship of trust with individuals before they are able to as them for information for the study. If this is successful, researchers will generally request that the first person pass on information about the survey to people they know. If successful, they may get more responses through acquaintances or colleagues of the first sex worker, and thus build out a network of people who are willing and interested in giving responses to the survey over time. Variations of this method, known as network sampling, have also been developed for cases where anonymity of individuals, or plausible deniability must be maintained. For example, if a survey is trying to gather information about abuse victims who do not wish to be identified or connected to the survey for their personal safety, anonymous tips or drop-offs are generally used to pass on information to researchers who then collate the responses without any identifying information. These methods rarely garner large samples, although they are rarely sampling large populations, but most importantly, these methods take a long time as the effort required to acquire every subject is quite high.

Another scenario where random sampling struggles to adequately gather data is when the population of interest is extremely small. In these cases, stratified random sampling will simply not work as the proportional size of the stratum will be so small that the overall sample size will need to be inordinately large. Even cluster sampling which hopes to capture members of this group living in a diverse area may struggle to find such a neighbourhood. For example, when trying to sample the Amish population in the USA, a religious group who vow to live away from the influences of technology, it will be inherently impossible to use many established online survey methods as members of the Amish community will likely not be connected to the internet or even have phones. Furthermore, Amish-identifying individuals who live in major cities will inherently be different from those who live in dedicated enclaves as they will have to make compromises to coexist in a modern metropolitan region, making them an unreliable sample of all Amish people. As such, researchers will simply have to accept that there are no shortcuts in this scenario and visit Amish communities to gather their samples at high cost. This is generally true for small populations, there is usually no easy way to fit them into a wider population sample and consequently they end up being left out and having their voices not heard on matters that still affect them.

Concluding Discussion

So far, the assumptions of *Sampling* and *Oxford* about what makes a study "good", whether it be the rigorousness of the random sample, or deviation from known true results through TSE have been taken at face value, but to truly answer the research question, it is important to interrogate these assumptions from a broader point of view to see if their assumptions of goodness really are what we should be using for polling research.

To start, let's ask the question: "Why do we conduct political polling?" The *Oxford* chapters covered in this paper answer this question by focusing on the uses of political polling for understanding election results and political opinions as they directly pertain to the electoral process. This makes sense because elections give a concrete result so future polls have a target to aim towards, making their accuracy relatively straightforward to appraise. Furthermore, political polling is a key tool for parties to understand the needs of their political bases and thus can tailor their campaigns and manifestos to appeal to the largest possible demographics. Both these cases are those where profit motives are inherently present. Exit polling and research conducted to predict election results is lucrative as it can be sold to media organizations who broadcast it to the wider public. Opinion polling of large demographics can be sold to political parties and interest groups who use it to try and gain public favour.

I would contend that neither of these cases serve the overall public interest or directly seek to understand the issues faced by society to work towards alleviating them. Both the survey methodologies and the mathematics focus on gathering data from the largest groups of people as cheaply as possible, because these people have an inherently larger influence in political results, especially in the US, where many elections can be won with a plurality rather than needing a strict majority. In contrast to the wealth of political polling research on and for elections, the census, a tool for understanding a country's population to improve governance, has a decennial frequency USA and the UK, and a quintennial frequency in Canada.

Based on this conclusion about the motives of political research, we can argue that methodologies are inherently biased towards people who are the easiest to poll. Large urban populations who live in proximity and are the generally more tech-literate are the easiest to poll, and therefore these populations become major participants in polls. Researchers are obviously not blind to this and do take measures to mitigate this in their analysis, but it is an inherent quirk of the sampling methodology bias that can render entire populations invisible and politically unimportant. Rural Americans are an example of such populations who felt aggrieved by the political establishment and thus stoked the rise of politicians like Donald Trump (Thompson, D). Even if farmers, coal miners, religious evangelicals and other rural and blue-collar labourers don't share many economic, religious, or political views, they were all united under the grievance of feeling forgotten, and Trump was able to unite them to sway an election in his favour (Thompson, D). Had these people been heard, perhaps through in-depth polling of their needs, we may have a less polarised political landscape today.

Works Cited

Atkeson, Lonna Rae, and R. Michael Alvarez. *The Oxford Handbook of Polling and Survey Methods*. Oxford University Press, 2018.

Thompson, Derek. “The Deep Story of Trumpism.” *The Atlantic*, Atlantic Media Company, 29 Dec. 2020, www.theatlantic.com/ideas/archive/2020/12/deep-story-trumpism/617498/.

Thompson, Steven K. *Sampling*. John Wiley & Sons, 2012.