



The Importance of Sampling in Survey



Abstract

Why is it so difficult to get accurate polling results? Because it is extremely hard to practice Random Sampling in reality.

The definition of Random Sampling is elegant and simple: *every individual in the target population needs to have an equal chance of being included in the sample.*

But in practice, the name doesn't do it justice. Yet virtually every standard statistical theory is derived from the assumption that random sampling is adhered to.

This article explains how to adhere to random sampling as closely as possible to ensure that your market insight is representative of your target population.

The Importance of Sampling in Survey

Authored by Dr. Kelvin Goh, PhD in Physics

In an ideal world, we would want to survey every consumer to get the most accurate insight about the market.

However, in reality, that is virtually impossible. Even if you can, it will cost you more than an arm and a leg! Thankfully there is an alternative called sampling, which is an act of gathering data from a section of your target market (sample) that represents the greater population of your target market.

Sampling is the most important aspect of market research because this is source of your data which will then generate your insights. If your sample is biased, your data will be biased and your insights will be inaccurate. As the saying goes: rubbish in, rubbish out.

To ensure that your sample represents your target market as closely as possible, we need to adhere to the most important rule in sampling theory: Random Sampling. The definition of Random Sampling is simple:

every individual in the target population needs to have an equal chance of being included in the sample.

But in practice, the name doesn't do it justice. Not everyone can have equal chance of being sampled, and in some survey method, there will be people who have zero chance of being sampled, which are the people outside of the *Sampling Frame*.

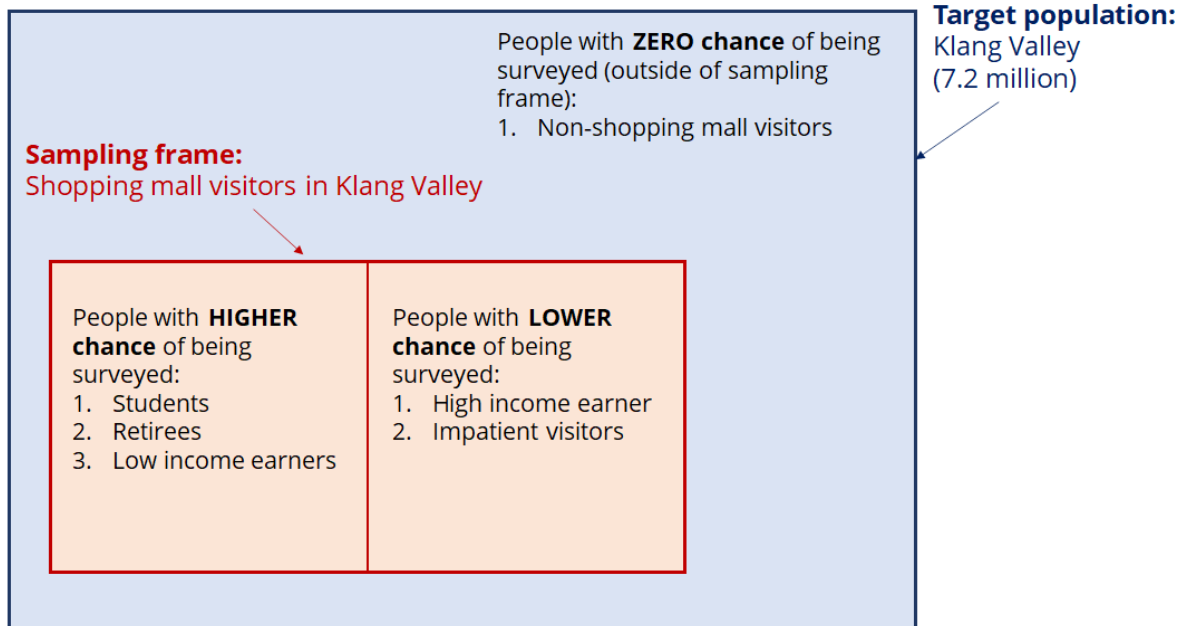
Example 1 (Face to Face survey method):

Consider a Face-to-Face (F2F) survey based on interception in all shopping malls in Klang Valley. The chances for a person to complete a survey is affected by factors such as length of the questionnaire and rewards for completing the surveys.

If the questionnaire is long and rewards aren't attractive, the people who will be attracted to complete the surveys will be biased to either people who see the reward as relatively higher in value (eg: low income earners, students), or people with a lot of free time (eg: retirees).

Conversely, if the questionnaire length is short and the rewards are attractive, then the survey completion rate will be high across all demographic groups, thus increasing the degree of Random Sampling.

Example 1: Face-to-Face survey with long questionnaires and unattractive rewards



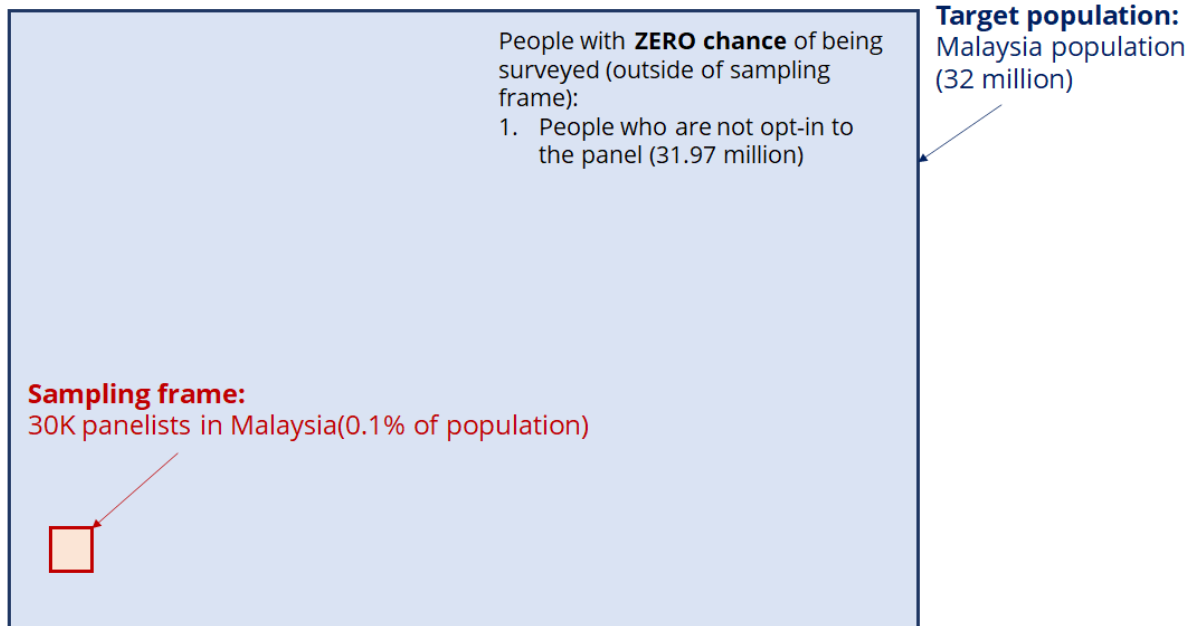
If the incidence rate of shopping mall visitors in Klang Valley (7.2 million population) is high enough (assuming 50% of the population), then the sampling frame of 50% of target population would be good enough for Random Sampling to be adhered to, assuming that factors that strongly correlates to shopping mall visits in Klang Valley has weak correlation to the topic of the study. For example, there is unlikely to be a difference in coffee consumption behaviors between coffee drinkers who visit shopping malls and coffee drinkers who don't.

However, F2F surveys are generally very expensive and takes a long time to complete. The sample size is also typically limited to no more than 2K.

Example 2 (Online panel survey method):

In the case of online panels which rely on opt-in panelists who complete surveys in exchange for rewards, their total active panelists are typically in the range of 1K to 30K. This makes the sampling frame no more than 0.3% of the population. This makes the level of biasness much higher than those in Example 1, because the odds of having factors that distinguishes 30K panelists in Malaysia from 32.34 million people outside of the sampling frame is very high. Therefore, samples drawn from sampling frame of 0.1% of the population are unlikely to adhere to Random Sampling, hence won't be representative of the population.

Example 2: Online panels



To reduce this apparent biasness, online panels often rely on stratified sampling to set quota on controllable variables (eg: Gender – 50% Male, 50% Female). However, there are only a handful of controllable variables (ie: demographics) that can be applied to benchmark against the national census. Other uncontrollable/unknown variables will still require high degree of Random Sampling to ensure the representativeness of the sample, in which online panels are lacking. Contrary to face-to-face, online surveys are much cheaper than F2F and its much faster to complete.

The Vodus Solution (One-Question Media Tracking Online Survey Method):

Ideally, we would like to have the large sampling frame of F2F combined with the low cost and speed of online surveys. Vodus is able to uniquely combine the best of both F2F and Online Panel to create an ideal survey solution which enables the sampling from a large sampling frame of 53% of the population like F2F, yet able to implement stratified sampling at a low cost and fast turnaround time like Online Panels. By using the One-Question Media Tracking Online Survey (OMTOS) method, online media consumers are only required to answer one question as they consume media content, thus encouraging high response rate across all groups of people in the sampling frame, similar to the short questionnaire scenario given in F2F Example 1. Therefore, OMTOS is able to gather samples with the highest degree of Random Sampling to generate insights which truly represents the population.

Vodus One-Question Media Tracking Online Survey (Vodus)

