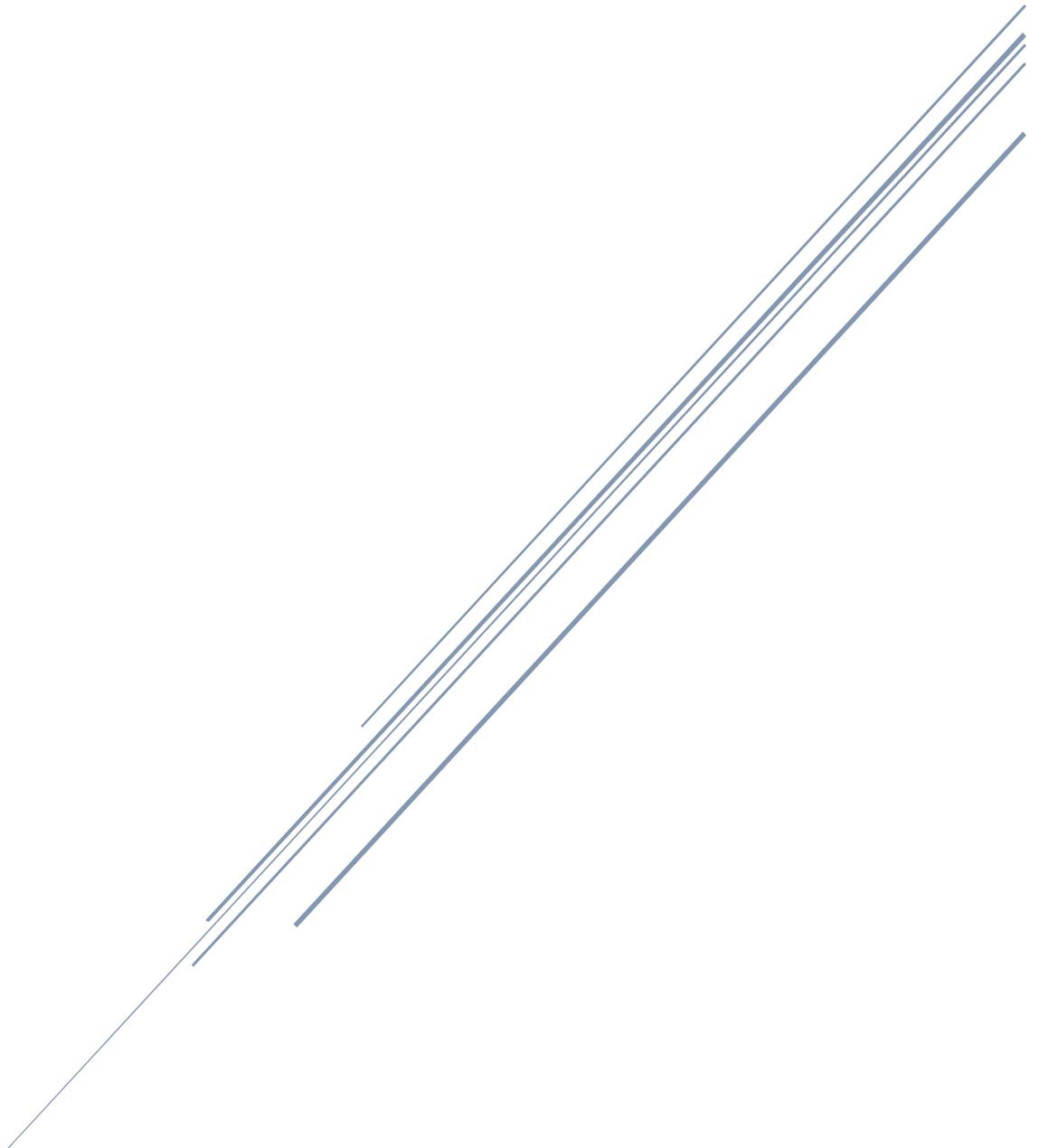


POST STRATIFICATION

The faster alternative to stratified sampling



BY DR. KELVIN GOH
CEO OF VODUS



Abstract:

As markets are evolving at a blistering pace, clients want market and consumer data to be gathered and analyzed as fast as possible. Yet the widely used *stratified sampling* method is time consuming.

To keep up with this demand, market researchers are adopting *Post-Stratification* as a faster and cheaper alternative to stratified sampling.

Learn more about *Post-Stratification* [here](#).

Post Stratification: The faster alternative to stratified sampling

Authored by Dr. Kelvin Goh, PhD in Physics

Stratified sampling is the most widely used sampling method in market research to offset the sampling bias due to having small sampling frames. However, it often requires an additional step of running an Incidence Rate Measurement to determine the representative strata of a target market (e.g.: regular smokers) prior to stratifying the sample of the intended survey. This often increases the fieldwork timeline and cost. To reduce the fieldwork time required from stratified sampling, some researchers would skip the Incidence Rate Measurement step and try to guess the demographic strata of their target market. However, this will cause the sample to be stratified wrongly, hence making the insights not representative of the target market.

There is a faster alternative to stratified sampling that has a faster fieldwork time without compromising on the representativeness of the intended target market. This alternative is *Post-Stratification*. Post-stratification is a powerful method that stratifies the data AFTER the fieldwork is completed to ensure that the data is representative of the target population. This removes the need for any pre-fieldwork assumptions on the strata of your target market. During fieldwork, both the incidence rate measurement survey and the intended survey are combined and executed in one-go. Hence, this reduces the need to run two separate fieldwork or studies, saving both time and cost.

Stratification of the data is done by applying weights onto the survey responses to ensure that the total response from each demographic group (e.g. rank-3 subgroup: Male-Chinese-Penang) matches the representative strata of the target market.

Let me give a simple illustration:

A sample has 1 Female and 2 Male, while the gender breakdown of a target market is 50-50. To ensure that the sample is representative of the target market, both Female and Male must have equal voice. By running post-stratification onto the data, the lone female will have her response weighted as 1, whereas both male's responses are weighted as 0.5.

How to does post-stratification works?

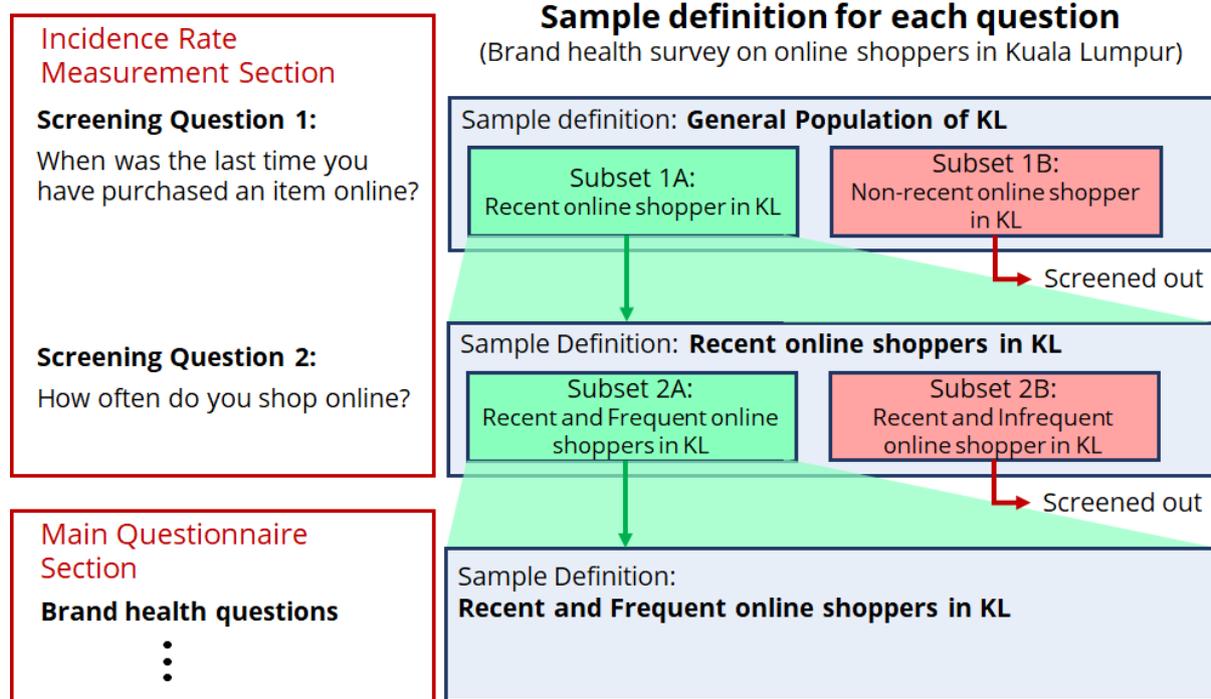
Now that you have an idea of what post-stratification is about, let me explain the step-by-step process of running a survey with post-stratification method using a brand health survey on online shoppers in Kuala Lumpur as an example.

1. Design the questionnaire and set the sample definition for each question

The questionnaire consists of two sections: *Incidence Rate Measurement* and the *Main Questionnaire Section*.

The Incidence Rate Measurement Section contains a set of screening questions that acts as a filter to screen the sample from the general population sample into the target market sample. Only those who pass the screening criteria for each screening question will continue the questionnaire. Eventually, those who passes all the screening

questions in this section will be a part of the target market sample who will proceed to answer the questions in the Main Questions Section.



2. Set the sample definition for each question

Each question in the survey may have a different sample definition due to the possibility of prior question with screener which excludes certain people from the sample. Since the survey always start with screening from the general population, the sample definition of the 1st screening question is the *General Population of KL*. After screening out non-recent online shoppers in KL, the sample of the 2nd screening question should contain only *Recent online shoppers in KL*. This process is repeated for the 3rd screening question until the sample definition of the Main Questions becomes the intended target market, which is *Recent and Frequent online shoppers*.

3. Stratify the 1st Question data to population census

Now that we have defined the sample for all our questions in the questionnaire, we can start to stratify the data from the 1st screening question to the population census. We begin by calculating the weight for all demographic group (e.g.: male, female) in a demographic profile (e.g.: gender):

$$w_s = census_s * \frac{N_{q1}}{n_s}$$

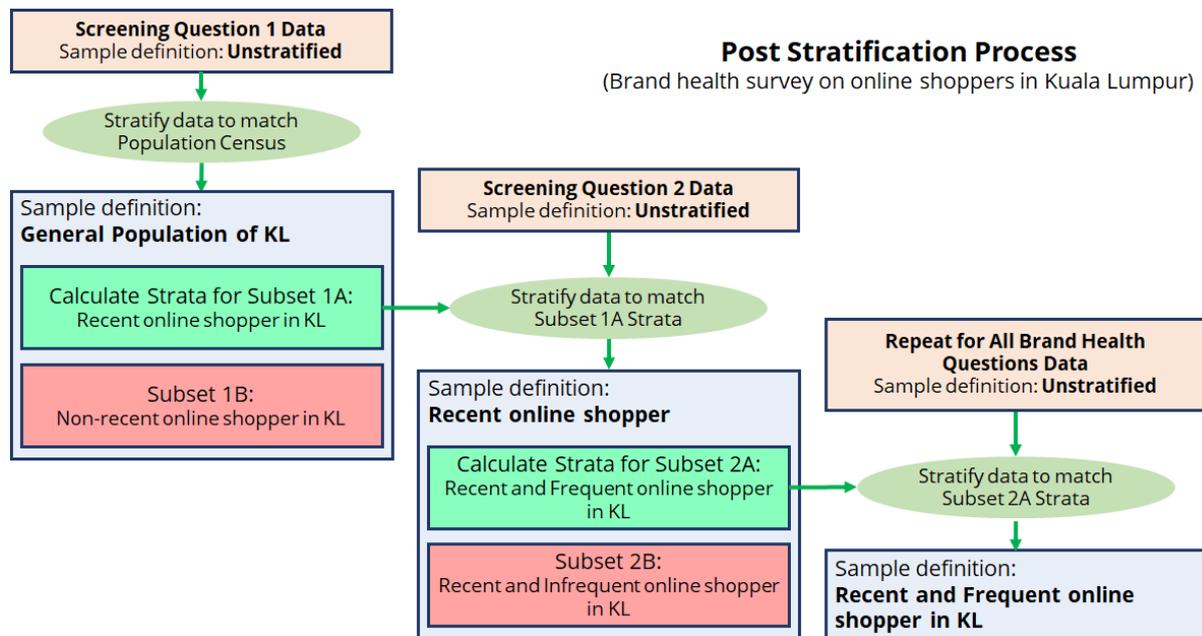
Where w_s is the weight of a demographic group s ; $census_s$ is the census strata of demographic groups; N_{q1} is the total sample size for the 1st screening question; n_s is the sample size of demographic group s for the 1st screening question. Repeat this for all demographic profiles (e.g.: Gender, Race, Age) available.

Next, stratify your data by multiplying the respective demographic group weights of every respondent to get their stratified weights $W_{respondent}$ (e.g.: $W_1 = W_{Male} * W_{Chinese} * W_{21to30}$) and assign the stratified weights to their responses. To verify that your data is stratified according to the census, sum the stratified weight of every Male in your question's sample and you should get the same value as the census.

4. Calculate the strata of the 1st Question

Now that the 1st screening question data has been stratified to be representative of the KL population, we can now calculate the Subset 1A strata of the 1st screening question, which is the demographic profile of recent online shoppers in KL.

To generate the insights from the post-stratified data, instead of counting the responses of all recent online shoppers, sum the stratified weight. For example: 50% of Males and 50% of Females in the unstratified sample are recent online shopper in KL; calculated $w_{Male} = 0.77$ and $w_{Female} = 1.43$; hence, the representative gender profile of recent online shoppers in KL is 38.5% Males and 71.5% Females.



5. Stratify the remaining question data by repeating Step 3 and 4

Repeat Step 3 to stratify the remaining question data to match the strata of their sample definition. If the question is a screening question, repeat Step 4 to generate the strata for the subsequent question to stratify against.

The above illustrates the steps involved on how Post Stratification can be used to generate a robust sample that is representative of the intended target market. In addition, this method is proven to be accurate, fast and cost efficient.

Vodus Methodology

Vodus online survey platform is purposely built to fully utilize the advantages of post-stratification to provide fast and representative insights to clients. Our online survey platform provides end-to-end DIY research services from survey scripting, to generating huge survey samples from a large sampling frame, automated data treatment which includes post-stratification to ensure that the final output are insights that are representative of your market.

For more info on how to utilize Vodus platform to generate fast and representative insights using post-stratification, please reach out at contact@vodus.com.