

SSI POV: Offline

What Is “Nat Rep” Sample?

- When researchers ask for a *nationally representative (“nat rep”)* sample, they mean that the population of interest is the entire population of the country in question and that the sample should reflect this in its structure.
- In the absence of other instructions, SSI will use age, gender and an economic wellbeing variable

The Theory of “Nat Rep”

At its best a nat rep sample will “look like” the population irrespective of how it is viewed. The numbers of men vs. women will match the national proportions, the percentage in each age group or each region will exactly match the population etc. On non-demographic measures (such as product ownership or psychographics) the sample should match the population.

To achieve this, textbook theory requires a large random sample and a high response rate to minimize systematic error and reduce the risk of unsystematic error resulting from bias. In the “real world” of marketing research, response rates are not high and differential response rates by demographic result in most pure random samples not looking like the population (however sophisticated the stratification techniques).

Quota Samples and Weighting

It is typical in marketing research therefore to use quota samples. A quota sample will be 100% guaranteed to look like the population on the demographics you choose to target. However, everything else is subject to sampling error.

Take the example of age: If the quotas are set at 16-34, 35-54, 55+ the sample will be representative within these proportions, but if analysis is done on age ranges 16-20, 21-30, 31-40, etc., there is no guarantee that the sample will still look correct. The extent to which it is possible to quota control a sample depends on the sample size and the reference data available. Six age breaks, two genders and 15 regions creates a grid of 180 cells. If the sample size is only 100 it is not possible to fill all the cells. Even with a larger sample size a cell may require only half of a person, and therefore will have no data in it.

To make a sample more representative, weighting can be used. As an alternative to interlocked cells, quota cells can be structured independently. The disadvantage here is that there may be large “holes” in the sample—if all the young people are men for example—and it won’t be possible to use weighting to correct the holes. What variables should be used to get a “nat rep” sample?

(continued)

SSI POV: Offline

The Variables for a “Nat Rep” Sample

There is no single definitive answer to what variables should be used to get a nat rep sample. It depends on the research objective and the geography. For example: age, gender, region and social class may be used in the UK; age, gender, region and ethnicity in the US; or age, gender, region and language spoken in Belgium. Age, gender and geographic region are often allied with something that differentiates by economic well-being. This could be income, education, social class or home ownership.

SSI Recommendations

In the absence of other instructions, in North America SSI will use age, gender and one additional variable which differentiates by economic well-being. (In North America, representation by geography falls naturally.) In Europe, in the absence of other instructions, SSI will use age, gender and region, since “economic well-being” data is not available from the census in all geographies.