Data Analysis and Statistical Inference
Dr. Çetinkaya-Rundel
Duke University

In-Video Quiz Questions for Unit 1: Part 1 – (3) Sampling and sources of bias

(04:19)

1. A retail store considering updates to their credit card policies randomly samples 1000 of their credit card holders to survey on the phone. The phone calls are made during business hours, therefore there is a lower rate of responses from members who work during these hours. What type of bias is this indicative of?

(a) convenience sample
(b) voluntary response
(c) non-response
(d) none of the above

(07:34)

2. A city council has requested a household survey be conducted in a suburban area of their city. The area is broken into many distinct and unique neighborhoods, some including large homes, some with only apartments. Which approach would likely be the least effective?

(a) Simple random sampling
(b) Cluster sampling, where each cluster is a neighborhood
(c) Stratified sampling, where each stratum is a neighborhood

Data Analysis and Statistical Inference
Dr. Çetinkaya-Rundel
Duke University

Answers:

1. c

Explanation: There is an initial random sample, but not everyone in this random sample is reached. Therefore the issue is non-response of the sampled individuals.

2. b

Explanation: The proposed cluster sampling method, "where each cluster is a neighborhood", will result in some neighborhoods being selected and others not being selected. Remember, in cluster sampling we divide the population into clusters and choose **a few** clusters and then sample from within those chosen clusters **only**. This can work well if the clusters are similar to each other (because then we don't lose much due to not having sampled from some of the clusters). But if the clusters are very different from each other (which is the case here, since the neighborhoods are said to be very distinct and unique) not being able to sample from some of those clusters will mean that you might lose some important information.

If we used process of elimination:

- Simple random sampling is always effective.
- Stratified sampling is more effective than cluster sampling.

Why is stratified sampling more effective than cluster sampling in this scenario? Remember that in stratified sampling the population is divided into homogenous groups (called strata), and then we sample from within **each** stratum. So, in this case, if we used stratified sampling we would sample from each neighborhood, while in cluster sampling we wouldn't get to sample from some of the neighborhoods (and those we don't get to sample from may be very different than those we do get to sample from).