

Respondent-driven sampling med viktade vänskapsband - en simuleringsstudie

Tuure Antonangeli *

Juni 2013

Sammanfattning

En metod för att göra urval från populationer som saknar urvalsram är snöbollsurval. Metoden går ut på att deltagare rekryterar individer från sitt sociala nätverk för att hitta nya individer i målgruppen, och den appliceras främst på populationer som är svåra att nå genom traditionella urvalsmetoder. Det anses dock allmänt ej vara möjligt att generera väntevärdesriktiga skattningar från ett stickprov taget med snöbollsurval, vilket exempelvis beror på att individer med stor social krets blir överrepresenterade i ett sådant stickprov. Respondent-driven sampling, RDS, är en utvidgning av snöbollsurval som kombinerar en effektiviserad och fördjupad samplingsmetodik med en matematisk modell som genom ett antal antaganden om samplingsprocessen möjliggör väntevärdesriktiga skattningar av populationsegenskaper. Ett av de antaganden som görs är att en deltagare i studien rekryterar sina bekanta i populationen med lika stor sannolikhet, vilket inte är troligt i de flesta undersökningar. Avsikten med denna studie är att undersöka hur RDS-skattningarna påverkas om en deltagare får rekrytera sina bekanta med olika stor sannolikhet, det vill säga om man tillåter relationer ha differentierade vikter. Studien utförs genom datorsimuleringar där vi undersöker andelen HIV-positiva i en fiktiv population som saknar urvalsram och har ett socialt nätverk med viktade kanter. Resultaten tyder på att RDS-skattningarna överskattar andelen HIV-positiva då vänskapsbanden är viktade och sannolikheten att vara HIV-positiv beror på summan av vikterna på ens sociala kontakter. Om sannolikheten att vara HIV-positiv däremot beror på antalet kontakter kan man inte observera några större avvikelser i RDS-skattningarna jämfört med det oviktade fallet.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige.
E-post: tuure.antonangeli@hotmail.com. Handledare: Jens Malmros.

Abstract

Snowball sampling is a method for collecting samples from a population that lacks a sampling frame, where in order to find new individuals in the target-group, participants recruit people into the study from their own social network. The method is mainly used for populations that are hard to reach with traditional sampling methods. It is however considered impossible to generate unbiased estimates from samples obtained by snowball sampling, e.g. because individuals with a large number of people in their social network are overrepresented in the study.

Respondent-Driven Sampling, RDS, is an extension of snowball sampling which combines a more efficient and advanced sampling methodology with a mathematical model, which through some assumptions about the sampling process can produce unbiased population estimates. One of the assumptions is that a participant recruits randomly from his peers in the population. This assumption is, however, implausible. The purpose of this study is to investigate how the RDS estimates are affected if a participant recruits from the social network with unequal probabilities, i.e. if the relationships have different weights.

The study is performed through computer simulation where we estimate the proportion that is HIV-positive in a fictitious population that lacks a sampling frame and has a social network with weighted edges. The results imply that RDS overestimates the proportion of HIV-positive individuals in the case where the relationships are weighted and the probability of being HIV-positive depends on the sum of the weights of the individuals social contacts. On the other hand, if the probability of being HIV-positive depends on the number of contacts, no notable difference can be observed in the RDS estimates when compared with the unweighted case.