

Metodeffekter i urvalsundersökningar där deltagarna får välja mellan pappers- och webbenkät

Martina Aksberg*

September 2010

Sammanfattning

Denna uppsats behandlar urvalsundersökningar där individerna i stickprovet själva får välja mellan att besvara undersökningens frågor via pappers- eller webbenkät. I undersökningar där svaren som inkommit genom de två insamlingsmetoderna skiljer sig signifikant från varandra har vi att göra med en så kallad metodeffekt, vilket innebär att insamlingsmetoden systematiskt påverkar de svar som ges. I uppsatsen tar vi fram förslag på hur hänsyn till metodvalet kan tas i de populationsskattningar som görs. Förslagen riktar sig till undersökningar där stickprovet har erhållits genom obundet slumpmässigt urval, stratifierat urval, klusterurval eller tvåstegsurval. I undersökningar utan bortfall eller där bortfallet är helt slumpmässigt fokuserar vi på regressions-skattningen, men ger även ett kortfattat förslag på hur hänsyn till metodvalet kan tas om skattningarna istället görs med hjälp av prediktion. Skattningar i undersökningar där bortfallet inte kan anses vara helt slumpmässigt gör vi genom att använda oss av en metod som kombinerar viktning och imputering. Våra förslag går i samtliga fall ut på att betrakta metodvalet som en bakgrundsvariabel som har inverkan på de svar som ges. Om populationsskattningar ska kunna göras med detta synsätt krävs det att vi skattar det totala antalet personer i populationen som skulle ha valt respektive insamlingsmetod om de hade blivit uttagna till stickprovet. När regressions-skattning av undersökningsvariabeln görs utgår vi från tidigare gjorda undersökningar för att skatta det totala antalet pappers- och webbundersökningar i populationen. Vid prediktions-skattning betraktar vi istället metodvalet som en undersökningsvariabel och använder oss av de metoder som finns för att skatta totalsumman av antalet personer i populationen som skulle ha valt respektive insamlingsmetod, vid de valda urvalsmetoderna. För att göra skattningar i undersökningar där bortfallet inte är helt slumpmässigt betraktar vi återigen metodvalet som en undersökningsvariabel och gör utifrån stickprovet skattningar av det totala antalet webb- och pappersundersökningar i populationen med hjälp av viktning. Skattningarna av vilka metodval som skulle ha gjorts vid en totalundersökning utan bortfall (som erhålls med hjälp av de ovan beskrivna metoderna) sätter vi sedan in i de formler som finns för att skatta populationstotalen av undersökningsvariabeln. Vi erhåller därmed populationsskattningar där hänsyn till insamlingsmetoden tas.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige.
E-post: ma.aks@hotmail.com . Handledare: Gudrun Brattström.

Abstract

This thesis deals with sample surveys in which the data is collected by giving the participants the opportunity to choose between paper-and-pencil and web-based questionnaires. In surveys where the answers received from the web respondents are significantly different from the answers received from the paper-and-pencil respondents there is a mode effect, which means that the survey mode systematically affects the answers given in the survey. In this thesis we present suggestions on how to take the choice of mode into account when estimating the population totals of the survey variables. The suggestions are applicable when the data is collected by simple random sampling, stratified sampling, cluster sampling or two-stage sampling. To make estimates in surveys without nonresponse or in surveys where data is missing completely at random (MCAR) we focus on using the regression estimator, but we also give a brief suggestion on how to make estimates when using the prediction approach. In surveys where data is missing at random (MAR) we use a method that combines weighting and imputation to make population estimates. The idea behind our suggestions is to consider the choice of mode to be a background variable that affects the answers given in the survey (i.e. the survey variables). To make population estimates (of the survey variables) from this point of view it is necessary to estimate the number of people in the whole population that would have chosen each of the two collection modes if they had been in the sample. When the regression estimator is used we estimate the total number of web and paper-and-pencil participants in the population by using data from former surveys. To make estimates when the prediction approach is used we analyze the data received in the survey by considering the choice of mode to be a survey variable. In surveys where data is missing at random (MAR) we again consider the choice of mode to be a survey variable, that is estimated for the whole population by weighting. The estimates of the total number of people that would have chosen each of the two collection modes in a census without nonresponse are put into the the formulas that estimate the population totals of the survey variables. Thus we receive population estimates that take the choice of mode into account.