

Varude imputeerimine Eesti Äriregistris 2011. aastal

Bakalaureusetöö

Cliona Georgia Dalberg

Kokkuvõte

Bakalaureusetöö tehti projekti “Metoodika väljatöötamine statistika tegemiseks kombineeritud administratiivsete andmeallikate ja uuringute andmete baasil” raames, mille tellis Eesti Statistikaameti käest Eurostat. Töös kasutati Äriregistrist pärit majandusaasta aruannete andmeid.

Bakalaureusetöö eesmärgiks oli imputeerida puuduvad väärtused varude osas 2011. aastal.

Töö esimeses osas anti lühiülevaade puudumisest ja selle mehhanismidest ning omadustest. Veel käsitleti mõningaid mitmese imputeerimise meetodeid nagu monotoonne regressioon, Monte Carlo Markovi ahelatega ja järjestikune regressioon.

Töö teises pooles räägiti läbiviidud simulatsioonist, kus võeti aluseks andmestikud, mis täitsid nõutud tingimusi. Neis andmestikes seati puuduvaks vastavalt ligikaudu 10%, 30% või 50% andmetest. Järgnes järjestikuse regressiooni rakendamine ning saadud tulemuste keskmiste võrdlemine. Lisaks vaadeldi simulatsiooni aspektist, et milline tunnuse tüüp varude komponentidel lisaaruandes annaks korrektsema tulemuse. Seejärel tegeleti andmestiku korrastamisega, eemaldades mittehuvipakkuvad vaatlused ning asendades puuduvad väärtused olemasolevate andmetega varasemast aastast. Järgnevalt püüti varemkirjeldatud meetodeid rakendada. MCMC teostati SAS Enterprise Guide abil ning järjestikust regressiooni praktiseeriti makro IVEwarega, viimane osutus ainsana tulemuslikuks variandiks. Seejärel leiti mitmeid koefitsiente, et tagada andmestikus nõutud tunnustevahelised seosed.

Andmetöötlus tehti SAS Enterprise Guide’i ning SAS 9.2-ga, töö kirjutati Microsoft Word 2007-ga.