

Mathematical Statistics Stockholm University Bachelor Thesis **2020:21** http://www.math.su.se

Approaches for manipulating censored data

Marc Roddis*

September 2020

Abstract

Data that is too uncertain to be reported as specific numbers is reported as censored. Censored data is generally very common from chemical analyses, so the manipulation of censored data plays a key role in statistical analysis of such data. The Swedish National Monitoring Programme for Contaminants (SNMPC) manipulates censored data by substitution; our main idea is to use imputation by censored regression instead. We show that this idea is viable and has relevance to real SNMPC data. We present its mathematical basis; formulate conjectures; outline the design, purpose and implementation of our simulation experiments; and report and discuss our experimental results. In our main experiments, substitution and imputation by censored regression are used to generate distinct manipulated datasets. From these datasets, estimates and predictions and their variance, squared-bias, and MSE are computed and reported. Our main finding is that imputation by censored regression generally results in much lower squared-bias than results from substitution.

^{*}Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: marc.roddis@gmail.com. Supervisor: Martin Sköld.