

# Statistical Models for Defective Count Data

G. Neubauer<sup>1</sup>, G. Djuraš<sup>1</sup>, H. Friedl<sup>2</sup>

<sup>1</sup>Statistical Applications, Joanneum Research

<sup>2</sup>Institute of Statistics, University of Technology  
Graz, Austria

**Abstract:** Count data are important in many fields of application: life sciences, administration, production and so on. Any structure or system designed to collect and record the occurrence of a certain event may make errors. If events that occur are not recognized by the system we speak under-reporting. And events that are reported although they never occurred cause over-reporting. The most prominent example of under-reporting are crime data: it is well known and widely accepted that official crime figures are lower than the actual number of crimes. On the other hand over-reporting is less considered, but insurance companies suspect that not every claim for a stolen good is based on theft. Typical goods are bicycles or skiing equipment. For health data like cause of death, wrong counts may occur due to errors in the diagnoses. Any misclassification will cause two errors in the counting system. An under-reported count in the true category, and an over-reported count in the actual category. Hence considering both kinds of error is of great importance when modelling count data. We present an overview on models for pure underreporting that are based on randomized binomial sampling schemes. Further we show how these models can be extended to cover over-reporting and finally give the results of application to real data.