

# Informative cluster size problems

Jaakko Nevalainen<sup>1</sup>, Somnath Datta<sup>2</sup>, and Hannu Oja<sup>3</sup>

<sup>1</sup>*University of Turku, Finland*

<sup>2</sup>*University of Louisville, USA*

<sup>3</sup>*University of Tampere, Finland*

## Abstract

In spite of many recent contributions to the literature informative (or nonignorable) cluster size settings are still not well known and understood. This is partly explained by the development of the theory to specific problems and mainly directly in the framework of modeling. The aim of this talk is to serve as a tutorial to these problems. We define the problem, describe it from different viewpoints in light of examples from the literature, and discuss different data generating mechanisms. We show how the classical statistical procedures (*t*-test, correlation, nonparametric location functionals,...) need to be modified in order to draw right conclusions efficiently from data with informative cluster size

## Keywords

Clustered data, Informative cluster size.