

Introduction to statistical inference 1

Lauri Mehtätalo
University of Eastern Finland
School of Computing

January 4, 2018

Contents

1	Preliminaries	1
----------	----------------------	----------

Chapter 1

Preliminaries

- Field of statistics builds on probability theory

“You can, for example, never foretell what any one man will do, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant. So says the statistician.” - Sherlock Holmes

- The paragraph includes the important ideas of the statistical model:
 - The percentage p = the model of the process or underlying population
 - The behavior of individuals = data
- Assuming a constant probability p may be a too simplistic or naive assumption, and may be replaced by more realistic one where p is a function of the probabilities of the individual and the context where s/he is.
- We also need to specify a model, for the variability of the individuals around the p to complete the model formulation.
 - A crude summary if the variance-covariance matrix of the observations.
 - A complete definition is done by specifying the joint distribution of all individuals.
- We also may want to estimate how accurately we finally estimated p writing the available data
- The theoretical process that generates the data is called
 - Statistical model or (tilastollinen malli)
 - Stochastic process or (statistinen prosessi)

- Random process or (satunnaisprosessi)
- The process is random/stochastic because the “man” do not behave exactly according to model.¹
 - Probability calculus and the theory of random variables provide tools to formulate and understand such models.
- Once model has been formulated or specified (muotoiltu), observed data can be used to²
 - estimate model parameters
 - evaluate the model fit (mallin sopivuus)
 - evaluate the inaccuracy related to the estimated model parameters
- When talking about models, we can talk about
 - True model (Tosi malli)
 - Estimated model (Estimoitu malli)
 - True model always stays the same, but as data used to formulate the estimated model gets larger, the estimated model gets closer to true model.
 - See example R-script `regsimu.R`

¹This is what is done on this part of the course (ISI1).

²This is what is done on second part of the course (ISI2)