

MTAT .08.043

Transportation Theory and Applications

Lecture II: Data and Sampling

A. Hadachi

Definitions

- ❖ Sample: a selected set of units that represents a larger population with certain attributes of interest.
- ❖ Sampling methods
 - ❖ Simple random sampling
 - ❖ Stratified random sampling
 - ❖ Choice-based sampling

Definition

- ❖ Sampling Error:
 - ❖ Possibility that the sample does not reflect the population (always present due the randomness in selecting the sample and has an impact on the variability only)
- ❖ Sampling bias:
 - ❖ consequence of a mistake in selection the sampling method or in defining population of interest.

Definition

- ❖ Sample size:
 - ❖ There is no straightforward answer to the calculation of the sample size.
 - ❖ (formula for computing sample size involves subjective parameters)
 - ❖ Sample size should be too large or too small

The Central Limit Theorem

- ❖ The estimates mean from a sample tend to become a normally distributed as the sample size increases.

Sample Extraction

❖ Example 2.1:

Consider a certain area the population of which may be classified in groups according to: automobile ownership (with and without a car); and household size (up to four and more than four residents).

Let us assume that m observations are required by cell in order to guarantee a 95% confidence level in the estimation of, say, trip rates; assume also that the population can be considered to have approximately the following distribution (i.e. from historic data).

Car ownership	Household size	% of population
With car	Four or less	9
	More than four	16
Without car	Four or less	25
	More than four	50

Sampling Problem



Check lecture note 1

Example 2.2

Assume that for the purposes of a transport study the population of a certain area has been classified according to two income categories, and that there are only two modes of transport available (car and bus) for the journey to work. Let us also assume that the population distribution is given by:

	Low income	High income	Total
Bus user	0.45	0.15	0.60
Car user	0.20	0.20	0.40
Total	0.65	0.35	1.00

- ❖ Let's apply previous three sampling techniques presented
- ❖ Check lecture note 2

Errors in modelling and forecasting

- ❖ Different type of errors
 - ❖ Measurement errors
 - ❖ Sampling Errors
 - ❖ Computational Errors
 - ❖ Specification Errors
 - ❖ Transfer Errors
 - ❖ Aggregation Errors

Errors in modelling and forecasting

- ❖ Different type of errors
 - ❖ Measurement errors

occurs due to the inaccuracies in measuring the data.

Such as: sensors' calibration not correct or defective, questions badly registered by the interviewees incase of

Errors in modelling and forecasting

- ❖ Different type of errors
 - ❖ Sampling Errors

this error rises because of the models used for
sampling

Errors in modelling and forecasting

- ❖ Different type of errors
 - ❖ Computational Errors

occurs during the process of running different iterations of the models

Errors in modelling and forecasting

- ❖ Different type of errors

- ❖ Specification Errors

this due the way the model was constructed and
it can be due to:

- ❖ Inclusion of irrelevant variable
 - ❖ Omission of relevant variable
 - ❖ etc .

Errors in modelling and forecasting

- ❖ Different type of errors
 - ❖ Transfer Errors

occurs when the model is developed in one context e.g time and or location and it is applied on a different one.

Errors in modelling and forecasting

- ❖ Different type of errors
 - ❖ Aggregation Errors

this case occurs when you are trying to model a group and actually you need to model at the level of individuals.

The model Complexity vs Data

❖ Check lecture note 3