STAB22 TUT21

Chong Chen
University of Toronto, Scarborough
Department of Computer and Mathematical Sciences

January 18, 2019

1   Basic information

Me: Chong Chen
Email: chong.chen@mail.utoronto.ca
Office hours: TBA

2   Data

"Five W's":Who, What, When, Where, and (if possible) Why.

2.1   Why

What purpose do the data have?

2.2   Who

What individuals do the data describe? How many individuals appear in the data?

2.3   What

How many variables do the data contain? What are the exact definitions of these variables? What are the units of measurements in which each variable is recorded?

2.4 Example

The Ministry of Education in Ontario requires local school districts to keep the following records on all students: age, ethnicity, days absent, current grade level, standardized test scores in reading and mathematics, and any disabilities or special needs. Determine the W's.

A.

　Who: the Ministry of Education;
　What: local school districts;
　When: information not provided;
　Where: information not provided;
　Why: required by the Ministry of Education in Ontario

B.

　Who: students in local school districts;
　What: age, ethnicity, days absent, current grade level, standardized test scores in reading and mathematics, and any disabilities or special needs;
　When: not specified;
　Where: Ontario;
　Why: required by the Ministry of Education in Ontario

C.

　Who: the Ministry of Education in Ontario;
　What: age, ethnicity, days absent, current grade level, standardized test scores in reading and mathematics, and any disabilities or special needs;
　When: not specified;
　Where: Ontario;
　Why: required by the Ministry of Education in Ontario

3 Variables

Def: A variable is any characteristic of an individual.
Mainly, they are two type of variables, categorical and quantitative.

3.1 Categorical variable

A variable that places an individual into one of several groups or categories.

3.2 Quantitative variable

A variable that takes numerical values for which arithmetic operations are defined.

3.3 Example

A consumer report article evaluating refrigerators listed 41 models, giving the brand, cost, type (top freezer, bottom freezer, side-by-side, French door), overall rating (excellent, good, average, unsatisfactory), and repair history for this brand (percentage requiring repairs over the last five years). Which of the following statements is true?

A.
  Model is an identifier; type is a categorical nominal variable; overall rating is a categorical ordinal variable, repair history for this brand is a categorical ordinal variable

B.
  Model is an identifier; type is a categorical nominal variable; overall rating is a categorical nominal variable, repair history for this brand is a quantitative variable

C.
  Model is an identifier; type is a categorical nominal variable; overall rating is a categorical ordinal variable, repair history for this brand is a quantitative variable