Software Measurement

Measurement is essential for a science.

Measurement

“the process by which numbers or symbols are assigned to attributes of entities in the real world in such a way as to describe them according to clearly defined rules”

Uses of Measurements

• assessment
• prediction
  – mathematical model
  – inherently stochastic

Types of Measurements

• Process measures
  – Identify essential characteristics of process
  – e.g. GQM - Goal - Question - Measure - if we want to improve customer satisfaction, measure number of defect reports
• Product measures
  – identify essential characteristics of document
  – e.g. LOC, McCabe’s Cyclomatic Number

Properties of Measures

• Monotonicity
• Scale Types
  – nominal
  – ordinal
  – interval
  – ratio
  – absolute

LOC

Why isn’t the Lines of Code measure very useful?

• It is not an essential characteristic of software
• It is not very useful as a predictor
Height and weight

- Why are height and weight useful in measuring people?

Using Measures

- Can we take the average of heights, temps, grades, shoe sizes, jersey numbers?
- Can we add two heights, temps, grades, jersey numbers, shoe sizes?
- Can we compare ratios of heights, temps, grades, jersey numbers, shoe sizes?

Software Measurement

Measurement Theory

- circa 1900 - applied to physics
- 1940’s - applied to psychology, sociology
- 1990’s - applied to software measurement

Representational TOM

- empirical relation system
  - (C,R)
- numerical relation system
  - (N,P)
- M maps (C,R) to (N,P)
- representation condition
  - \( x < y \implies M(x) < M(y) \)

Empirical

- A set of entities, E
- A set of relationships, R
  - often “less than” or “less than or equal”
  - note that not everything has to be related
Numerical

- A set of entities
  - also called the “answer set”
  - usually numbers - natural numbers, integers or reals
- A set of relations
  - usually already exists
  - often “less than” or “less than or equal”

The Mapping

- The representation condition
  - $M(x) \text{ rel } M(y)$ if $x \text{ rel } y$
  - $x \text{ rel } y$ iff $M(x) \text{ rel } M(y)$
- Both have been used by classical measurement theory authors
- I prefer the first definition

Height

\[ \text{Height} \]

\[ 6' 0'' \]

Representation Condition

- If Fred is shorter than or equal to Bill, then Fred’s height is less than or equal to Bill’s height
- if $x \leq y$, then $m(x) \leq m(y)$

Classic Example

- McCabe’s Cyclomatic Number
- Empirical - Control Flow Graph
- Numerical (Answer Set) - integers
- mapping - E-N+2

“BIG” example

- We want to measure people as BIG
  - i.e. height and weight
- Empirically, if two people are the same height, the heavier is bigger
  - if two people are the same weight, the taller is bigger
- Answer set is real or tuple?
Statistics

- The scale type implies usable statistics
  - Mean of values – at least interval
  - Ratios between values – at least ratio

Piles of blocks

Average of measures

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>3</th>
<th>4</th>
<th>4 ave</th>
</tr>
</thead>
<tbody>
<tr>
<td>3x+2</td>
<td>5</td>
<td>11</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>x^2</td>
<td>1</td>
<td>9</td>
<td>16</td>
<td>16</td>
</tr>
</tbody>
</table>