The Basics of Measurements

SE 611

Representation Theory of Measurement

- The *representational theory of measurement* seeks to formalize our intuition about the way the world works.
- The core of this theory is that *measurement* is a process of assigning *numbers* to *attributes* or *characteristics* of the empirical world in such a way that the relevant *qualitative empirical relations* among these *attributes* or *characteristics* are reflected in the numbers themselves.
- According to representational theory, *measurement* is possible only because the empirical system represented and the numerical system representing it, possess the same mathematical structure.

Empirical Relations

- We perceive the real world by comparing things. Not always by assigning number to them.
- For example, *taller than* is a binary relation defined on the set of pairs of people. Given any two people, *x* and *y*, we can observe that
    - *x* is taller than *y*, or
    - *y* is taller than *x*

    Therefore *taller than* is an **empirical relation**
- We define *measurement* as the mapping from the empirical world to the formal, relational world. Consequently, a *measure* is the number or symbol assigned to an *entity* by this mapping in order to characterize an attribute.

Empirical Relations (cont..)

Sampling of 100 users to express preference among product A,B,C, and D (pairwise)

	More Functionality				More User-Friendly			
	A	B	C	D	A	B	C	D
A	—	80	10	80	—	45	50	44
B	20	—	5	50	55	—	52	50
C	90	95	—	96	50	48	—	51
D	20	50	4	—	54	50	49	—

Subjective Rating Formats of E.R.

Likert Scale

- Give the respondent a statement with which to agree or disagree.
- Example: This software program is reliable

Strongly Agree	Agree	Neither agree nor disagree	Disagree	Strongly Disagree

Subjective Rating Formats of E.R.

Forced Ranking

- Give n alternatives, ordered from 1 (best) to *n* (worst).
- Example: Rank the following five software modules in order of maintenance difficulty

  with 1 = least complex, 5 = most complex:

—	Module A
—	Module B
—	Module C
—	Module D
—	Module E

Subjective Rating Formats of E.R.

Verbal Frequency Scale

- Example: How often does this program fail?

Always	Often	Sometimes	Seldom	Never

Subjective Rating Formats of E.R.

Ordinal Scale

- List several ordered alternatives and have respondent select one.
- Example: How often the software fail?
  - 1. Hourly
  - 2. Daily
  - 3. Weekly
  - 4. Monthly
  - 5. Several times a year
  - 6. Once or twice a year
  - 7. Never

Subjective Rating Formats of E.R.

Comparative Scale

- Compare at least two entities and assign a numeric value

Very superior			About the same				Very inferior
1	2	3	4	5	6	7	8

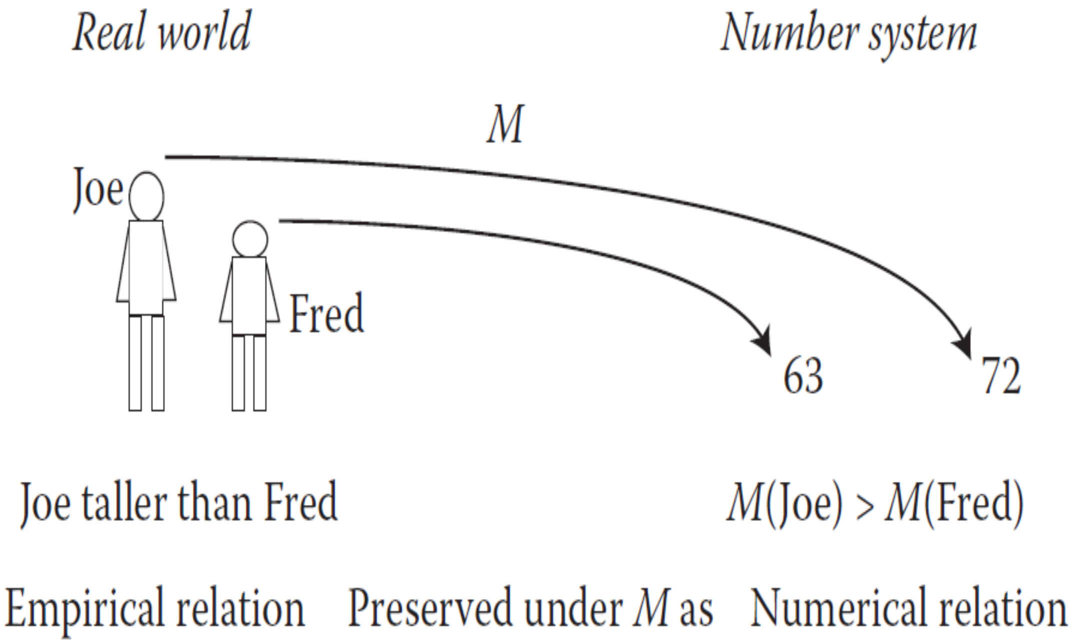
Subjective Rating Formats of E.R.

Numerical Scale

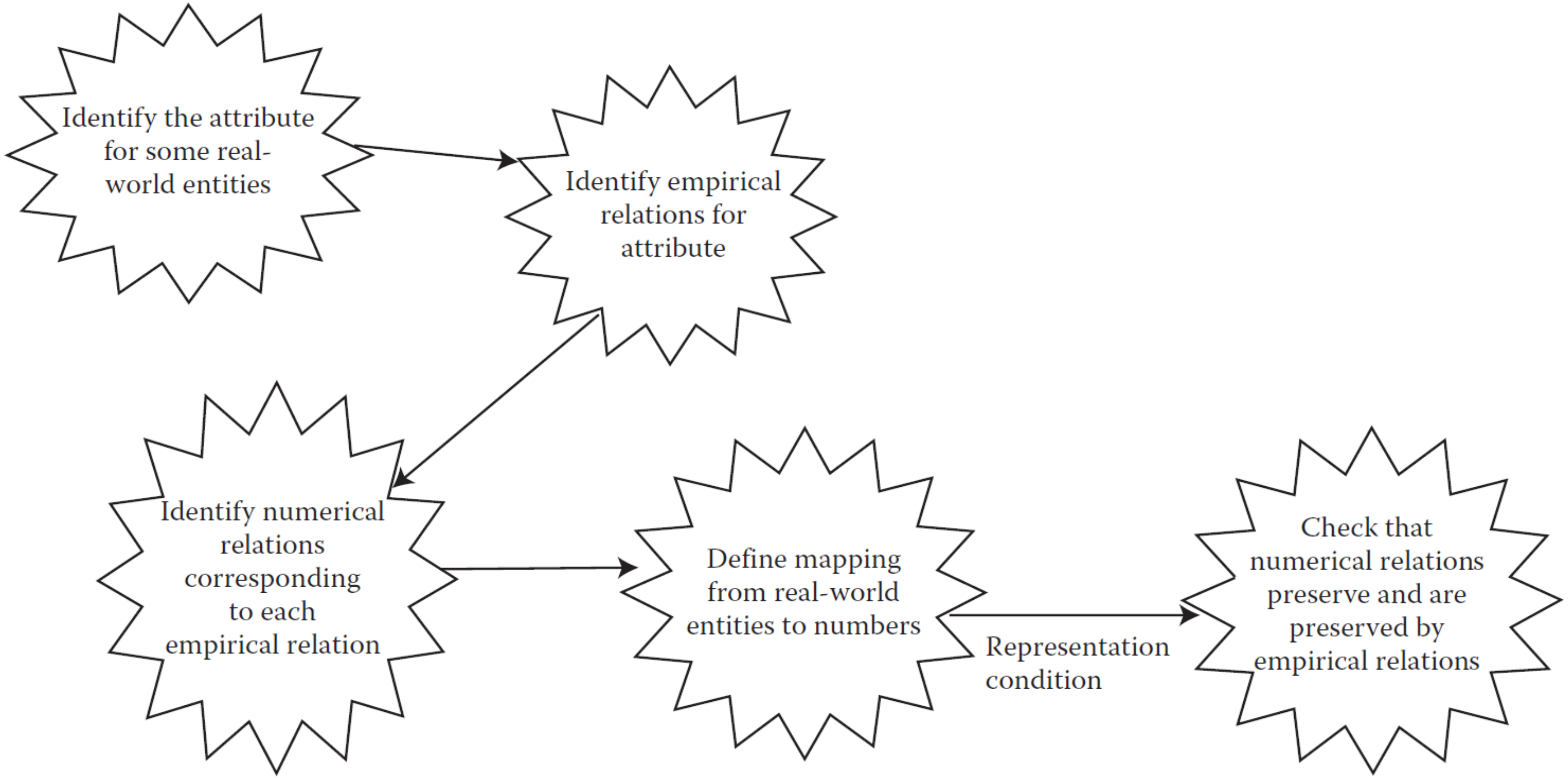
Unimportant							Important
1	2	3	4	5	6	7	8

The Representation Condition of Measurement

- The real world is the domain and the mathematical world is the range
- Each relation in the empirical relational system corresponds via the measurement to an element in a number system
- We want the mapping to preserve the relation. This rule is called the representation condition (see figure)
- The mapping we call a measure is sometimes called a *representation* or *homomorphism*



Key Stages of Formal Measurement



Some Specific Measurements in Software

Entity	Attribute	Measure
Completed project	Duration	Months from start to finish
Completed project	Duration	Days from start to finish
Program code	Length	Number of lines of code (LOC)
Program code	Length	Number of executable statements
Integration testing process	Duration	Hours from start to finish
Integration testing process	Rate at which faults are found	Number of faults found per KLOC (thousand LOC)
Test set	Efficiency	Number of faults found per number of test cases
Test set	Effectiveness	Number of faults found per KLOC (thousand LOC)
Program code	Reliability	Mean time to failure (MTTF) in CPU hours
Program code	Reliability	Rate of occurrence of failures (ROCOF) in CPU hours

- *There is nothing wrong with using the same representation in different ways, or using several representation for the same attribute.*

Measurement and Model

- A model is an abstraction of reality. It can view an entity or concept from a particular perspective

- Models come in many different forms: as equations, mapping, or diagram for instance

- For example, to measure the length of a program using LOC, we need a model of a program which would specify how a program differs from a subroutine, whether or not to treat separate statements on the same line as distinct LOC, whether or not to count comment lines, data declaration, etc.

Direct and Derived Measurement

- *Direct measurement* of an attribute of an entity involves no other attribute or entity. For example, *length* of a physical object can be measured without reference to any other object or attribute. Commonly user direct measure in SE:
    - E.g. *Size* of source code (measured by LOC)
    - *Schedule* of the testing process (measured by elapsed time in hours)
    - *Number of defects* discovered
    - *Time* a programmer spends on a project (measure by months worked)
- *Derived Measures* can be a combination of *direct measures*. It is often useful in making visible the interactions between direct measures.

Example of Derived Measures

Programmer productivity	LOC produced/person-months of effort
Module defect density	Number of defects/module size
Defect detection efficiency	Number of defects detected/total number of defects
Requirements stability	Number of initial requirements/total number of requirements
Test coverage	Number of test requirements covered/total number of test requirements
System spoilage	Effort spent fixing faults/total project effort

Measurement Scale and Scale Types

- Three important questions concerning representation and scales:

  - How do we determine when one numerical relation system is preferable to another? (*The answer is pragmatic)*
  - How do we know if a particular empirical relation system has a representation in a given numerical relation system? *(A representation problem)*
  - **What do we do when we have several different possible representation in the same numerical relation system? *(The uniqueness problem)***

Measurement Scale and Scale Types

- Five major types of measurement scale

1. *Nominal*
2. *Ordinal*
3. *Interval*
4. *Ratio*
5. *Absolute*

Nominal Scale Type

Nominal scale measurement places elements in a classification scheme. The classes are not ordered. It has two major characteristics:

- 1. The empirical relation system consists only of different classes; there is no notion of ordering among the classes.
- 2. Any distinct numbering or symbolic representation of the classes is an acceptable measure

Example: *Suppose we are investigating the set of all known software faults.*

$$M_1(x) = \begin{cases} 1, & \text{if } x \text{ is specification fault} \\ 2, & \text{if } x \text{ is design fault} \\ 3, & \text{if } x \text{ is code fault} \end{cases}$$

$$M_2(x) = \begin{cases} 101, & \text{if } x \text{ is specification fault} \\ 2.73, & \text{if } x \text{ is design fault} \\ 69, & \text{if } x \text{ is code fault} \end{cases}$$

Ordinal Scale Type

We can often augment the nominal scale with information about an ordering of the classes or categories creating an ordinal scale. The ordering leads to analysis not possible with nominal measures. The ordinal scale has the following characteristics:

- The empirical relation system consists of classes that are ordered with respect to the attribute.
- Any mapping that preserves the ordering (i.e., any monotonic function) is acceptable.
- The numbers represent ranking only, so addition, subtraction, and other arithmetic operations have no meaning.

Ordinal Scale Type Example

Suppose we want to capture the attribute *Complexity* of a software *X*, quantitatively:

$$M_1(x) = \begin{cases} 1 \text{ if } x \text{ is trivial} \\ 2 \text{ if } x \text{ is simple} \\ 3 \text{ if } x \text{ is moderate} \\ 4 \text{ if } x \text{ is complex} \\ 5 \text{ if } x \text{ is incomprehensible} \end{cases} \qquad M_2(x) = \begin{cases} 1 \text{ if } x \text{ is trivial} \\ 2 \text{ if } x \text{ is simple} \\ 3 \text{ if } x \text{ is moderate} \\ 4 \text{ if } x \text{ is complex} \\ 10 \text{ if } x \text{ is incomprehensible} \end{cases}$$

$$M_3(x) = \begin{cases} 0.1 \text{ if } x \text{ is trivial} \\ 1001 \text{ if } x \text{ is simple} \\ 1002 \text{ if } x \text{ is moderate} \\ 4570 \text{ if } x \text{ is complex} \\ 4573 \text{ if } x \text{ is incomprehensible} \end{cases}$$

However, neither M_4 nor M_5 is valid:

$$M_4(x) = \begin{cases} 1 \text{ if } x \text{ is trivial} \\ 1 \text{ if } x \text{ is simple} \\ 3 \text{ if } x \text{ is moderate} \\ 4 \text{ if } x \text{ is complex} \\ 5 \text{ if } x \text{ is incomprehensible} \end{cases} \qquad M_5(x) = \begin{cases} 1 \text{ if } x \text{ is trivial} \\ 3 \text{ if } x \text{ is simple} \\ 2 \text{ if } x \text{ is moderate} \\ 4 \text{ if } x \text{ is complex} \\ 10 \text{ if } x \text{ is incomprehensible} \end{cases}$$

Interval Scale Type

The interval scale carries more information and it more powerful than nominal or ordinal. This scale captures information about the size of the intervals that separate the classes, so that we can in some sense understand the size of the jump from one class to another.

- ❑An interval scale preserves order, as with an ordinal scale.
- ❑An interval scale preserves differences but not ratios. That is, we know the difference between any two of the ordered classes in the range of the mapping, but computing the ratio of two classes in the range does not make sense.
- ❑Addition and subtraction are acceptable on the interval scale, but not multiplication and division

Example:

$$M_1(x) = \begin{cases} 1 \text{ if } x \text{ is trivial} \\ 2 \text{ if } x \text{ is simple} \\ 3 \text{ if } x \text{ is moderate} \\ 4 \text{ if } x \text{ is complex} \\ 5 \text{ if } x \text{ is incomprehensible} \end{cases} \qquad M_2(x) = \begin{cases} 0 \text{ if } x \text{ is trivial} \\ 2 \text{ if } x \text{ is simple} \\ 4 \text{ if } x \text{ is moderate} \\ 6 \text{ if } x \text{ is complex} \\ 8 \text{ if } x \text{ is incomprehensible} \end{cases}$$

$$M_3(x) = \begin{cases} 3.1 \text{ if } x \text{ is trivial} \\ 5.1 \text{ if } x \text{ is simple} \\ 7.1 \text{ if } x \text{ is moderate} \\ 9.1 \text{ if } x \text{ is complex} \\ 11.1 \text{ if } x \text{ is incomprehensible} \end{cases}$$

Ratio Scale Type

- It is a measurement mapping that preserves ordering, the size of intervals between entities, and ratios between entities.
- There is a zero element, representing total lack of the attribute.
- The measurement mapping must start at zero and increase at equal intervals, known as units.
- All arithmetic can be meaningfully applied to the classes in the range of the mapping.
- In general, any acceptable transformation for a ratio scale is a mapping of the form

$$M = aM'$$

Absolute Scale Type

The absolute scale is the most restrictive of all scale types. For any two measures, *M* and *M′*, there is only one admissible transformation: *the identity transformation*.

That is, there is only one way in which the measurement can be made. The *absolute scale* has the following properties:

- The measurement for an absolute scale is made simply by counting the number of elements in the entity set.
- The attribute always takes the form "number of occurrences of *x* in the entity."
- There is only one possible measurement mapping, namely the actual count, and there is only one way to count elements.
- All arithmetic analysis of the resulting count is meaningful.

Example:

- Number of failures observed during integration testing can only measured in one way: by counting the failures
- Number of people working in a project can be only measured by counting people

Summary of Scales

Scale Type	Admissible Transformations (How Measures M and M' must be Related)	Examples
Nominal	1–1 mapping from M to M'	Labeling, classifying entities
Ordinal	Monotonic increasing function from M to M', that is, $M(x)$ $M(y)$ implies $M'(x)$ $M'(y)$	Preference, hardness, air quality, intelligence tests (raw scores)
Interval	$M' = aM + b$ $(a > 0)$	Relative time, temperature (Fahrenheit, Celsius), intelligence tests (standardized scores)
Ratio	$M' = aM$ $(a > 0)$	Time interval, length, temperature (Kelvin)
Absolute	$M' = M$	Counting entities

Meaningfulness in Measurement

- The type of scale determine what kind of analysis we can perform on the measurements
- Consider these statements:
  1. The # of errors discovered during integration testing was 100.
  2. The cost of fixing each error is at least 100.
  3. A semantic error takes twice as long to fix than an syntactic error.
  4. A semantic error is twice as complex as a syntactic error.

- **A measurement statement is meaningful if its truth value is invariant of *admissible transformations* of the scale**

Statistical Operations on Measures

- The scale type of a measure affects the types of operations and statistical analyses
- Consider we have measured an attribute for 13 entities, and the resulting data points in ranked order are:

  2, 2, 4, 5, 5, 8, 8, 10,11, 11, 11, 15, 16

  - The *mean* of this set of data (i.e., the sum divided by the number of items) is *8.3*.
  - The *median* (i.e., the value of the middle-ranked item) is *8*.
  - The *mode* (i.e., the value of the most commonly occurring item) is *11*.

End of Chapter 2