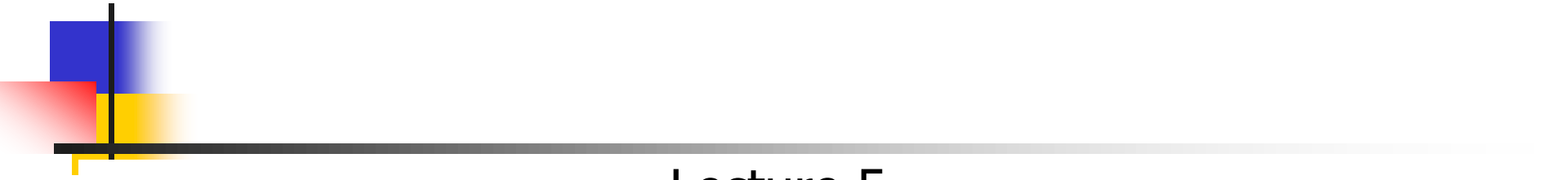
SOEN 384
Management, Measurement and Quality Control

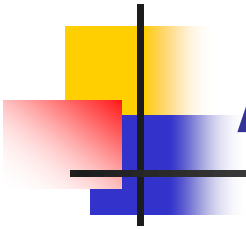
http://users.encs.concordia.ca/~s384_2/



Lecture 5

Evaluate Measurement: **Representational Theory of Software Measurement**

- *Reading material Lecture 5 sections 2.1, 2.2.1, 2.3, 2.4.1 , 2.5, 2.6*



Agenda

- **Review**
- Representational theory of measurement
- Scale Types
- Exercises
- Next?

Software Measurement Key Tasks

Define goals
- Goal-Question-(Indicator)-Metric
- process objectives

Define measures
- document definitions
- storage
- change procedures

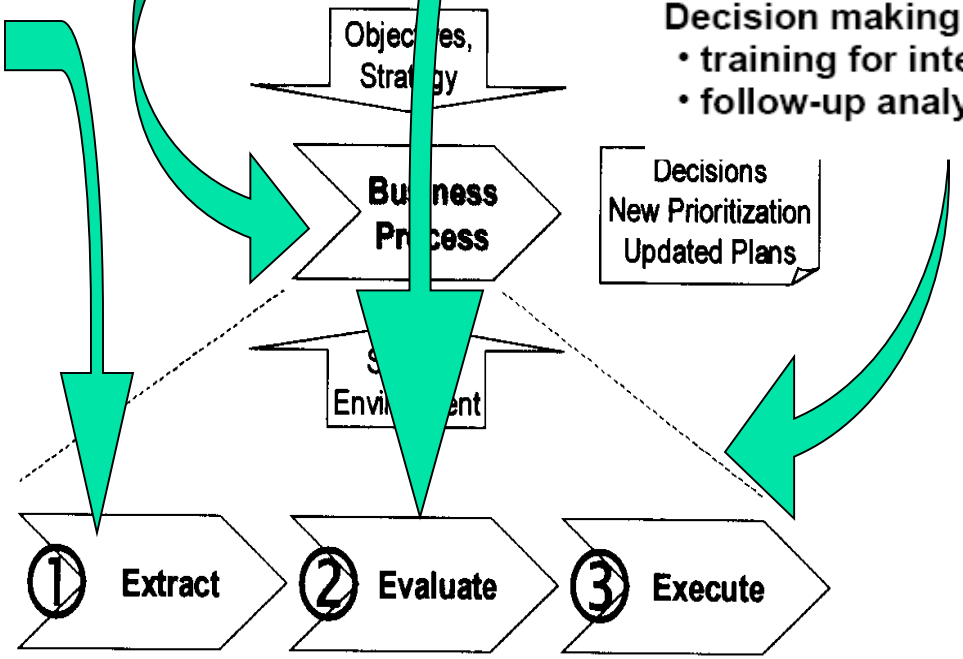
Data collection
- forms
- training
- automation
- procedures

Data storage

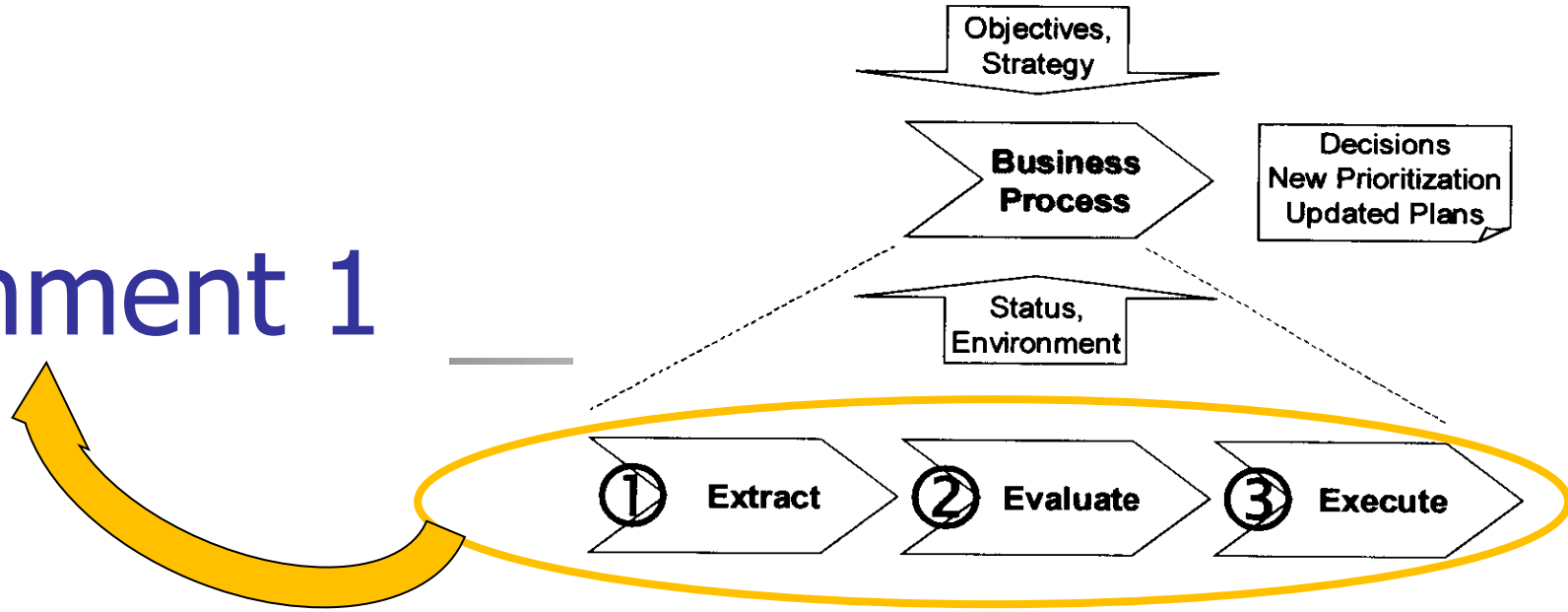
Data analysis
- must support decision making
- must be appropriate for type of measure

Data reporting
- meet timeliness requirements
- understandable content and format

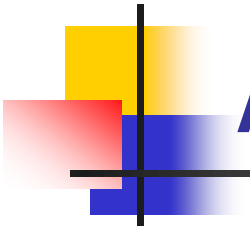
Decision making
- training for interpretation
- follow-up analysis capability



Assignment 1



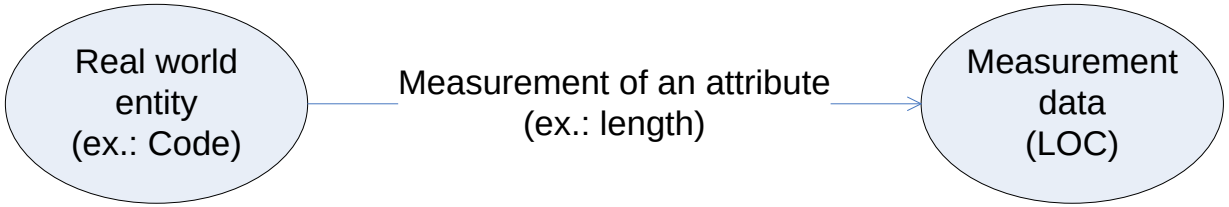
- **Part 1:** Collect and analyze quality data with McCabe
  - Quality trend
- **Part 2:** Collect and analyze quality data with Logiscope
- **Source**: soen390 projects



Agenda

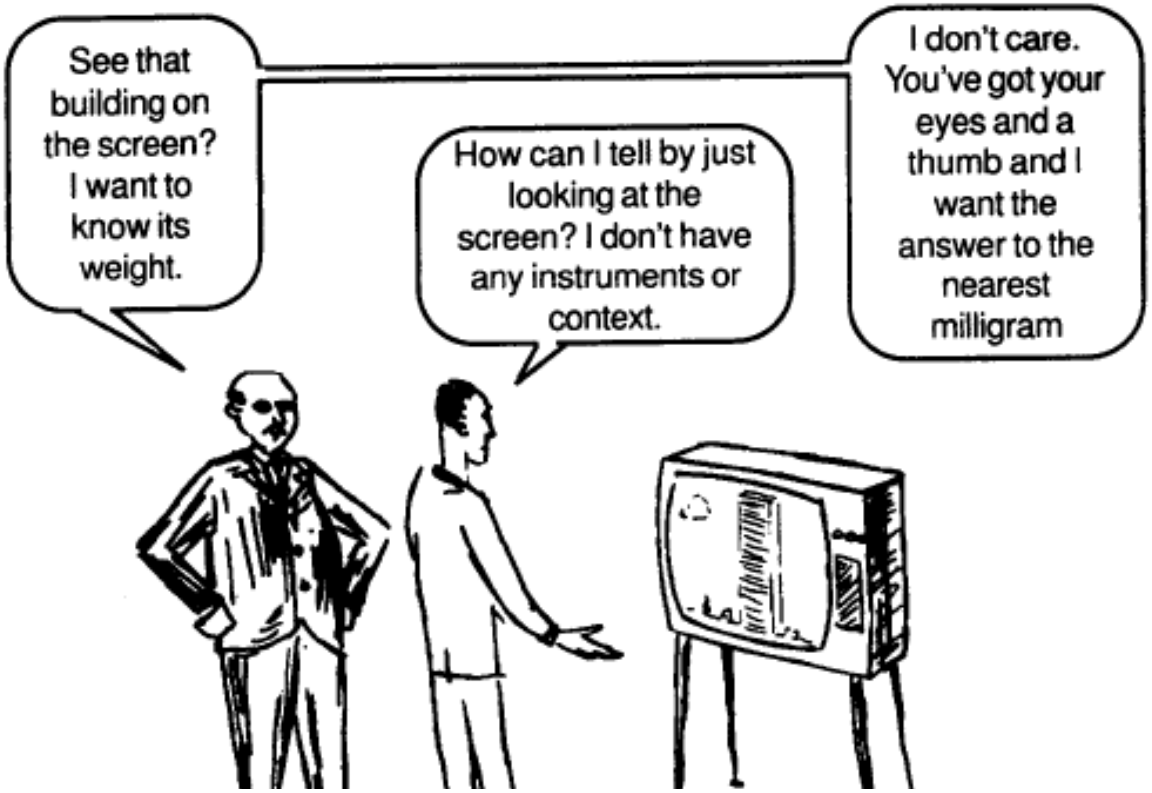
- Review
- **Representational theory of measurement**
- Scale Types
- Exercises
- Next?

What is a Measurement?



- *Measurement* is the process by which numbers or symbols are assigned to attributes of entities in the real world in such a way as to characterize them according to clearly defined rules.

How much must we know about the phenomena before measuring it?

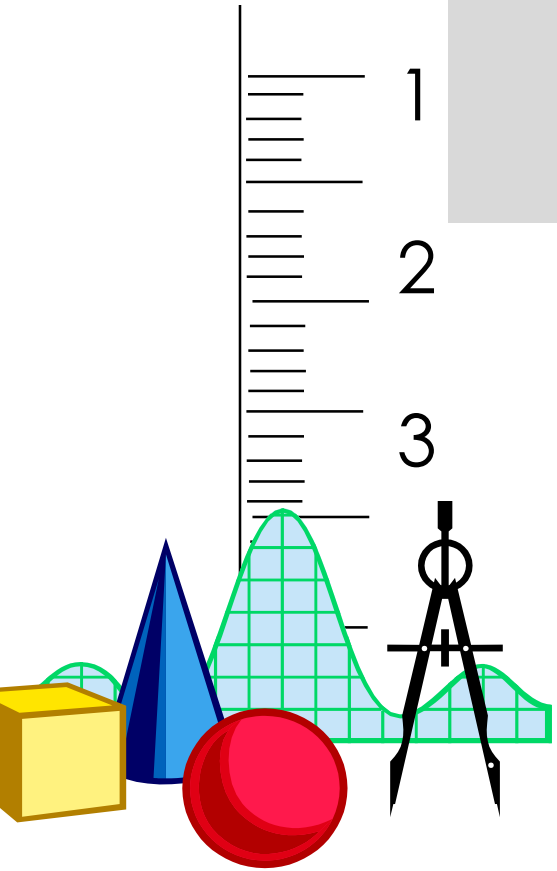


You can neither predict nor control what you cannot measure.

Representational Measurement Theory: Link between Measures & Empiricism

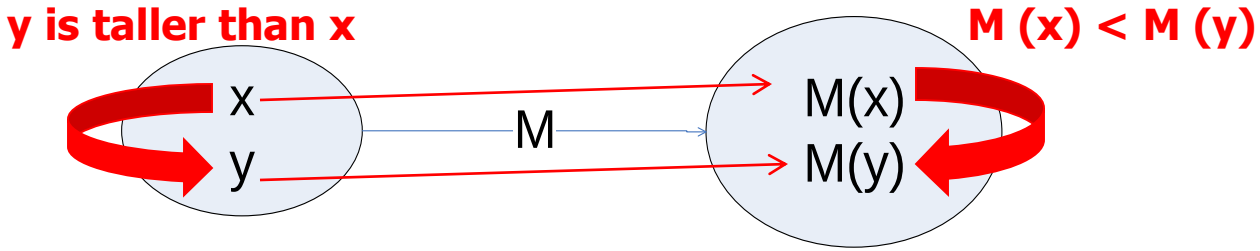
Representational theory of measurement:

rigorous framework for determining when a proposed measure really does characterize the attribute it is supposed to.



x is taller than y ⇔ M (x) > M (y)

Where M(x) is the height of a person x.



Representation condition for derived measurement

- Software A's quality is *good*. **3 defects /10 LOC**
- Software B's quality is ~~*better*~~ ***worse***. **15 defects /1 KLOC**

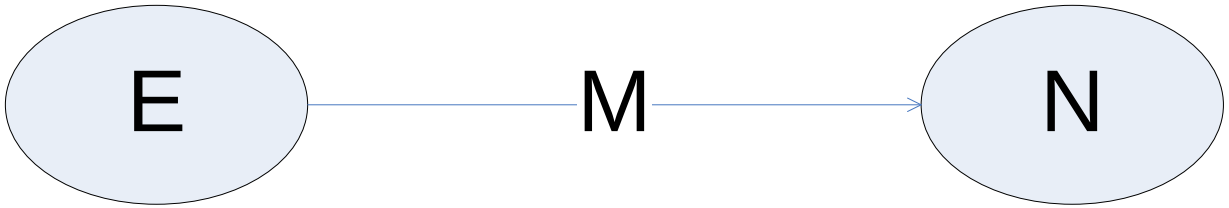
•to measure is to assign numbers to ***quality*** attribute of Software A and Software B according to clearly defined rules;

•"defect", "KLOC" are measurement units

•"reliable", "more reliable" are relations.

Measurement Scale – Formal Definition

- **Scale:** a triple < E, N, M> where
  - E is the set of *empirical (observed) phenomena,*
  - N is the set of *numerical (mathematical) objects,*
  - M is the *mapping between them using a valid and repeatable mechanism* (homomorphism)



Measurement is not about numbers!

- **Example 1:**
  - **Assume there are four football players with the numbers 2, 4 , 6, 8 on their shirts.**
  - We now calculate the arithmetic mean

    AR=(2+4+6+8)/4 = 5
  - Can we say that **the player with number 5 represents the average performance of the four players**?
  - **Answer: *no*** because the numbers are only identification of the players …

Example on Meaningfulness

- Categories of income sources of students:
  - 1=full-time work
  - 2=part-time work
  - 3=scholarship
- **Is the arithmetic mean of the scores meaningful?**

- John and Joe take math test
  - John receives 80
  - Joe receives 40.
- **Is the following statement meaningful?**

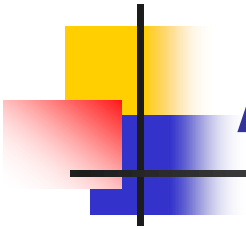
"John knows twice as much as Joe"

Meaningfulness of the analysis of the measurement data and mathematical operations

- **Measurement data is a collection of numbers.**

  One can always perform mathematical operations on numbers (add them, average them, take logarithms, etc.)

- **What kind of mathematical manipulations are meaningful?**
  - **Depends on the scale type of the measure**



Agenda

- Review
- Representational theory of measurement
- **Scale Types**
- Exercises

Meaningfulness and Scale Types of measurement data

- **Five Major Types:**

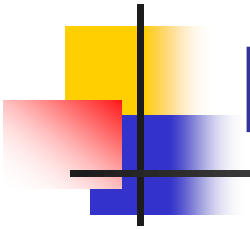
  Nominal, Ordinal, Interval, Ratio, Absolute

  - Represent differing degrees of empirical knowledge about entities and their attributes
  - Shown in increasing level of **richness** of empirical knowledge
  - Restrict analysis of measurement data [on the corresponding scale] to meaningful statistics

Nominal: classification scheme

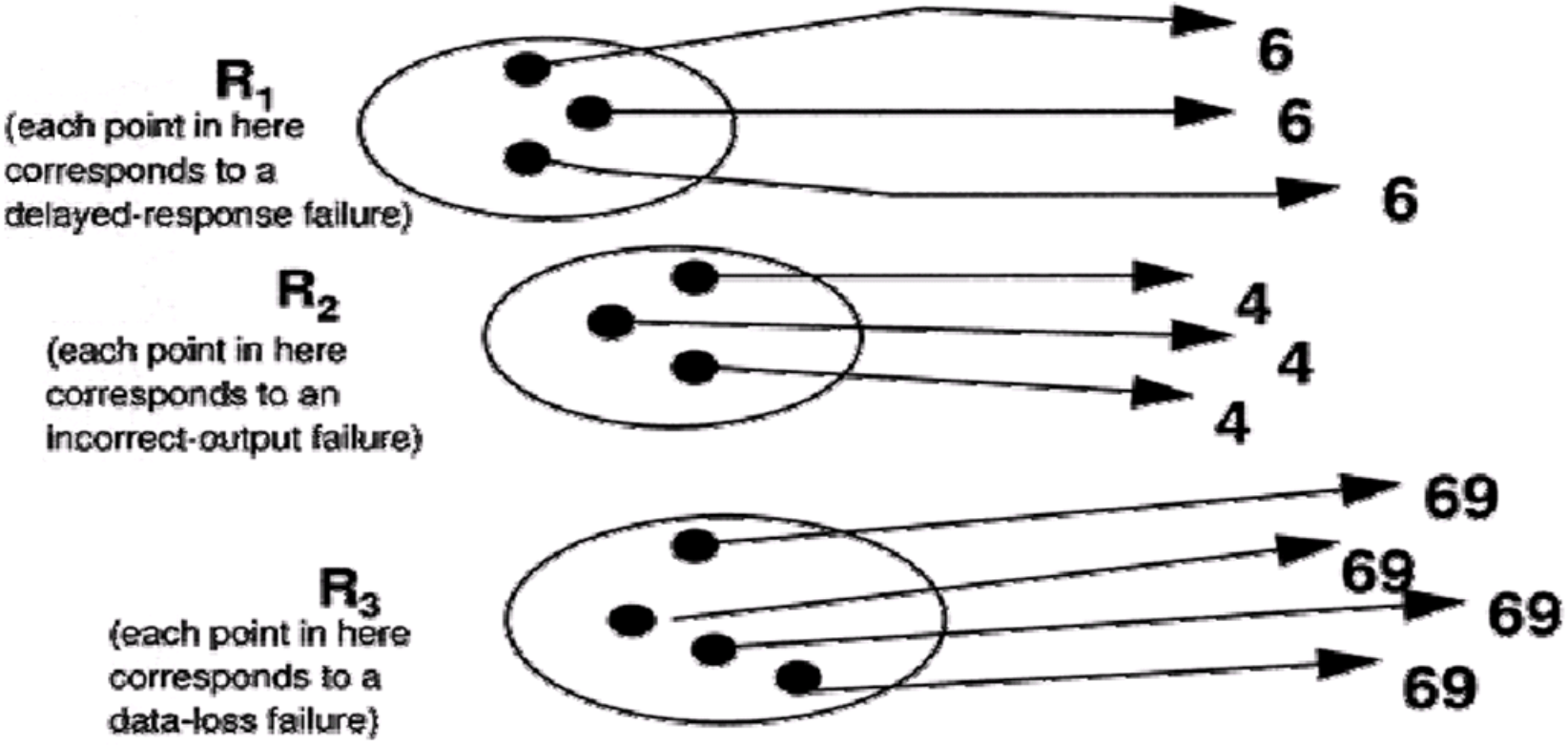
- **Classification**
  - Partitions the set of empirical entities into categories (also called **equivalence classes**) with respect to a certain attribute
  - Equivalence operation: $\sim$
  - No knowledge about relations among categories
- **Empirical Requirements for the partitioning:**
  - Empirical Classes are **jointly exhaustive**
    - ALL CATEGORIES TOGETHER SHOULD COVER ALL POSSIBLE CATEGORIES FOR THE ATTRIBUTE
  - Empirical Classes are **mutually exclusive**
    - A SIBJECT CAN BE CLASSIGIED INTO ONE AND ONLY ONE CATEGORY
- **Numerical Requirements:**
  - Math. Operation: **=**
  - Stats: **Frequency, Mode**
  - Each category is represented by a different number

Nominal Scale Type: Example 2.6, Fenton



- Entity: **software failure**
- Attribute: **criticality**
- Categories of software failures according to their criticality:
  - **R1:** delayed response
  - **R2:** incorrect output
  - **R3:** data loss

Example 2.6: Measurement Method



Back to the 1st example: more formal answer

- ...
  - **Assume there are four football players with the numbers 2, 4 , 6, 8 on their shirts.**
  - Scale type?
  - We now calculate the arithmetic mean

    AR=(2+4+6+8)/4 = 5
  - Is AR a meaningful statistical data?
  - **Answer: *no*** because the arithmetic mean is not a meaningful mathematical operation on the nominal scale.

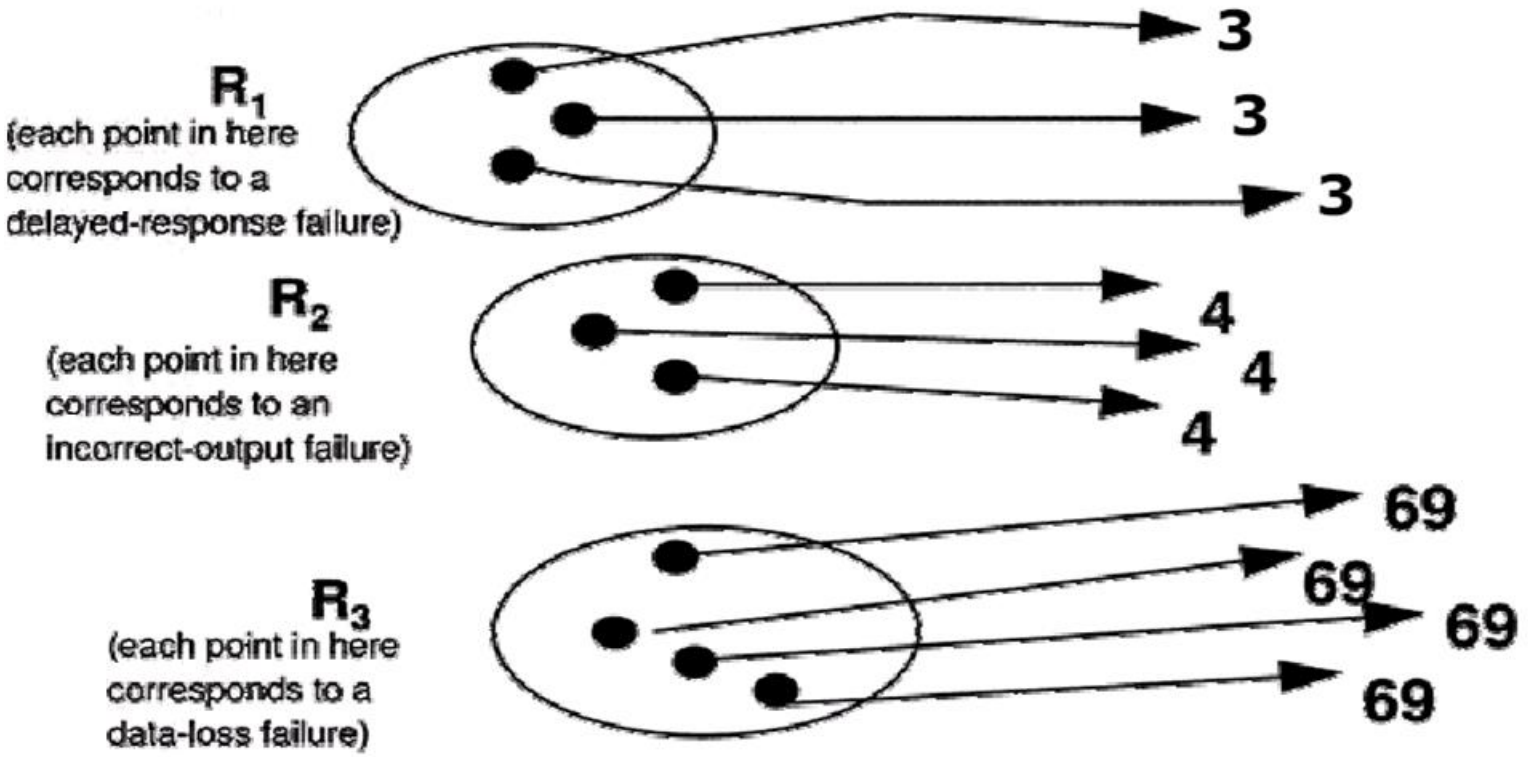
Ordinal:

- comes from Latin name for order;
- Useful to augment the nominal scale with information about an ordering of the categories.
- involves rank ordering items (“>”, “<”, “=”)
- higher number indicates more of the characteristic being measured;
- the intervals between adjacent values are indeterminant.
- The numbers represent ranking only, so **addition, subtraction, and other arithmetic operations have no meaning.**

Example 2.6: from Nominal to Ordinal

- **Next**: add a new binary relation *x* ***more critical than*** *y* :
  - Each data loss failure **(R3)** is more critical than incorrect output **(R2)** and delayed response failure **(R1)**
  - Each incorrect failure **(R2)** is more critical than delayed response failure **(R1)**
- The measurement method shall be revised to account for this new binary relations

Example 2.6: Revised Measurement Method



Ordinal Scale Type: Meaningful mathematical operations

- **Examples of Permissible Statistics & Statistical tests**:
  - All nominal scale statistics
  - Median
  - Rank order statistics (f.e, Spearman's correlation coefficient)
  - Non-parametric

Exercises on Nominal and Ordinal Scale Types

- **Determine the scale type of:**
  - Gender
  - Marital status
  - Distance
  - Intelligence Score
  - "Yes-No" voting system
  - Blood types (A,B,AB,O)
  - Letter grade system

Interval Scale Type

- Requires well-defined **unit** **of measurement**
- **Empirical requirements:**
  - preserves equivalence classes, as with the nominal scale
  - Preserves order, as with the ordinal scale.
  - Scale has **arbitrary** zero, therefore, ratio is meaningless
  - An interval scale **preserves differences** but not ratios.
- **Numerical Requirements:**
  - Math. Operations: **=, <, >, +, -**

Ratio Scale Type

- **Well-defined unit of measurement**
- **Empirical requirements:**
  - preserves ordering, the size of intervals between entities, and ratios between entities.
  - There is a zero element, representing total lack of the attribute.
- **Numerical requirements:**
  - 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 he mapping. =, <, >, +, - , *, /
- **Examples of Statistics & Permissible Statistical tests:**
  - All that apply to interval scale, percentage, coefficient of variation
  - Non-parametric and parametric

Absolute Scale

- The absolute scale is the most restrictive of all. There is **only one way to measure it** and there is **only one unit per absolute measure**!
- **Properties:**
  - The measurement on the absolute scale is made **simply by counting** the number of elements in the entity set.
  - There is only one possible measurement mapping.
  - All arithmetic analysis of the resulting count is meaningful.

More on admissible transformations

Table 2.5: Scales of measurement

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, $M(x) \geq M(y)$ implies $M'(x) \geq M'(y)$	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

Revisiting the Representation Condition

- The condition that, if one software entity is less than another entity in terms of a selected attribute, then any software metric for that attribute must associate a smaller number to the first entity than it does to the second entity.
[http://encyclopedia2.thefreedictionary.com/representation+condition]
  - Example: LOC (next)

Theoretical Validation: LOC Example

Length of joint programs should be

$$m(p_1 ; p_2) = m(p_1) + m(p_2)$$

If length of p_1 is greater that p_2, any measure of length should satisfy $m(p_1) > m(p_2)$

Base Measures v.s. Derived Measures

- A measure is **base (direct)** if it directly characterizes an empirical property and does not require the prior measurement of some property
- **Derived (indirect) measures**: use one or more fundamental measures of one or more attributes to measure, indirectly, another supposedly related attribute.
  1. Requires first the measurement of two or more attributes
  2. Then it combines them using a mathematical model of some kind, according to the laws imposed by the empirical model.

Scale Types of Derived Measure. Example

Scale of an indirect measure *M* will generally be the weakest of the scale types of the direct measures

A commonly used indirect measure of programmer productivity P is:

$$P = L / E$$

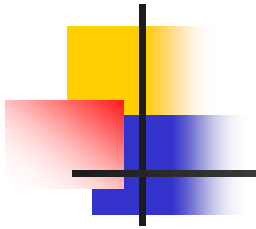
Where L is the number of executable lines of code, and E is effort in man-months (i.e., the number of months required for one developer to write the L lines of code).

Determine the scale type for *P*. Justify your answer.

Exercises on Scale Types

- **Determine the scale type of:**
  - US quality of beef rating
  - Annual Income in $
  - Distance
  - Weight
  - Number of developers in a team
  - Length
  - Duration of time in a project
  - Kelvin temperature scale

More exercises on meaningfulness (in class)



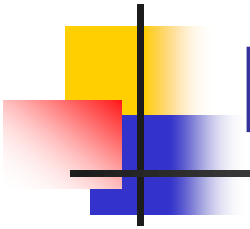
Determine which of the following statements are meaningful:

i. The length of Program A is 50.
ii. The length of Program A is 50 executable statements.
iii. Program A took 3 months to write.
iv. Program A is twice as long as Program B.

v. Program A is 50 lines longer than Program B.
vi. The cost of maintaining program A is twice that of maintaining Program B.
vii. Program B is twice as maintainable as Program A.
viii. Program A is more complex than Program B.

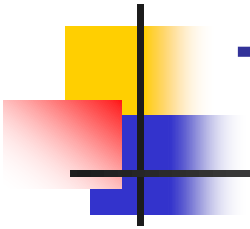
Meaningfulness of the analysis of measurement data

- Given two distances D1=10km and D2=5km:
- Can we say that the distance D1 is twice as long as the distance D2? (**D1=D2*2**)
- Will this statement be meaningful if the length is measured in meters or yards?



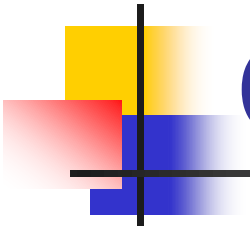
Next?

- Analyzing Software Measurement Data



TO-DO list:

- READ Chapter 2, Fenton (reading material Lecture 5)
-



Questions?

- …