Personality Research Part 2
Psychometric Theory

- Part I: Descriptive and causal explanations of personality
- Part II: How do we measure personality

Lord Kelvin's dictum

In physical science a first essential step in the direction of learning any subject is to find principles of numerical reckoning and methods for practicably measuring some quality connected with it. I often say that when you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the stage of science, whatever the matter may be. (Thomsom, 1891)

Taken from Michell (2003) in his critique of psychometrics:
Michell, J. The Quantitative Imperative: Positivism, Naïve Realism and the Place of Qualitative Methods in Psychology, Theory & Psychology, Vol. 13, No. 1, 5-31 (2003)

Psychometric Theory

- ‘The character which shapes our conduct is a definite and durable ‘something’, and therefore ... it is reasonable to attempt to measure it. (Galton, 1884)
- “Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quantity as well as its quality” (E.L. Thorndike, 1918)

Psychology and the need for measurement

- The history of science is the history of measurement (J. M. Cattell, 1893)
- We hardly recognize a subject as scientific if measurement is not one of its tools (Boring, 1929)
- There is yet another [method] so vital that, if lacking it, any study is thought ... not be scientific in the full sense of the word. This further an crucial method is that of measurement. (Spearman, 1937)
- One's knowledge of science begins when he can measure what he is speaking about and express in numbers (Eysenck, 1973)

Michell, J. The Quantitative Imperative: Positivism, Naïve Realism and the Place of Qualitative Methods in Psychology, Theory & Psychology, Vol. 13, No. 1, 5-31 (2003)

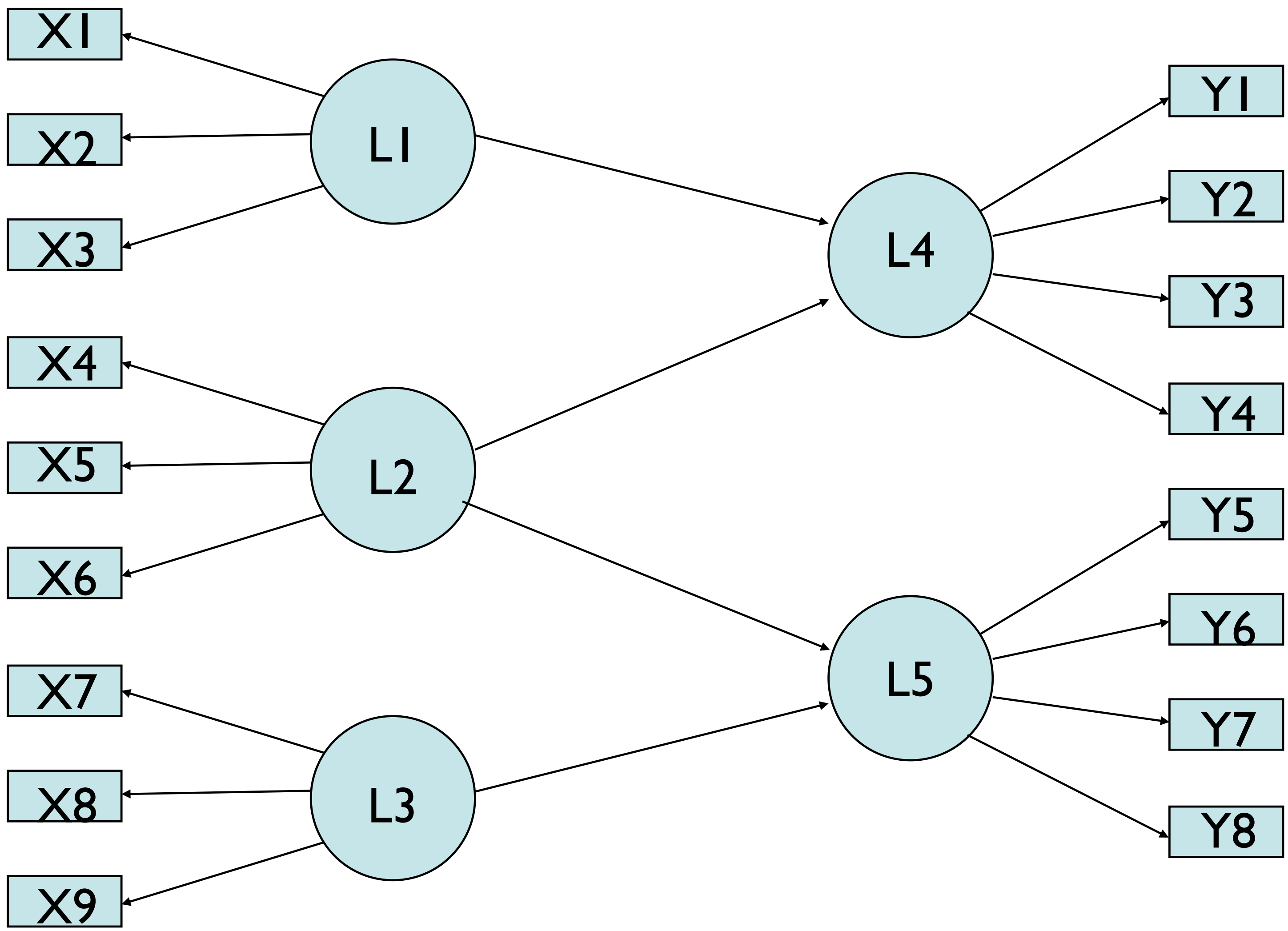
Psychometric Theory: Goals

1. To acquire the fundamental vocabulary and logic of psychometric theory.
2. To develop your capacity for critical judgment of the adequacy of measures purported to assess psychological constructs.
3. To acquaint you with some of the relevant literature in personality assessment, psychometric theory and practice, and methods of observing and measuring affect, behavior, cognition and motivation.

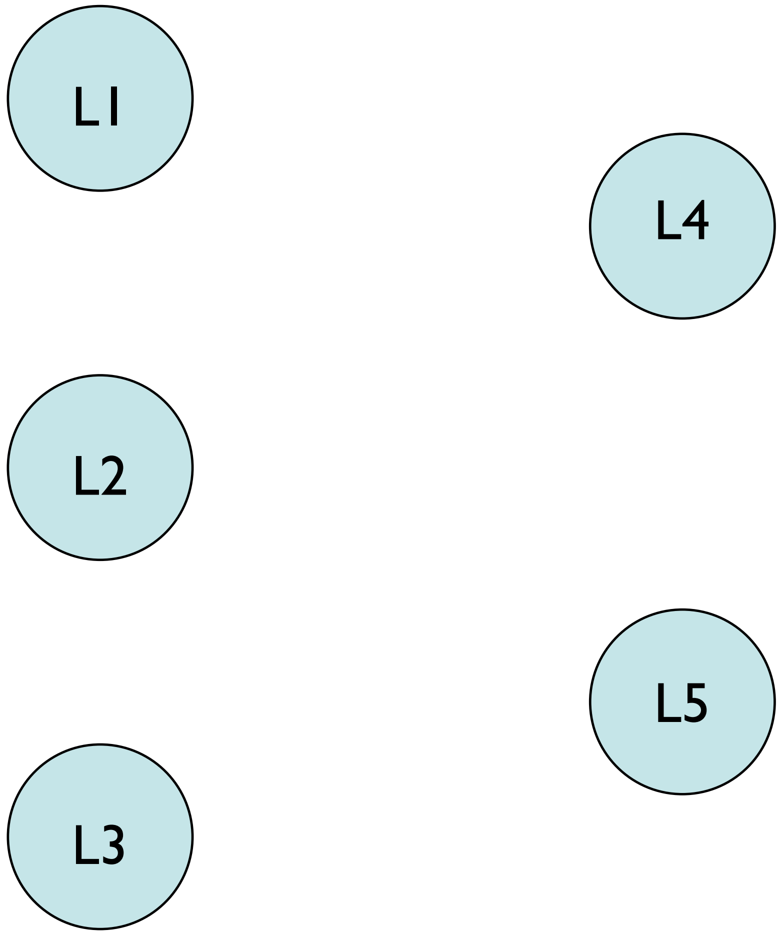
Psychometric Theory: Goals II

1. To instill an appreciation of and an interest in the principles and methods of psychometric theory.
2. This course is not designed to make you into an accomplished psychometist (one who gives tests) nor is it designed to make you a skilled psychometrician (one who constructs tests)
3. It will give you limited experience with psychometric computer programs (examples using R).

Psychometric Theory: A conceptual Syllabus



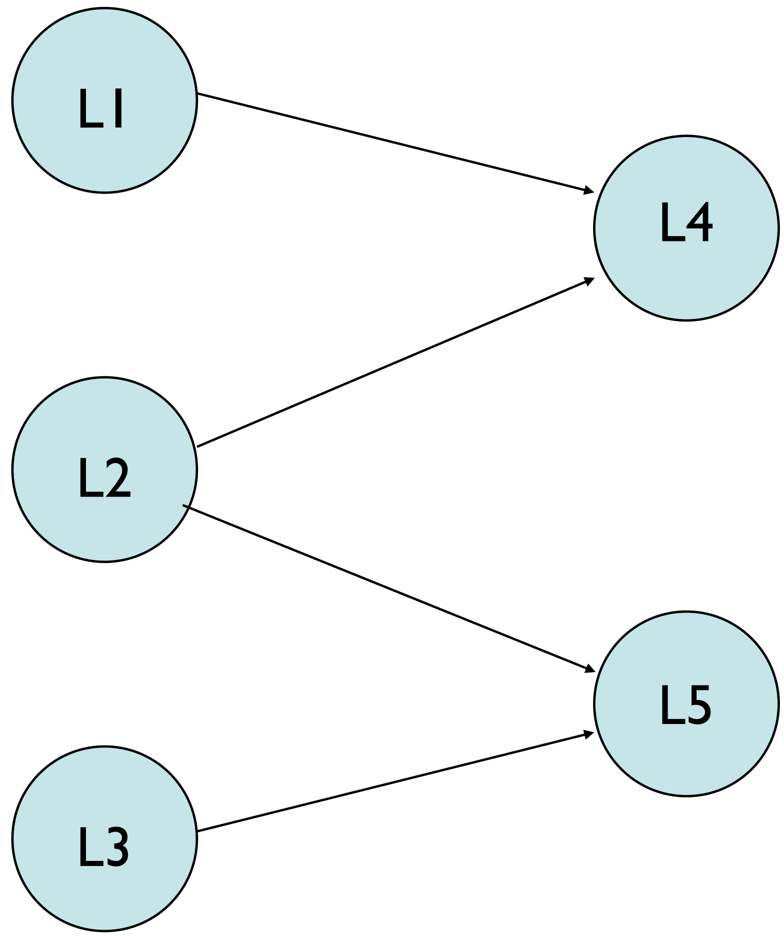
Constructs/Latent Variables



Examples of psychological constructs

- Anxiety
  - Trait
  - State
- Love
- Conformity
- Intelligence
- Learning and memory
  - Procedural - memory for how
  - Episodic -- memory for what
    - Implicit
    - explicit
- ...

Theory as organization of constructs



Theories as metaphors and analogies-1

- Physics
  - Planetary motion
    - Ptolemy
    - Galileo
    - Einstein
  - Springs, pendulums, and electrical circuits
  - The Bohr atom
- Biology
  - Evolutionary theory
  - Genetic transmission

Theories as metaphors and analogies-2

- Business competition and evolutionary theory
  - Business niche
  - Adaptation to change in niches
- Learning, memory, and cognitive psychology
  - Telephone as an example of wiring of connections
  - Digital computer as information processor
  - Parallel processes as distributed information processor

Models and theory

- Formal models
  - Mathematical models
  - Dynamic models - simulations
- Conceptual models
  - As guides to new research
  - As ways of telling a story
    - Organizational devices
    - Shared set of assumptions

Observable or measured variables

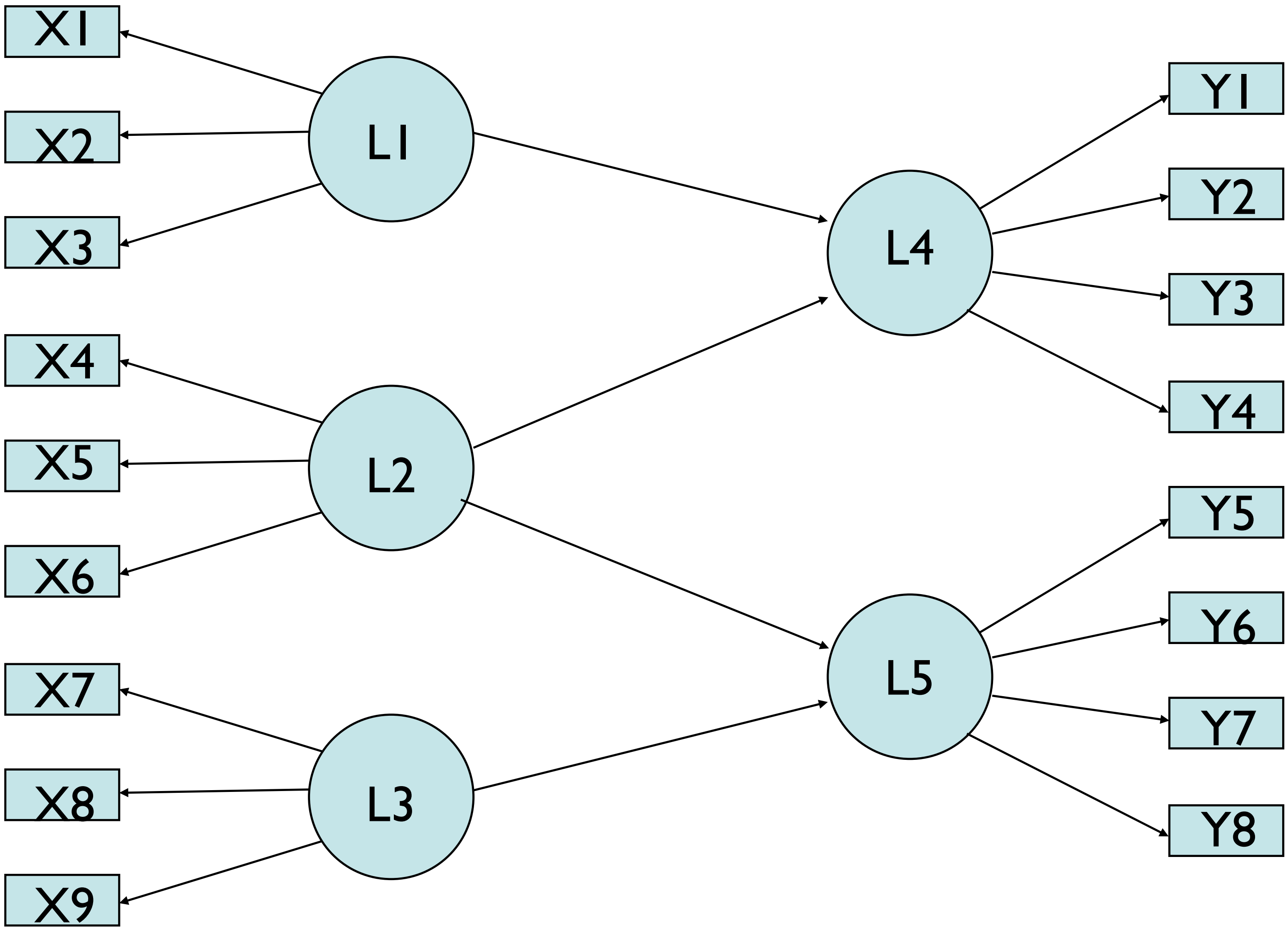
Observed Variables

- Item Endorsement
- Reaction time
- Choice/Preference
- Blood Oxygen Level Dependent Response
- Skin Conductance
- Archival measures

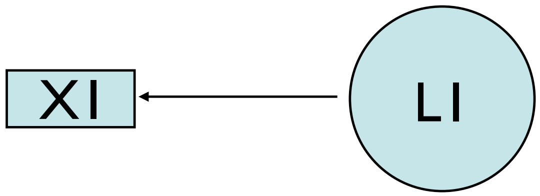
Theory development and testing

- Theories as organizations of observable variables
- Constructs, latent variables and observed variables
  - Observable variables
    - Multiple levels of description and abstraction
    - Multiple levels of inference about observed variables
  - Latent Variables
    - Latent variables as the common theme of a set of observables
    - Central tendency across time, space, people, situations
  - Constructs as organizations of latent variables and observed variables

Psychometric Theory: A conceptual Syllabus



A Theory of Data: What can be measured

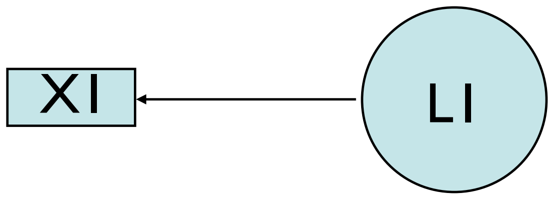


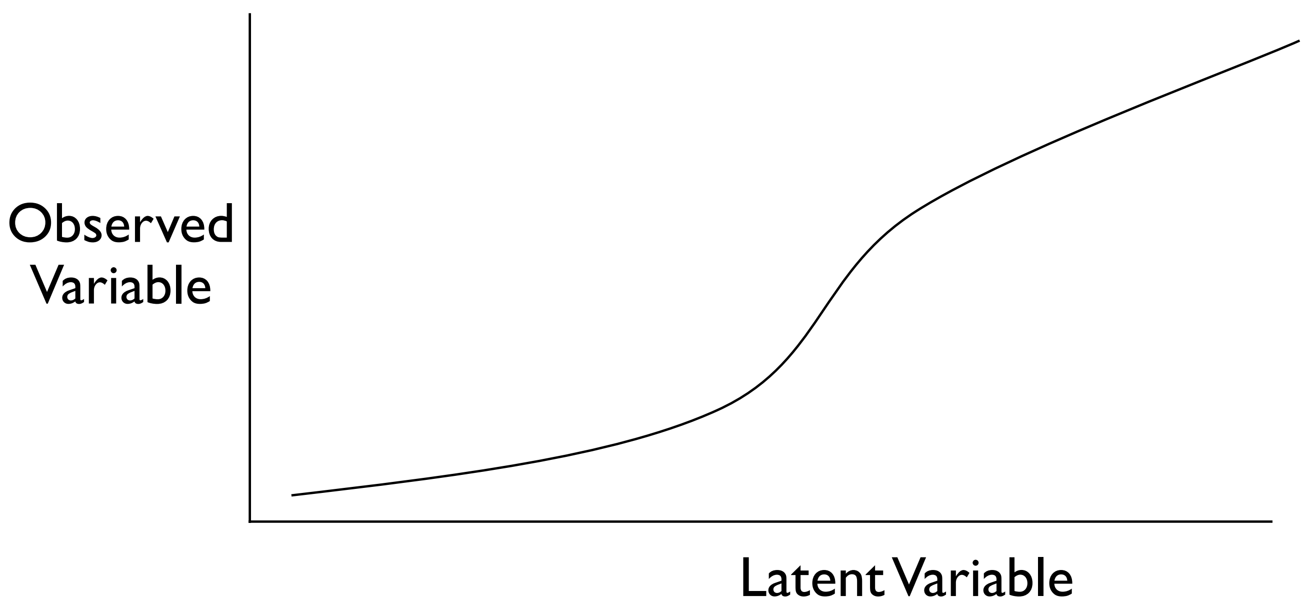
What is measured?
- Objects
- Individuals

What kind of measures are taken?
- Order
- Proximity

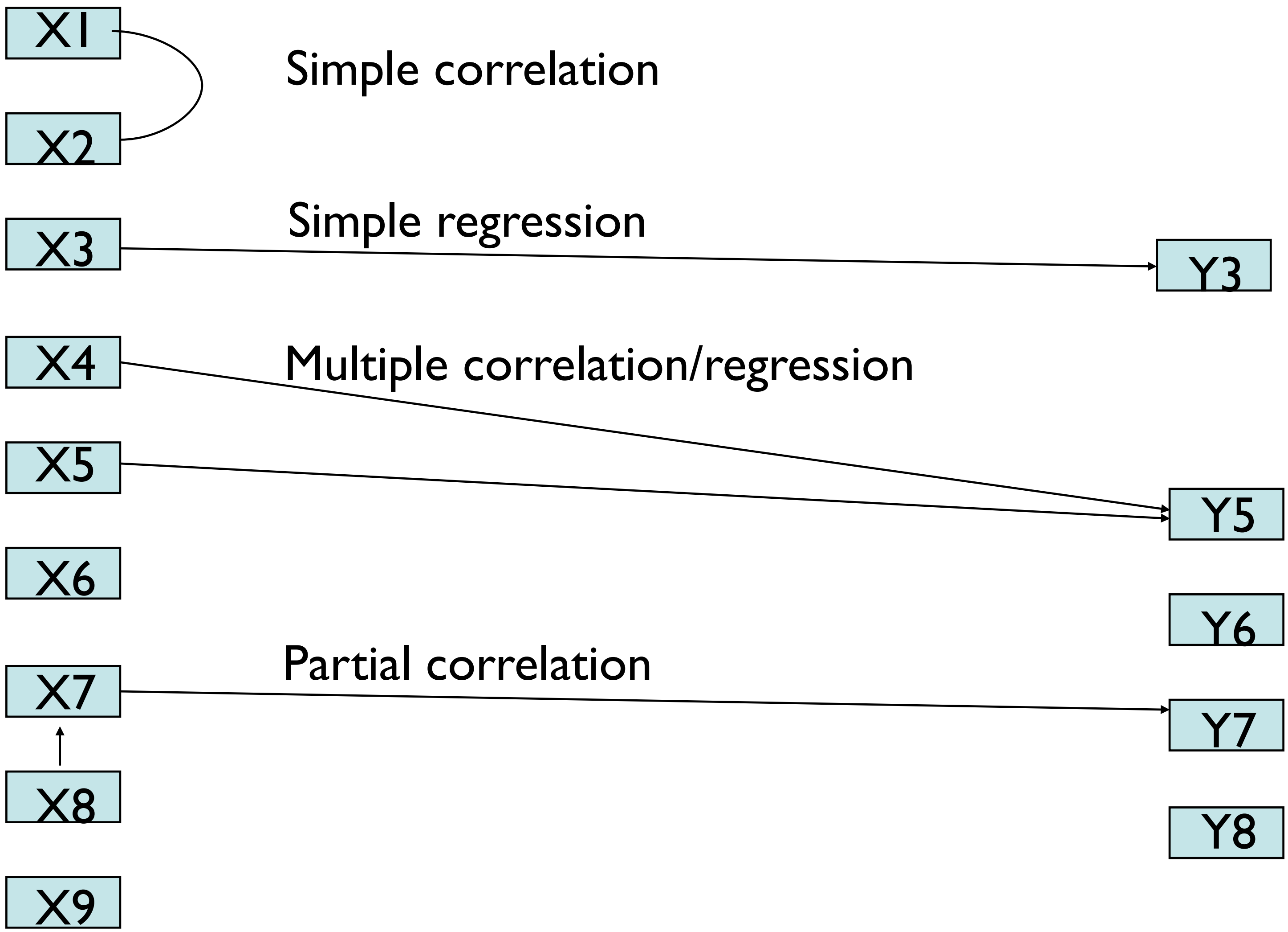
What kind of comparisons are made?
- Single Dyads
- Pairs of Dyads

Scaling:
the mapping between observed and latent variables

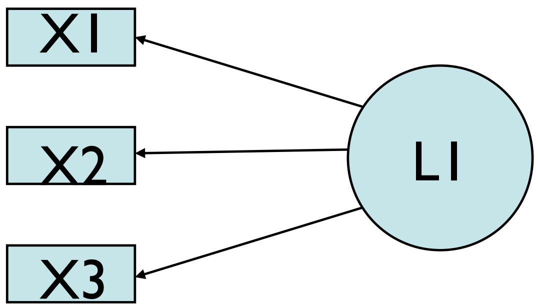




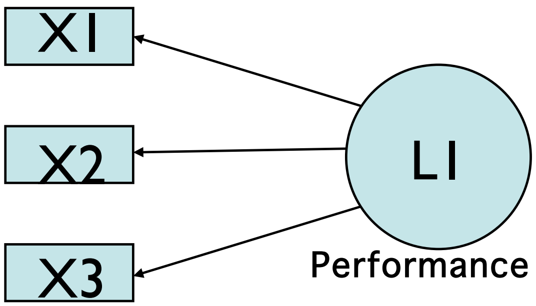
Variance, Covariance, and Correlation



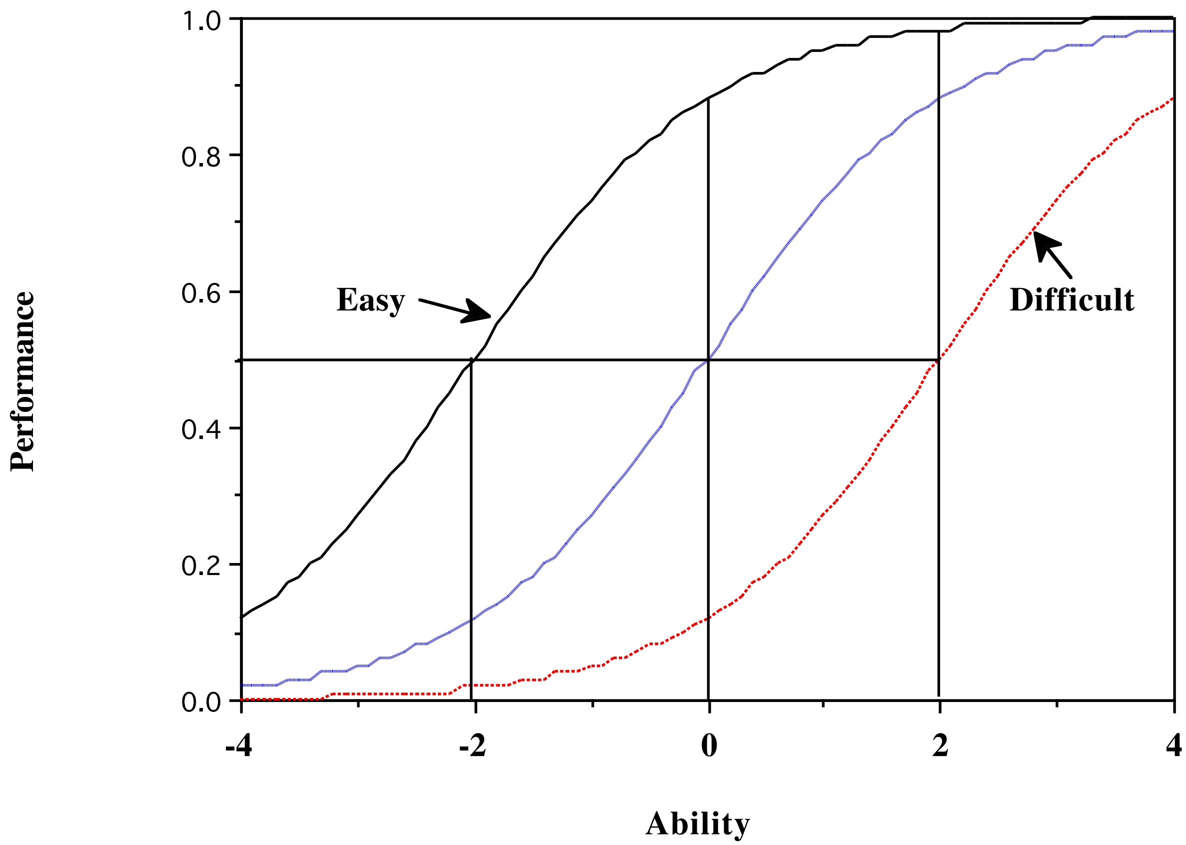
Classic Reliability Theory: How well do we measure what ever we are measuring



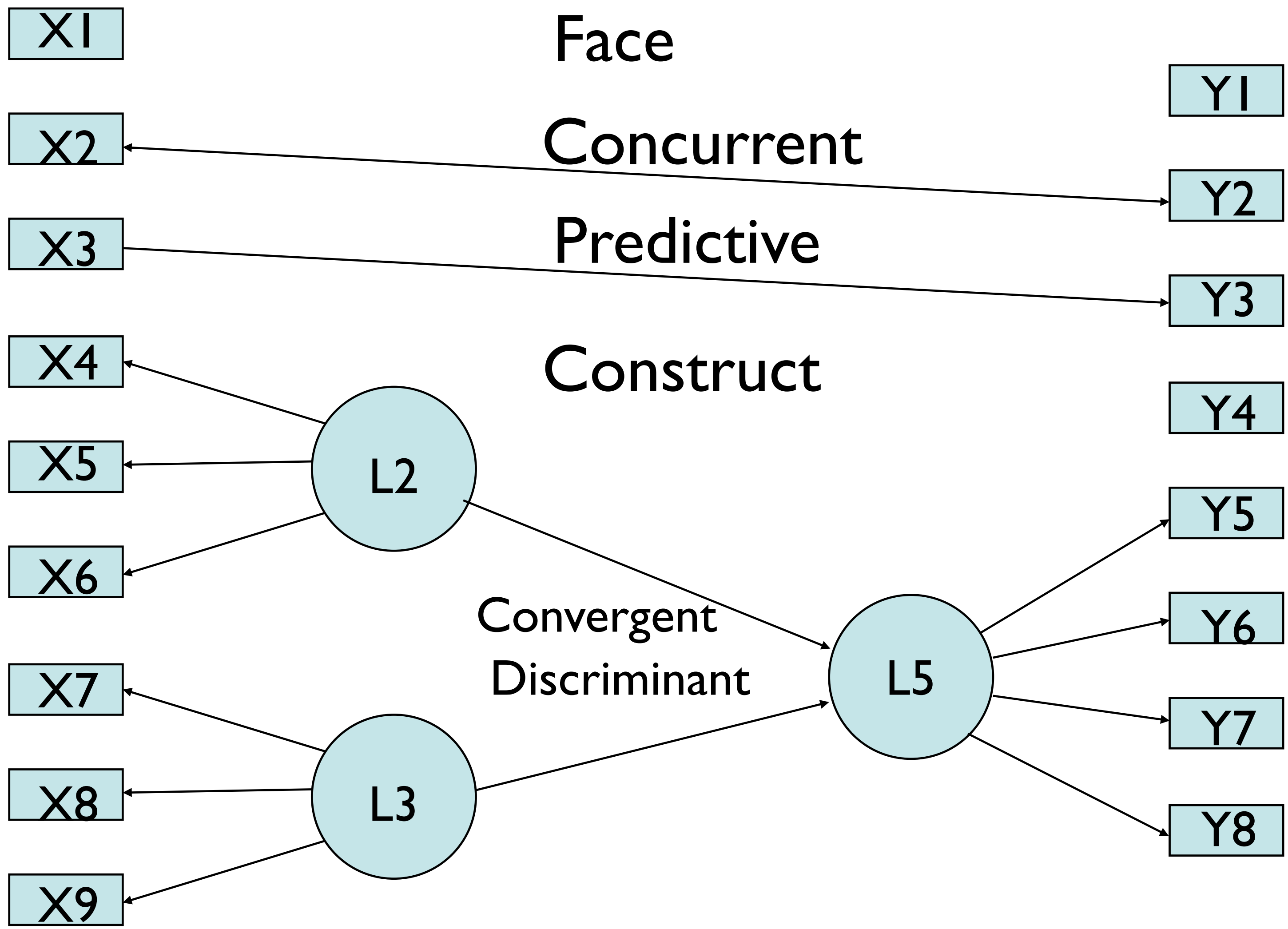
Modern Reliability Theory: Item Response Theory
How well do we measure what ever we are measuring



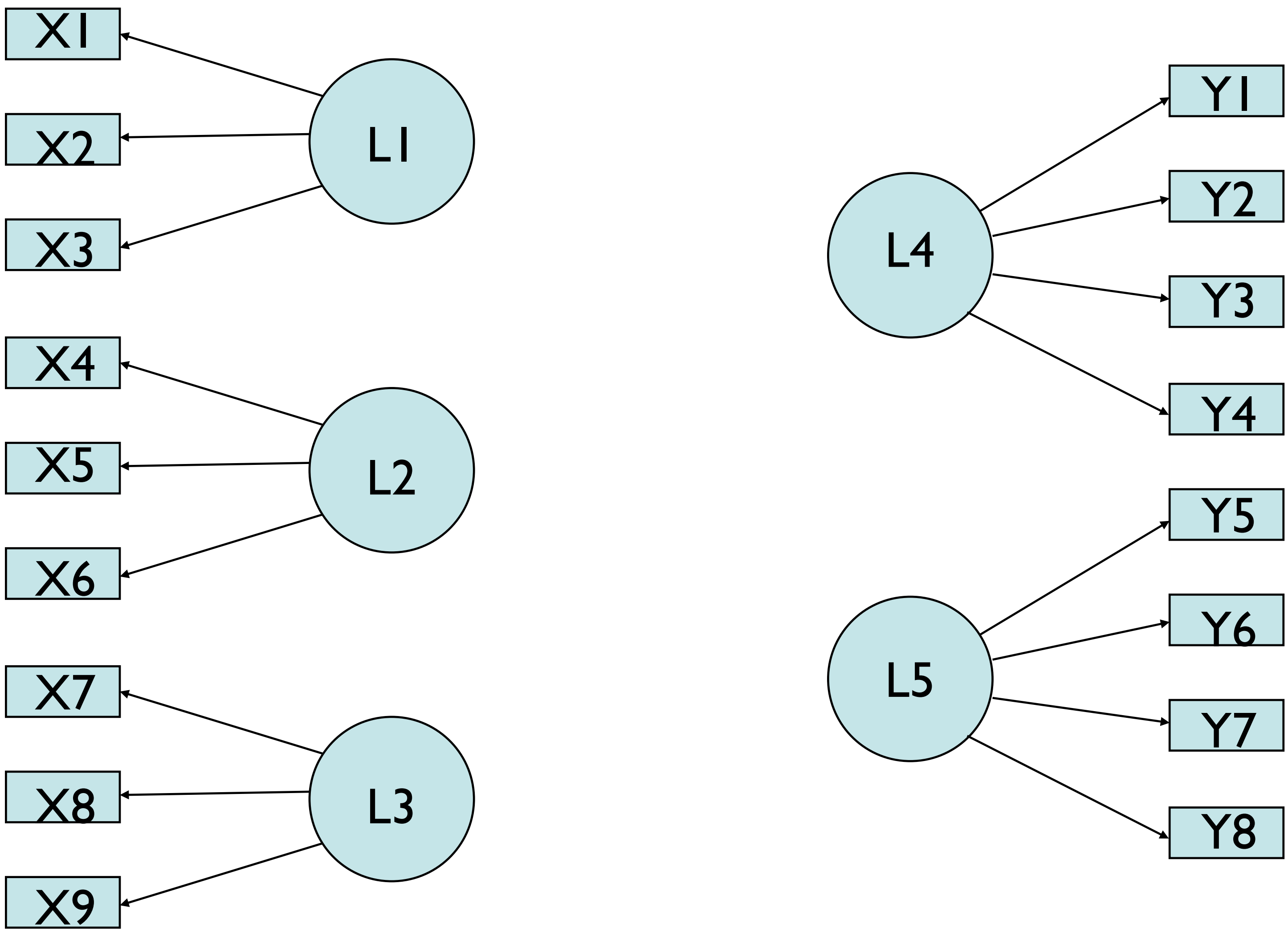
Performance as a function of Ability and Test Difficulty



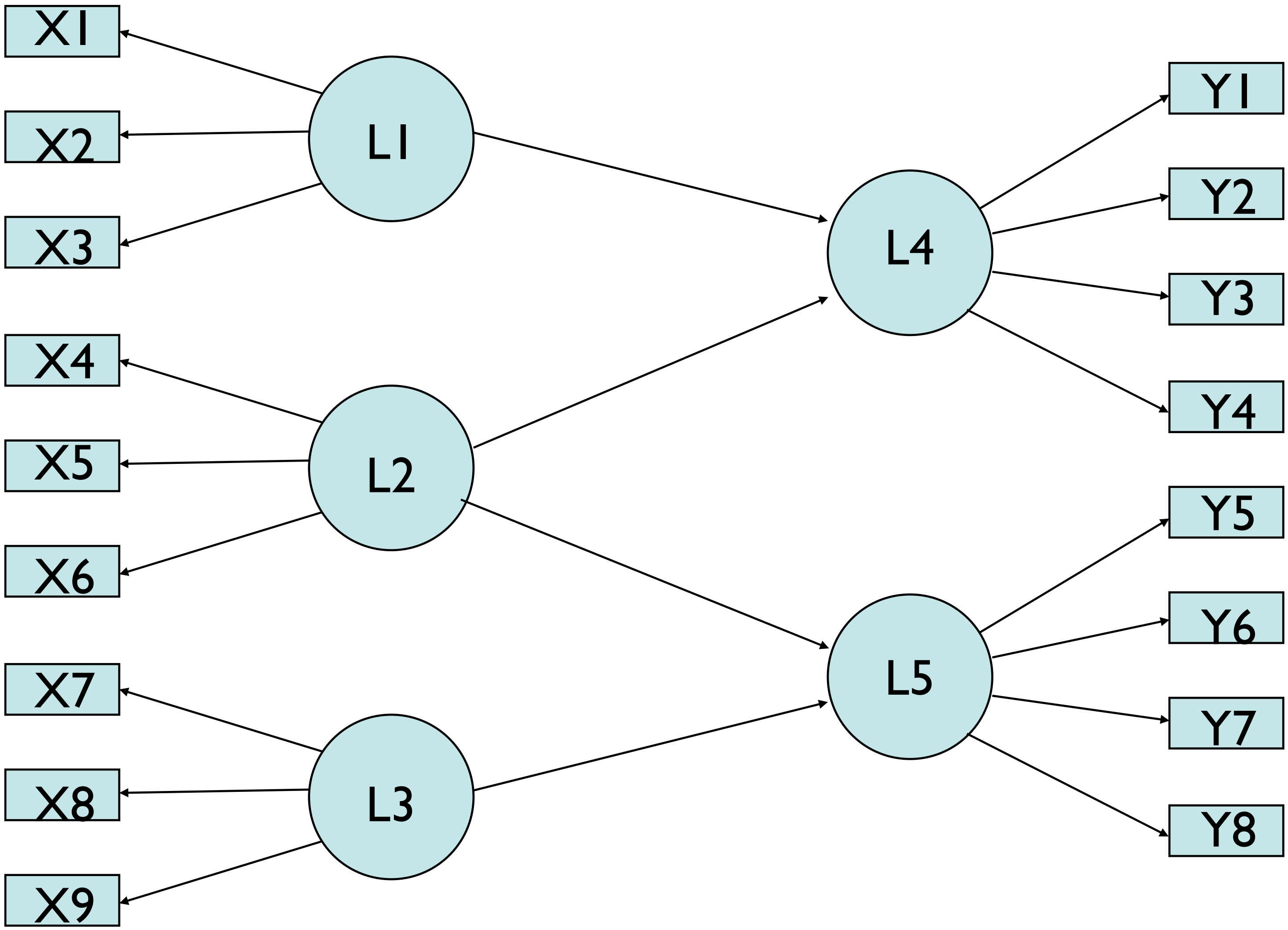
Types of Validity: What are we measuring



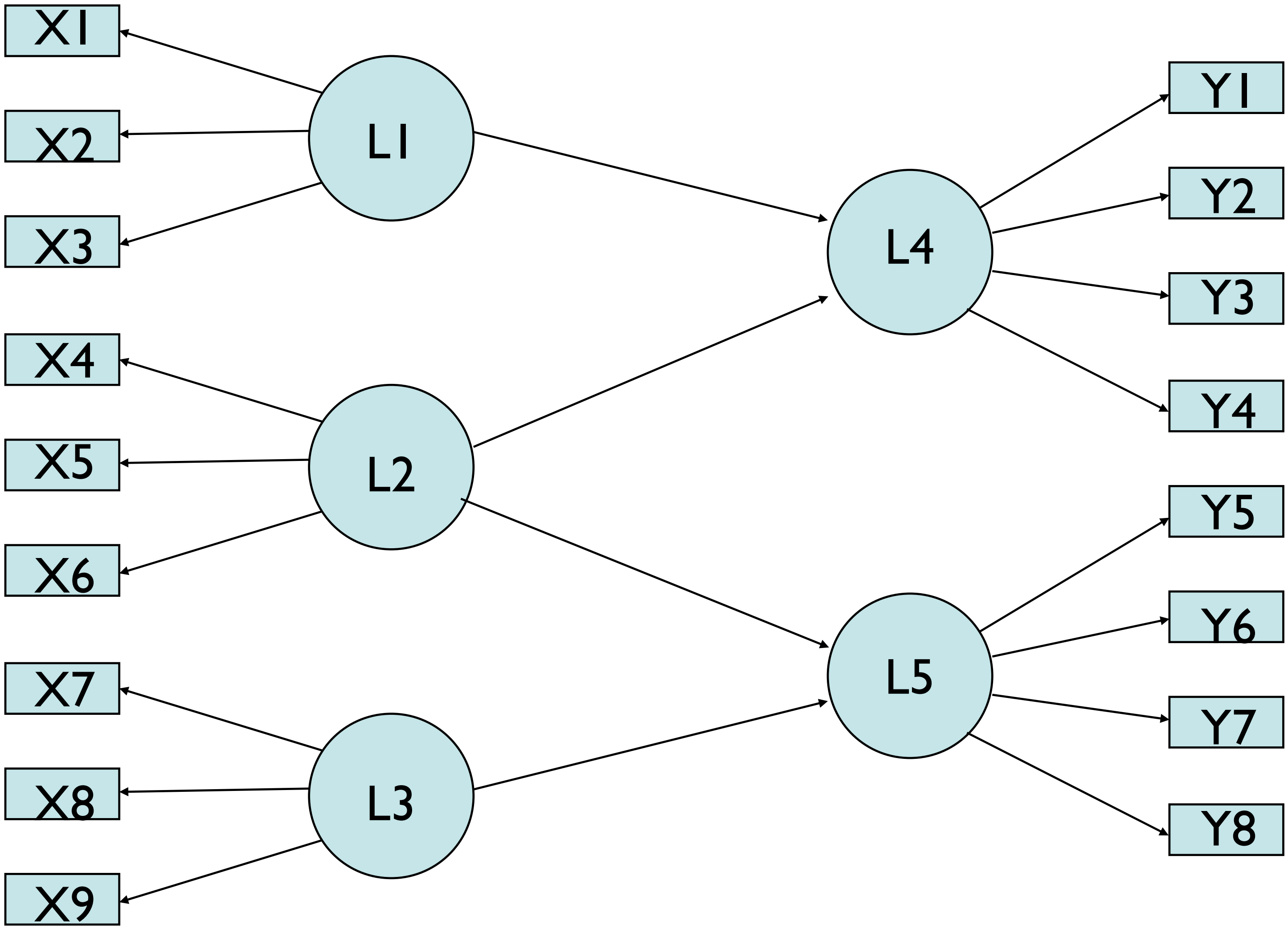
Techniques of Data Reduction: Factor and Components Analysis



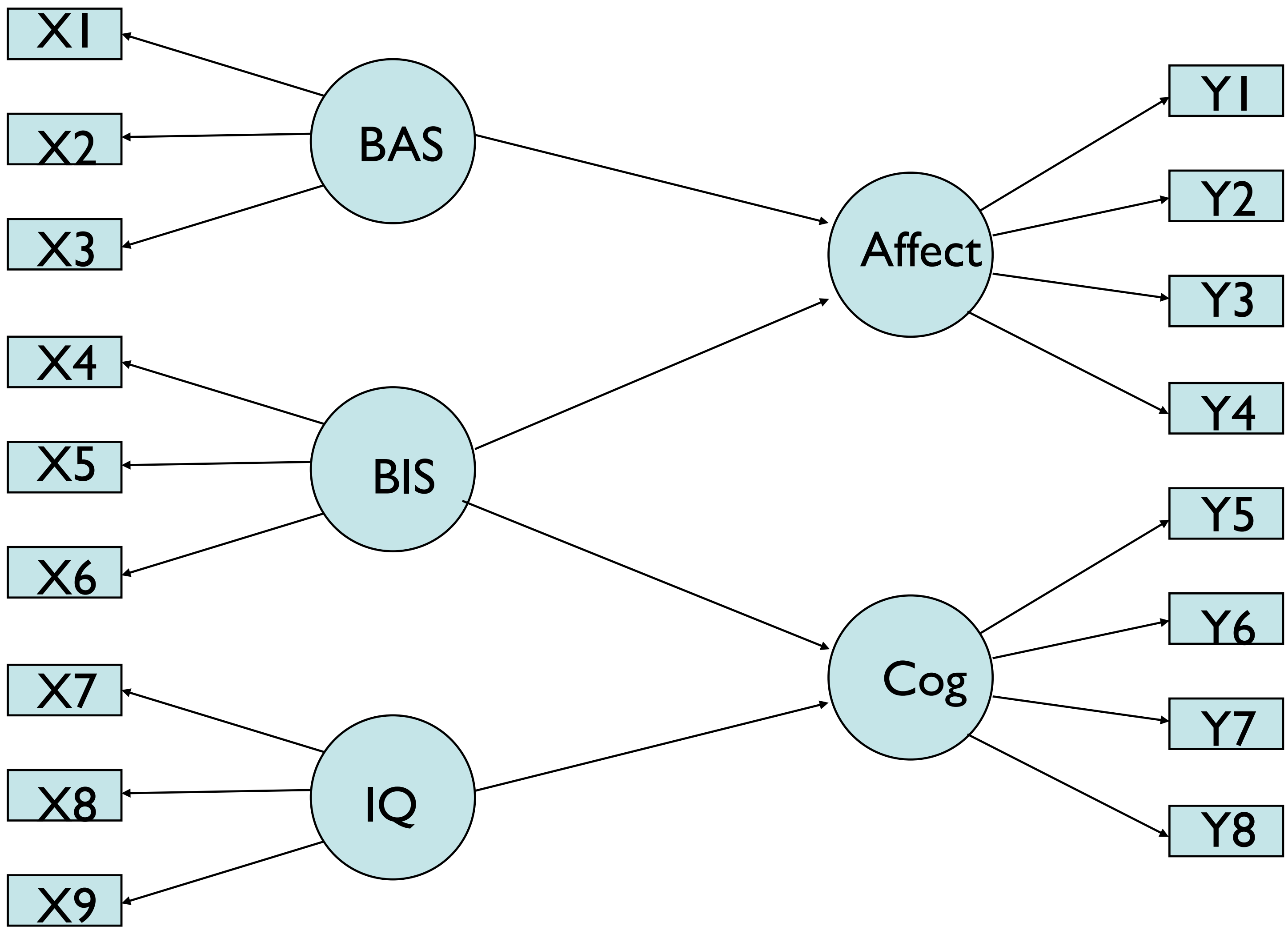
Structural Equation Modeling: Combining Measurement and Structural Models



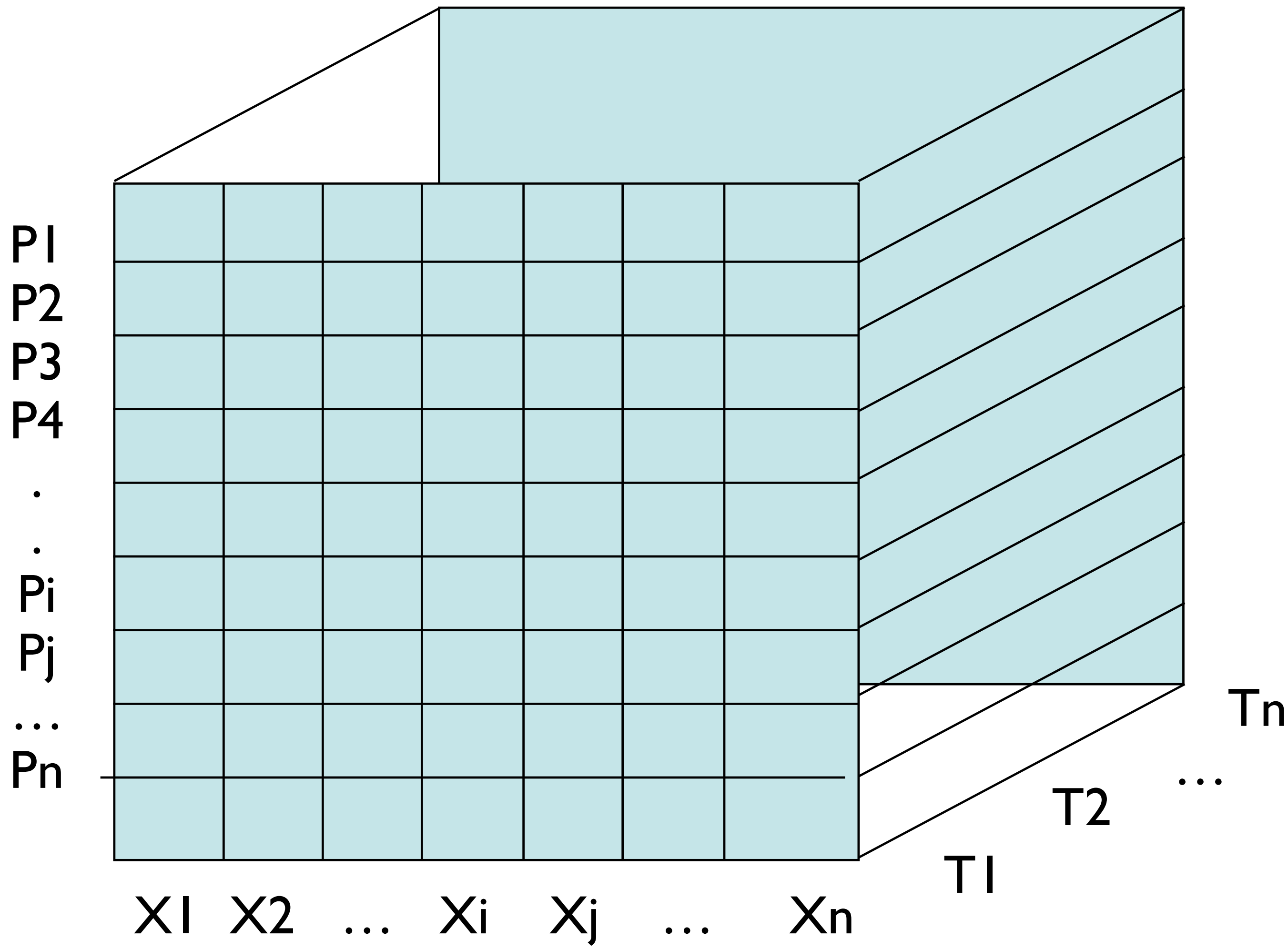
Scale Construction: practical and theoretical



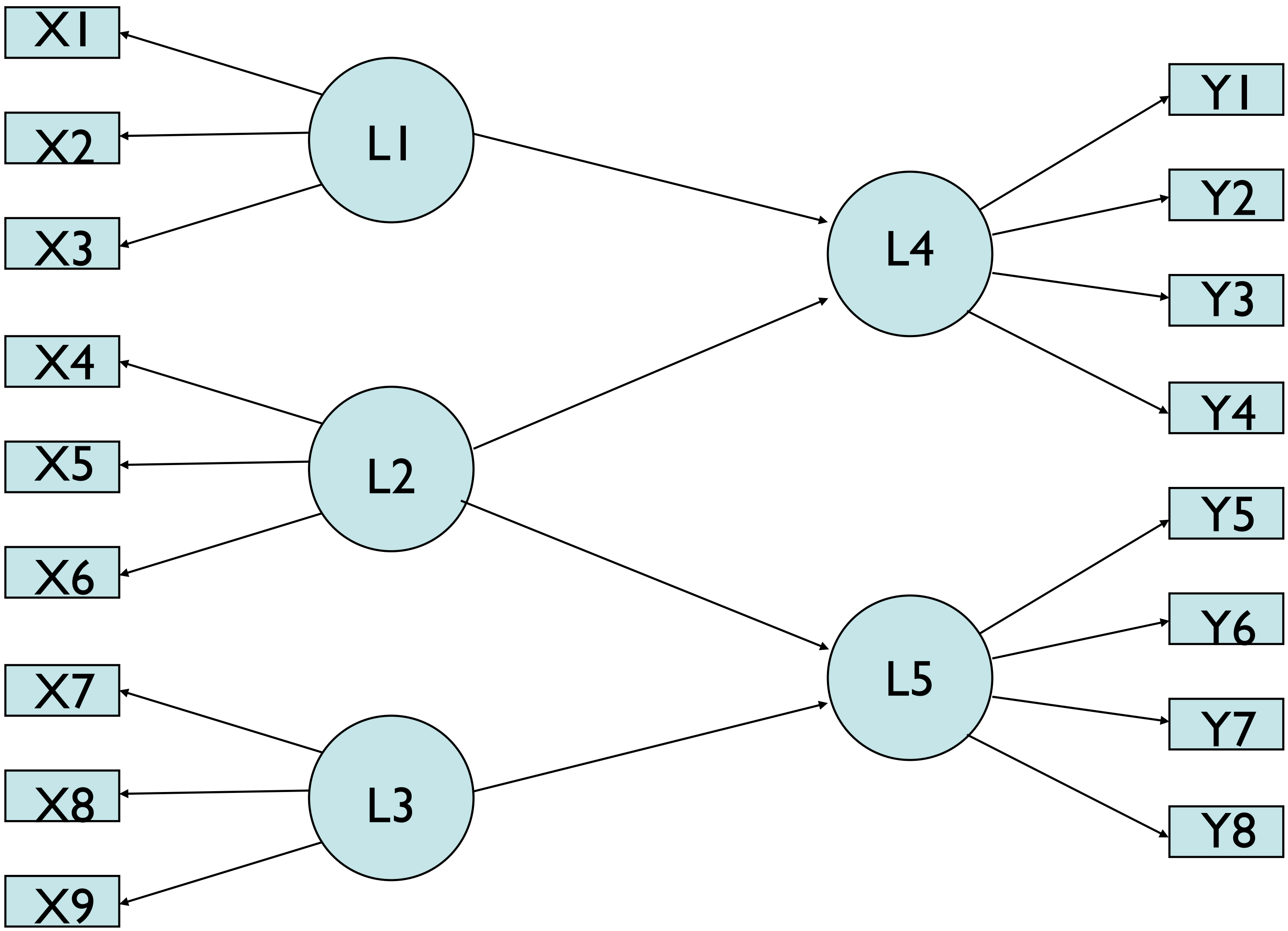
Traits and States: What is measured?



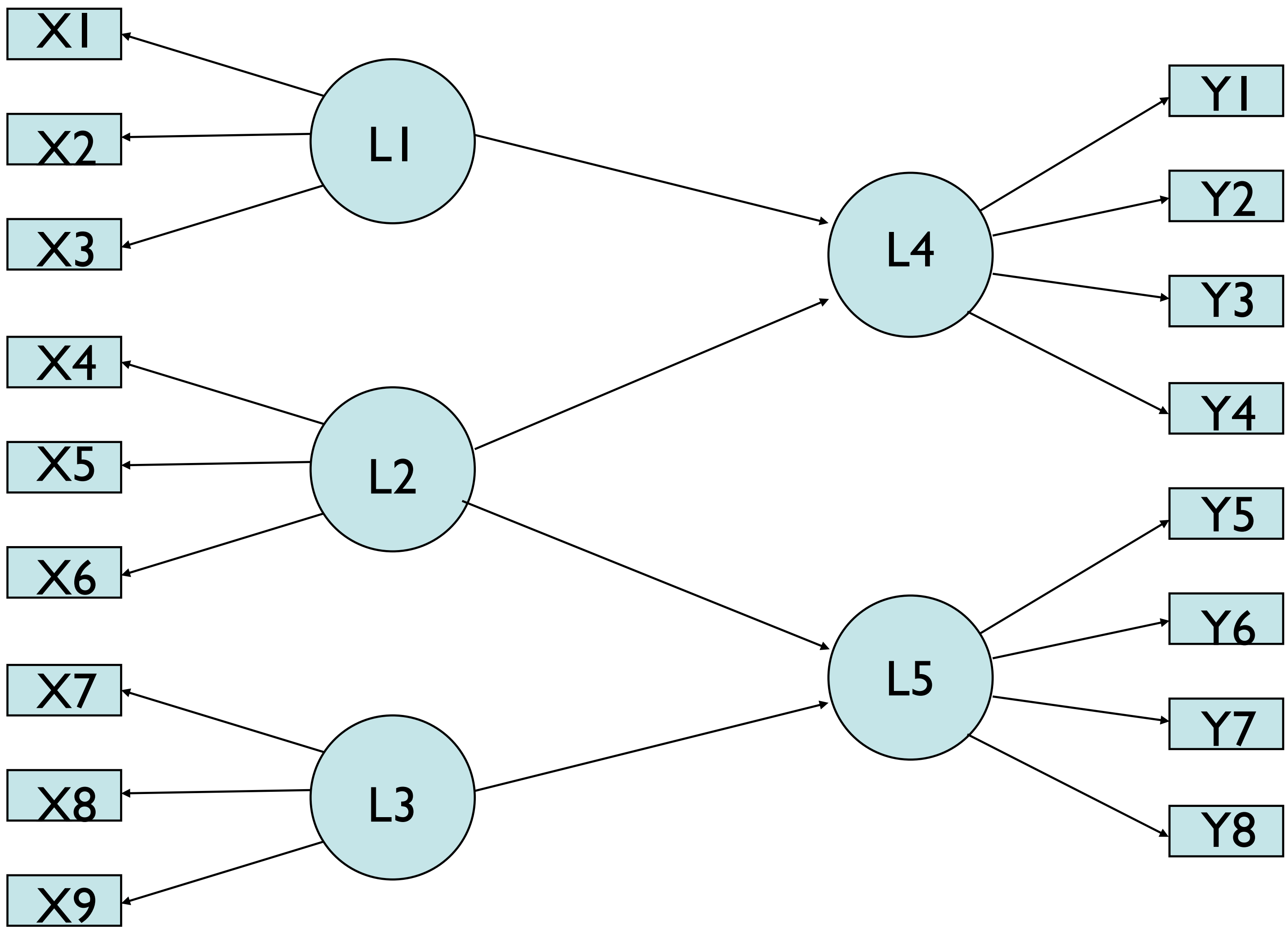
The data box: measurement across time, situations, items, and people



Psychometric Theory: A conceptual Syllabus



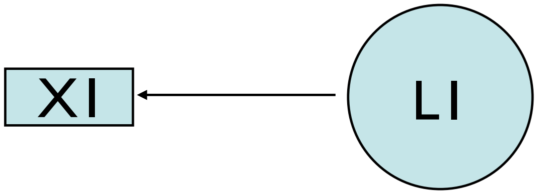
Psychometric Theory: A conceptual Syllabus



Data = Model + Error

- In all of psychometrics and statistics, five questions to ask are:
  - What is the model?
  - How well does it fit?
  - What are the plausible alternative models?
  - How well do they fit?
  - Is this better or worse than the current fit?

A Theory of Data: What can be measured



What is measured?
- Objects
- Individuals

What kind of measures are taken?
- Proximity (- distance)
- Order

What kind of comparisons are made?
- Single Dyads
- Pairs of Dyads

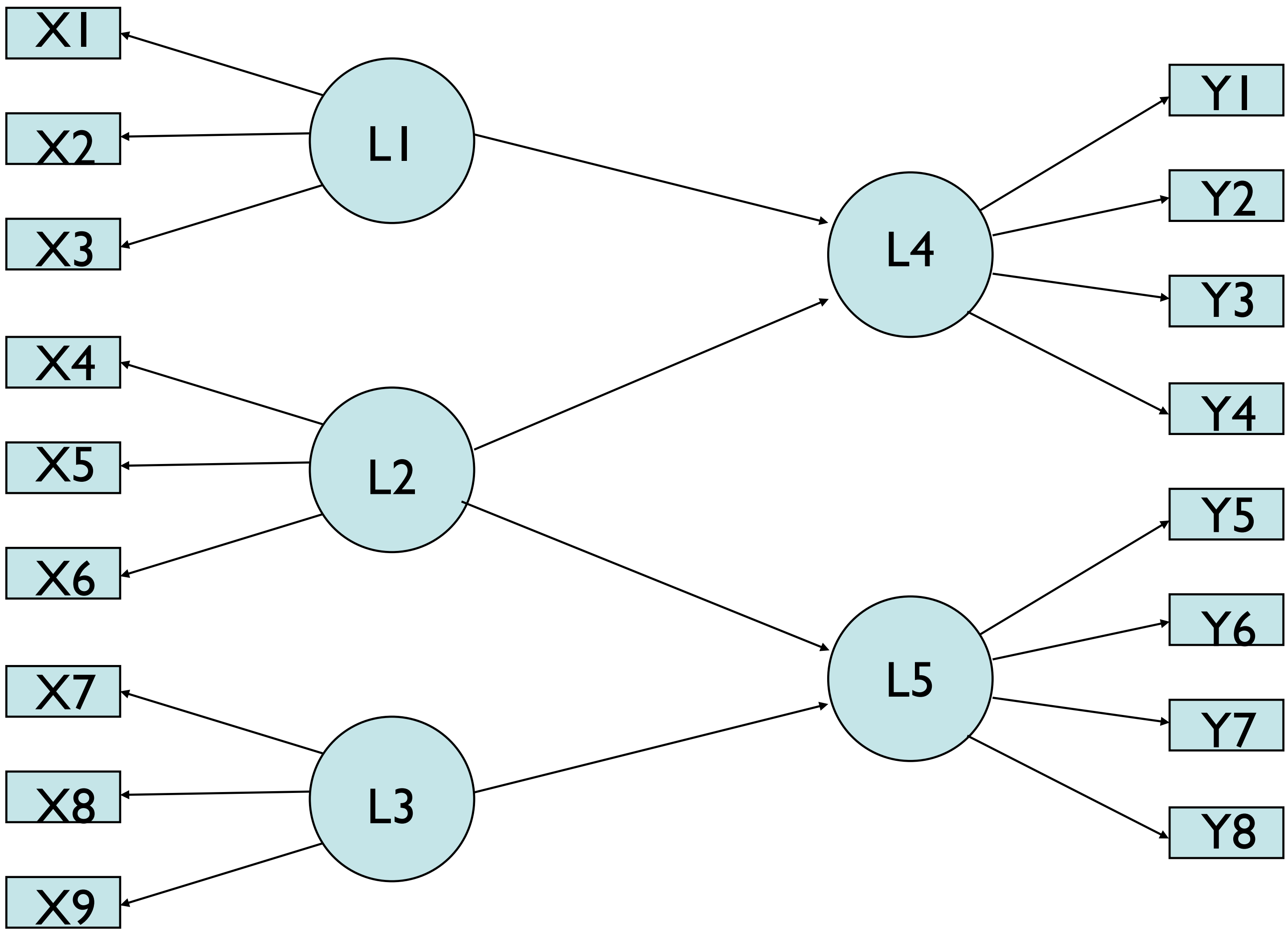
Assigning numbers to observations

2.718282	3,413
3.1415929	86,400
24	31,557,600
37	299,792,458
98.7	$6.022141 * 10^{23}$
365.25	42
365.25636305	X

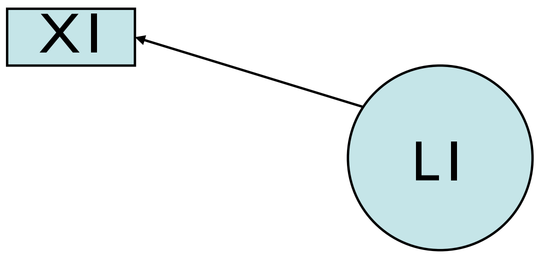
Assigning numbers to observations: order vs. proximity

- Suppose we have observations X, Y, Z
- We assume each observation is a point on an attribute dimension (see Michell for a critique of the assumption of quantity) .
- Assign a number to each point.
- Two questions to ask:
  - What is the order of the points?
  - How far apart are the points?

Psychometric Theory: A conceptual Syllabus



Measurement and scaling



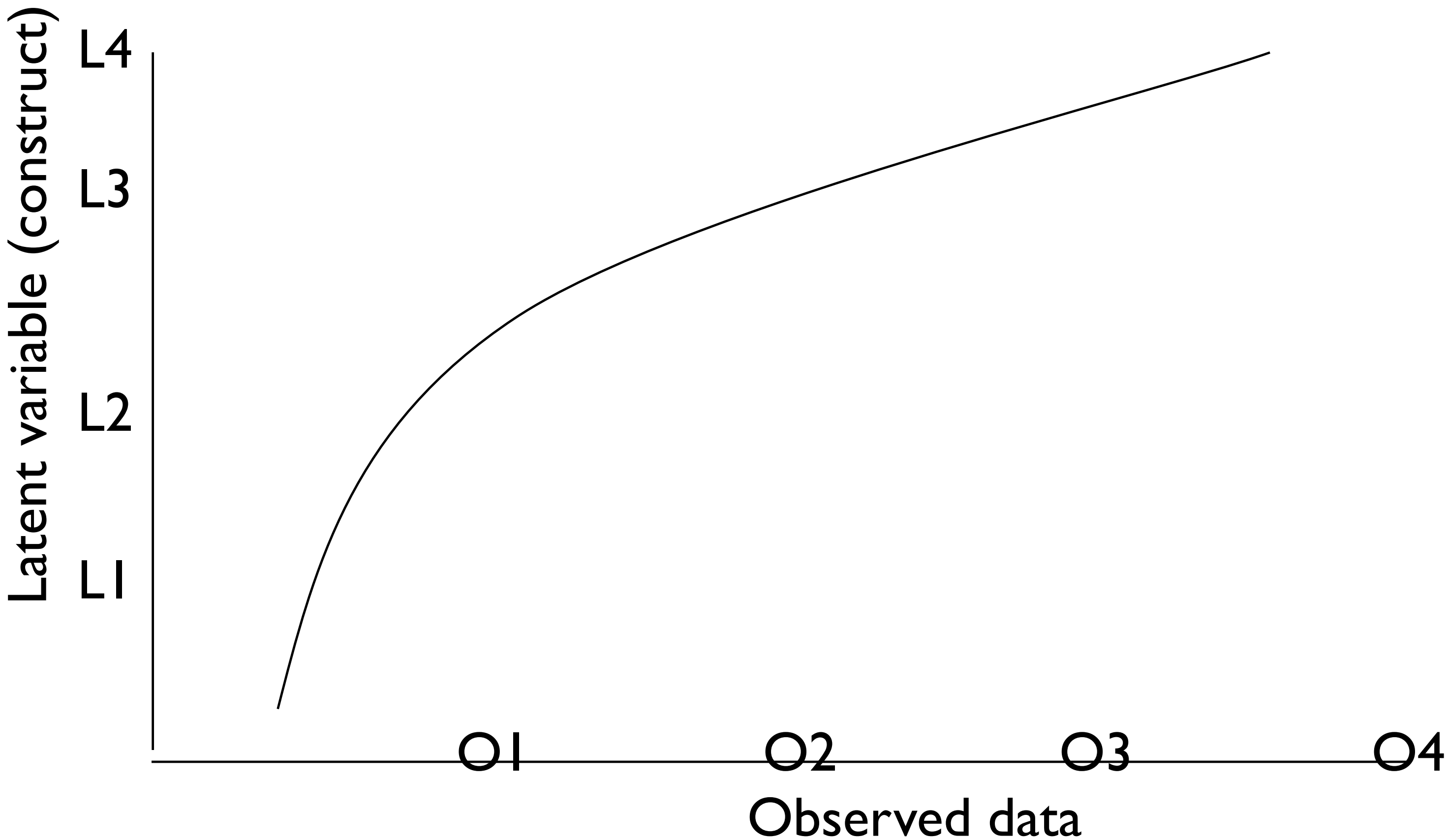
Inferring latent values from observed values

Types of Scales: Inferences from observed variables to Latent variables

- Nominal
- Ordinal
- Interval
- Ratio

- Categories
- Ranks (x > y)
- Differences
  - X-Y > W-V
- Equal intervals with a zero point =>
  - X/Y > W/V

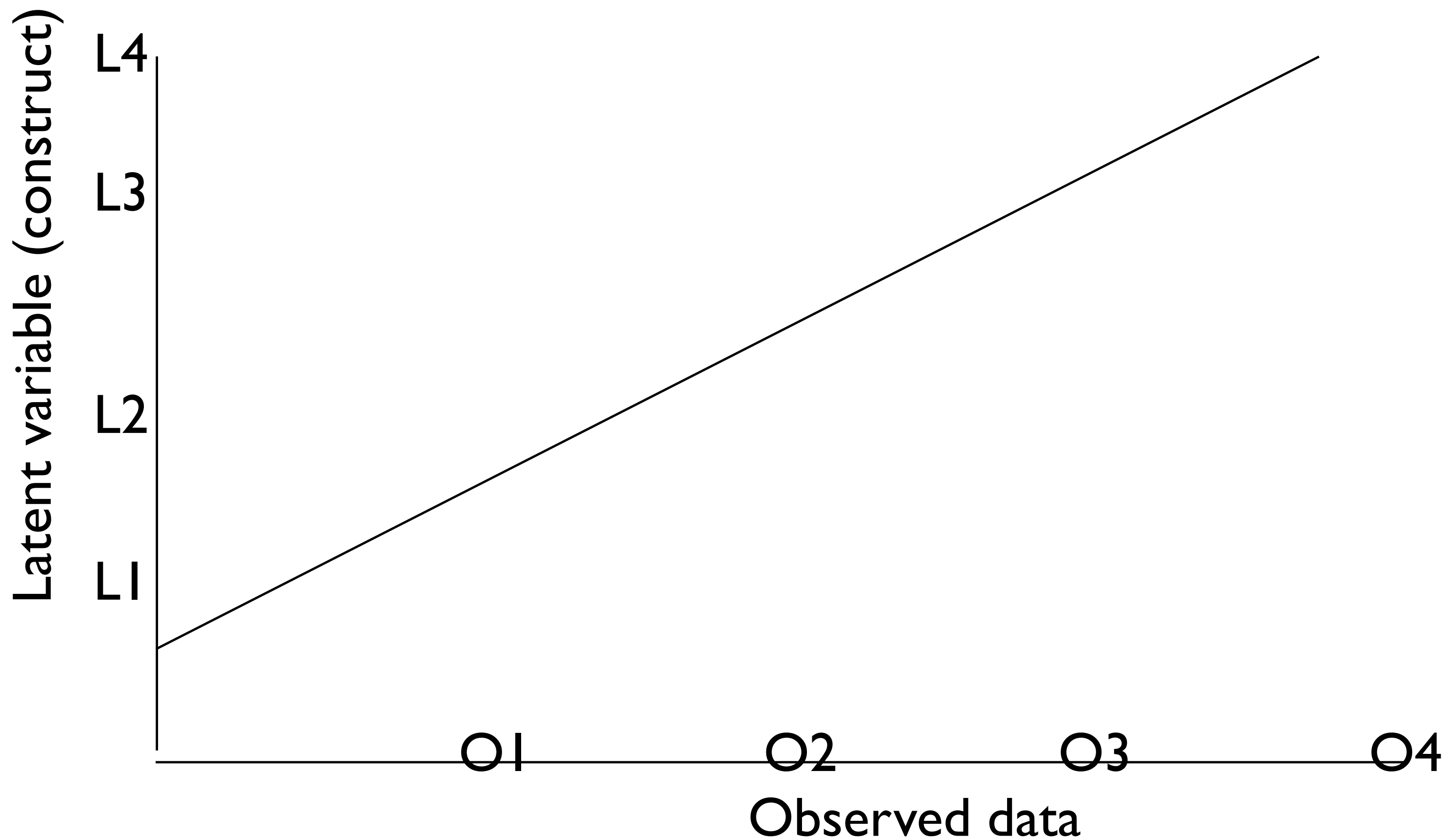
Mappings and inferences



Ordinal Scales

- Any monotonic transformation will preserve order
- Inferences from observed to latent variable are restricted to rank orders
- Statistics: Medians, Quartiles, Percentiles

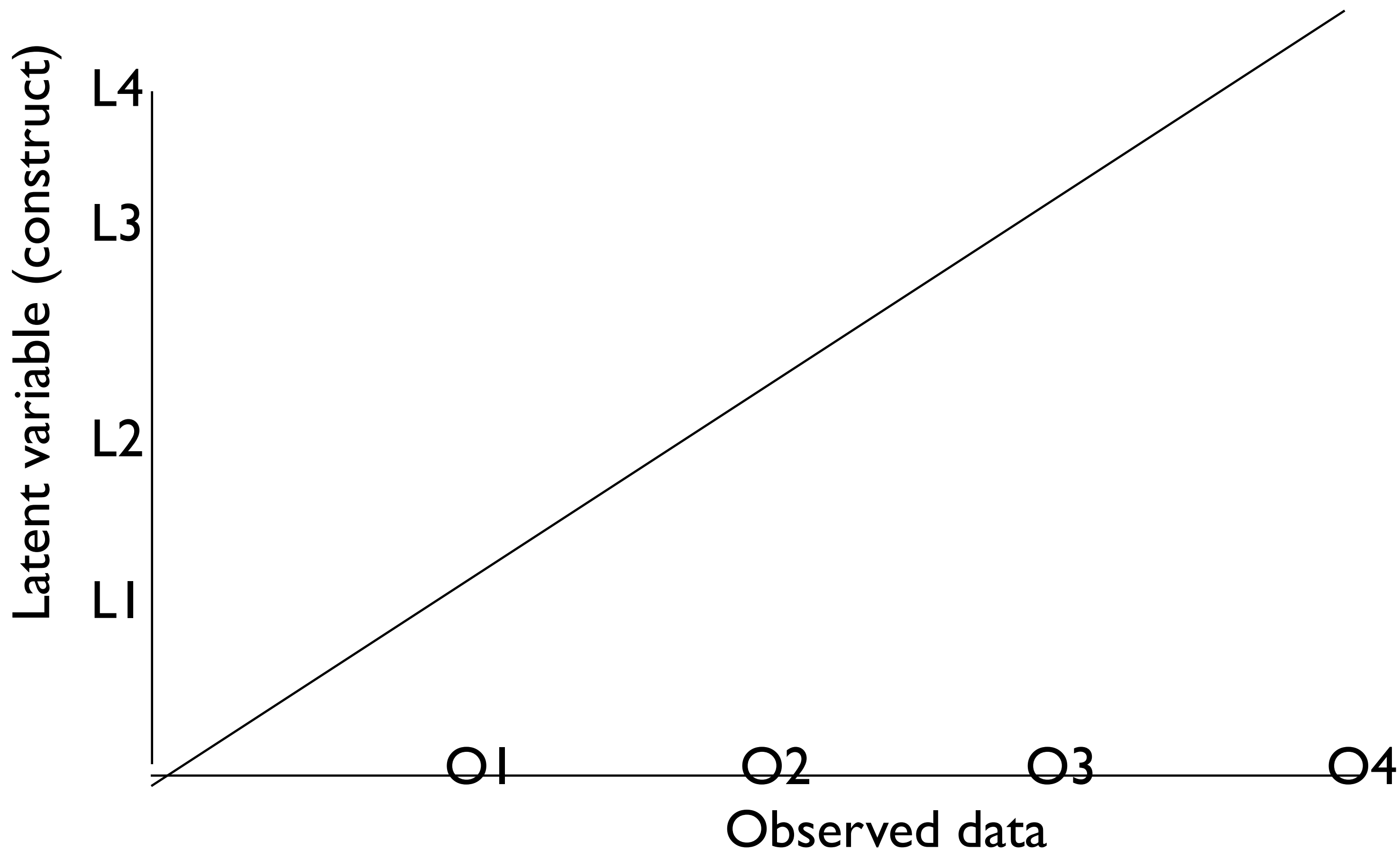
Mappings and inferences



Interval Scales

- Possible to infer the magnitude of differences between points on the latent variable given differences on the observed variable
  X is as much greater than Y as Z is from W
- Linear transformations preserve interval information
- Allowable statistics: Means, Variances

Mappings and inferences



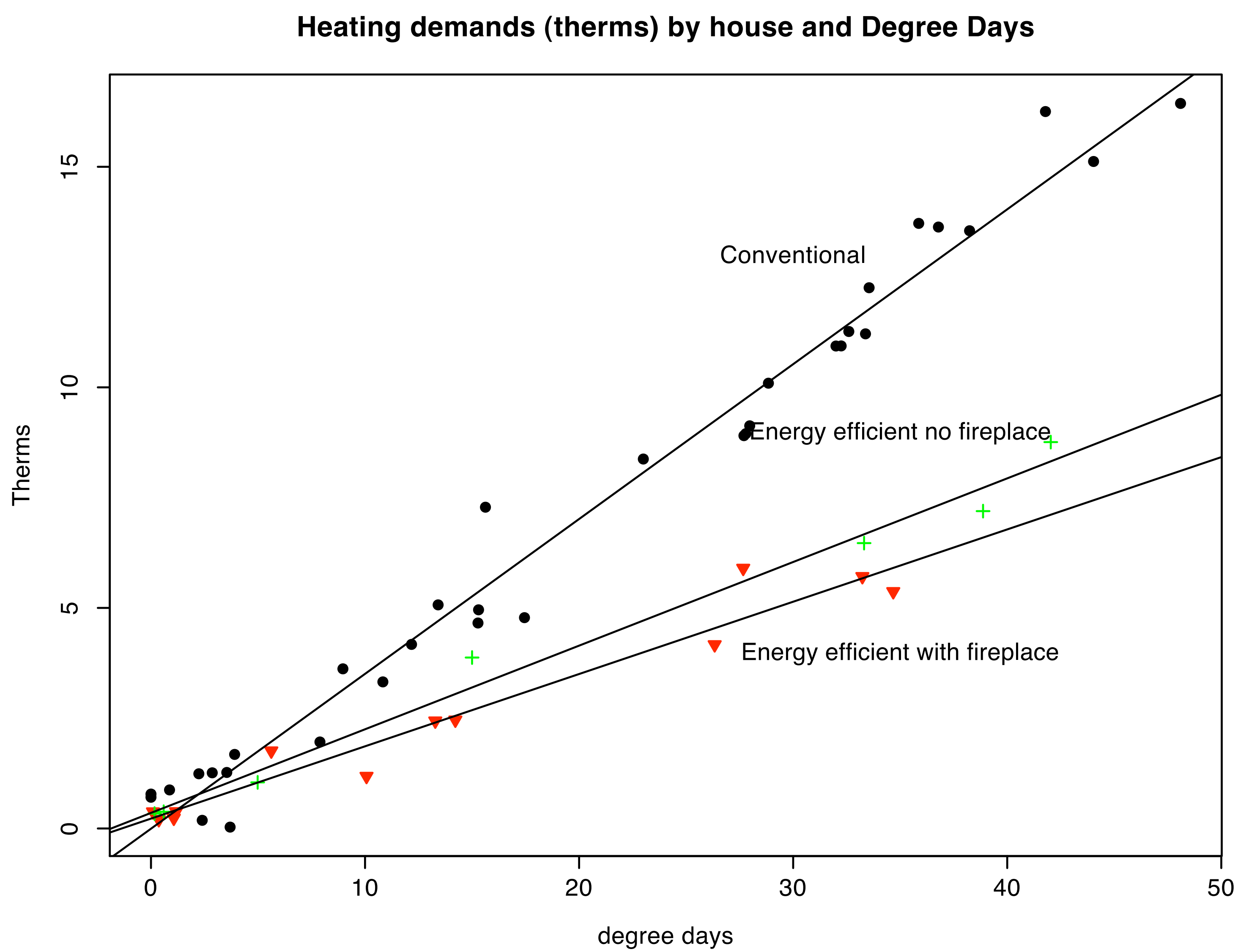
Ratio Scales

- Interval scales with a zero point
- Possible to compare ratios of magnitudes (X is twice as long as Y)

The search for appropriate scale

- Is today colder than yesterday? (ranks)
- Is the amount that today is colder than yesterday more than the amount that yesterday was colder than the day before? (intervals)
  - 50 F - 39 F < 68 F - 50 F
  - 10 C - 4 C < 20 C - 10 C
  - 283K - 277K < 293K -283K
- How much colder is today than yesterday?
  - (Degree days as measure of energy use)
  - K as measure of molecular energy

Gas consumption by degree days (65-T)



Latent and Observed Scores The problem of scale

Much of our research is concerned with making inferences about latent (unobservable) scores based upon observed measures. Typically, the relationship between observed and latent scores is monotonic, but not necessarily (and probably rarely) linear. This leads to many problems of inference. The following examples are abstracted from real studies. The names have been changed to protect the guilty.

Effect of teaching upon performance

A leading research team in motivational and educational psychology was interested in the effect that different teaching techniques at various colleges and universities have upon their students. They were particularly interested in the effect upon writing performance of attending a very selective university, a less selective university, or a two year junior college.

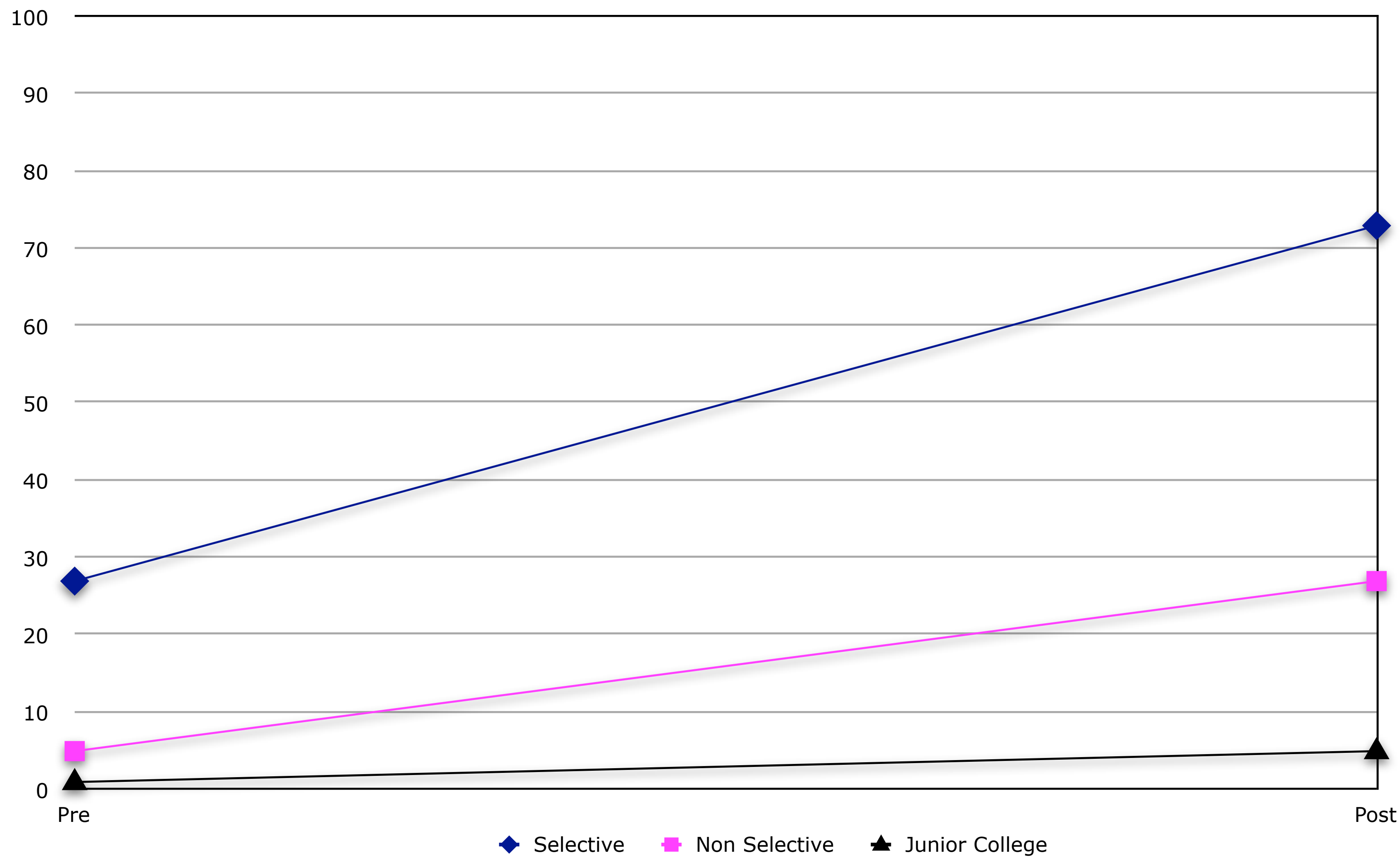
A writing test was given to the entering students at three institutions in the Boston area. After one year, a similar writing test was given again. Although there was some attrition from each sample, the researchers report data only for those who finished one year. The pre and post test scores as well as the change scores were as shown below:

Effect of teaching upon performance

	Pretest	Posttest	Change
Junior College	1	5	4
Non-selective university	5	27	22
Selective university	27	73	45

From these data, the researchers concluded that the quality of teaching at the very selective university was much better and that the students there learned a great deal more. They proposed to study the techniques used there in order to apply them to the other institutions.

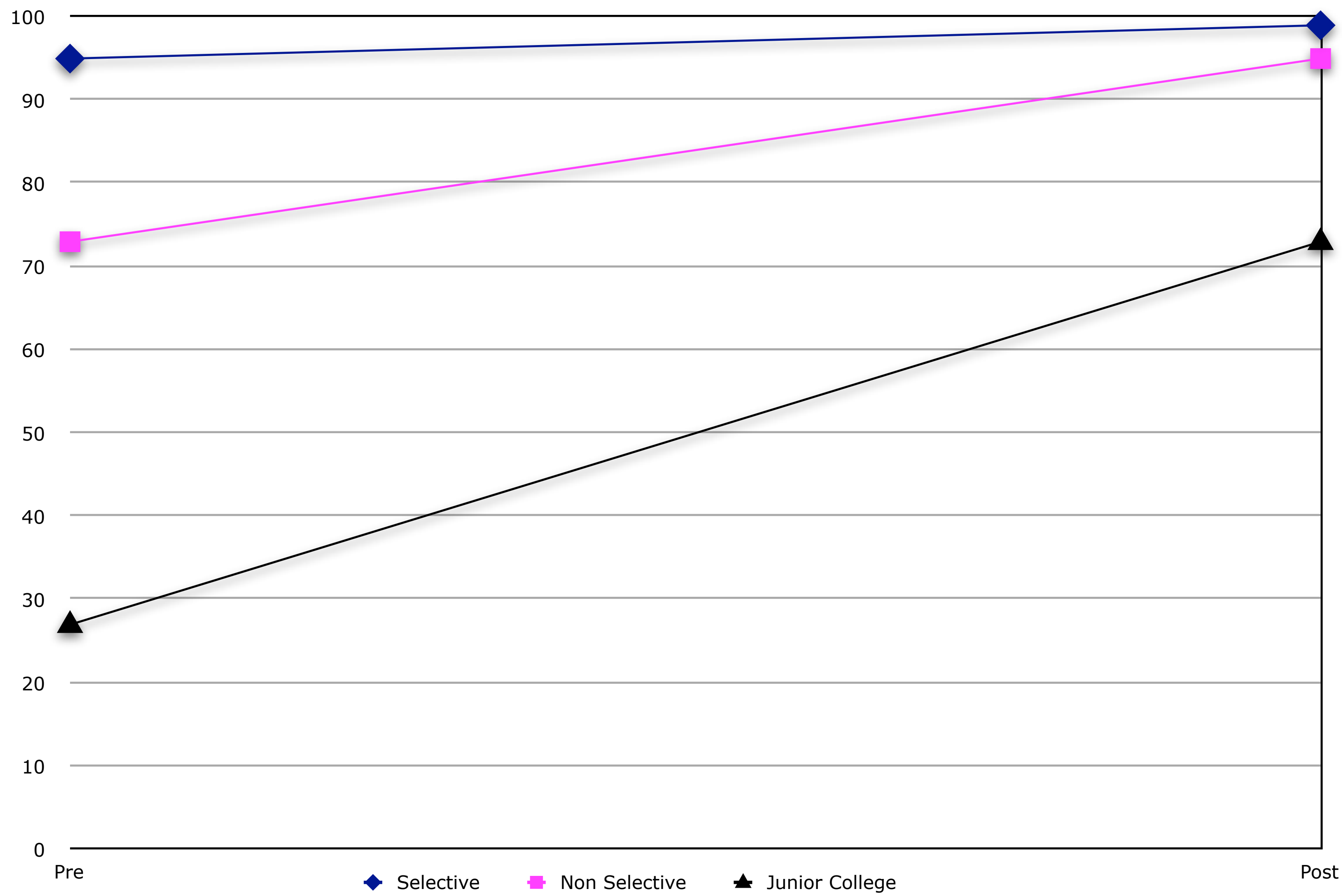
Effect of Teaching upon Performance?



Another research team in motivational and educational psychology was interested in the effect that different teaching techniques at various colleges and universities have upon their students. They were particularly interested in the effect upon mathematics performance of attending a very selective university, a less selective university, or a two year junior college. A math test was given to the entering students at three institutions in the Boston area. After one year, a similar math test was given again. Although there was some attrition from each sample, the researchers report data only for those who finished one year. The pre and post test scores as well as the change scores were:

	Pretest	Posttest	Change
Junior College	27	73	45
Non-selective university	73	95	22
Selective university	95	99	4

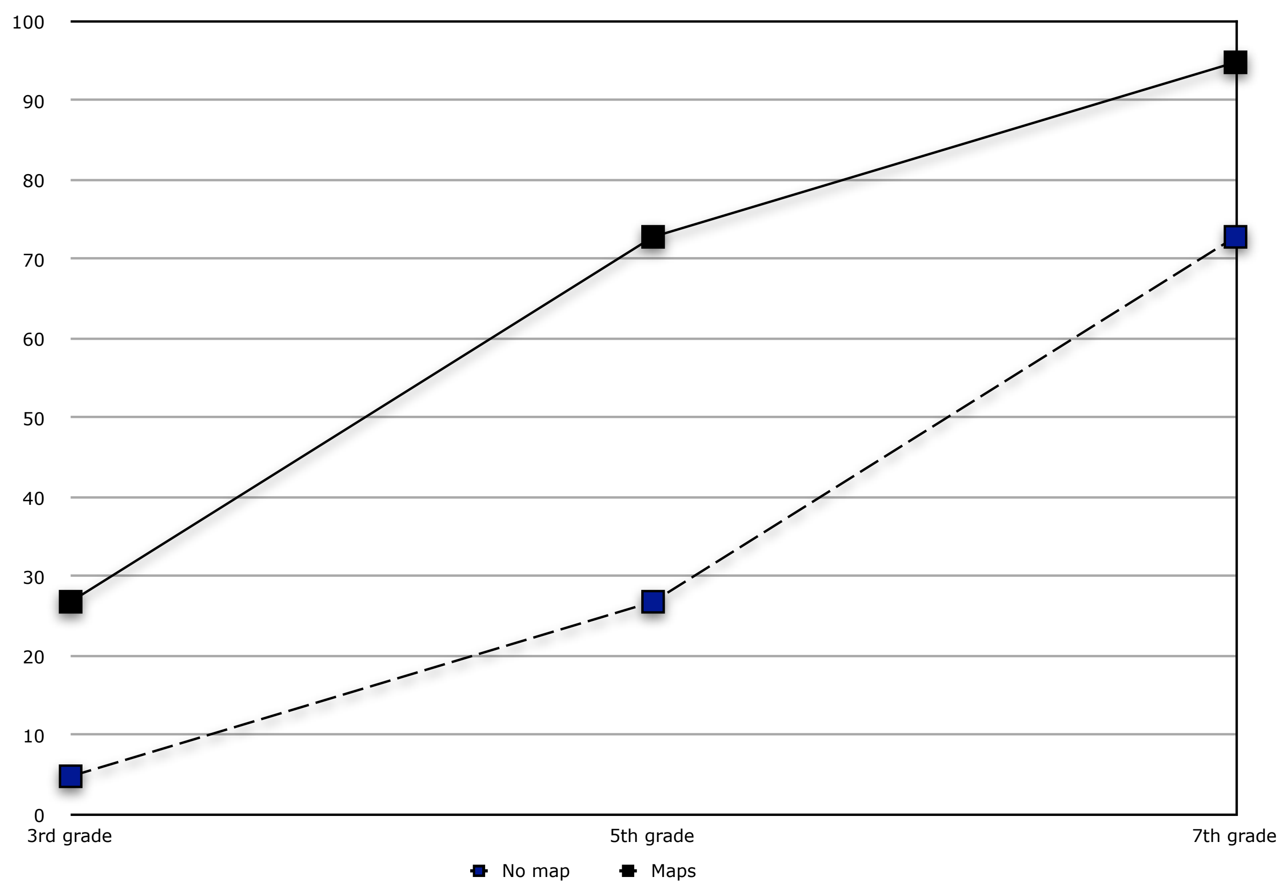
Effect of Teaching upon Performance?



A leading cognitive developmentalist believed that there is a critical stage for learning spatial representations using maps. Children younger than this stage are not helped by maps, nor are children older than this stage. He randomly assigned 3rd, 5th, and 7th grade students into two conditions (nested within grade), control and map use. Performance was measured on a task of spatial recall (children were shown toys at particular locations in a set of rooms and then asked to find them again later. Half the children were shown a map of the rooms before doing the task.

	No map	Maps
3rd grade	5	27
5th grade	27	73
7th grade	73	95

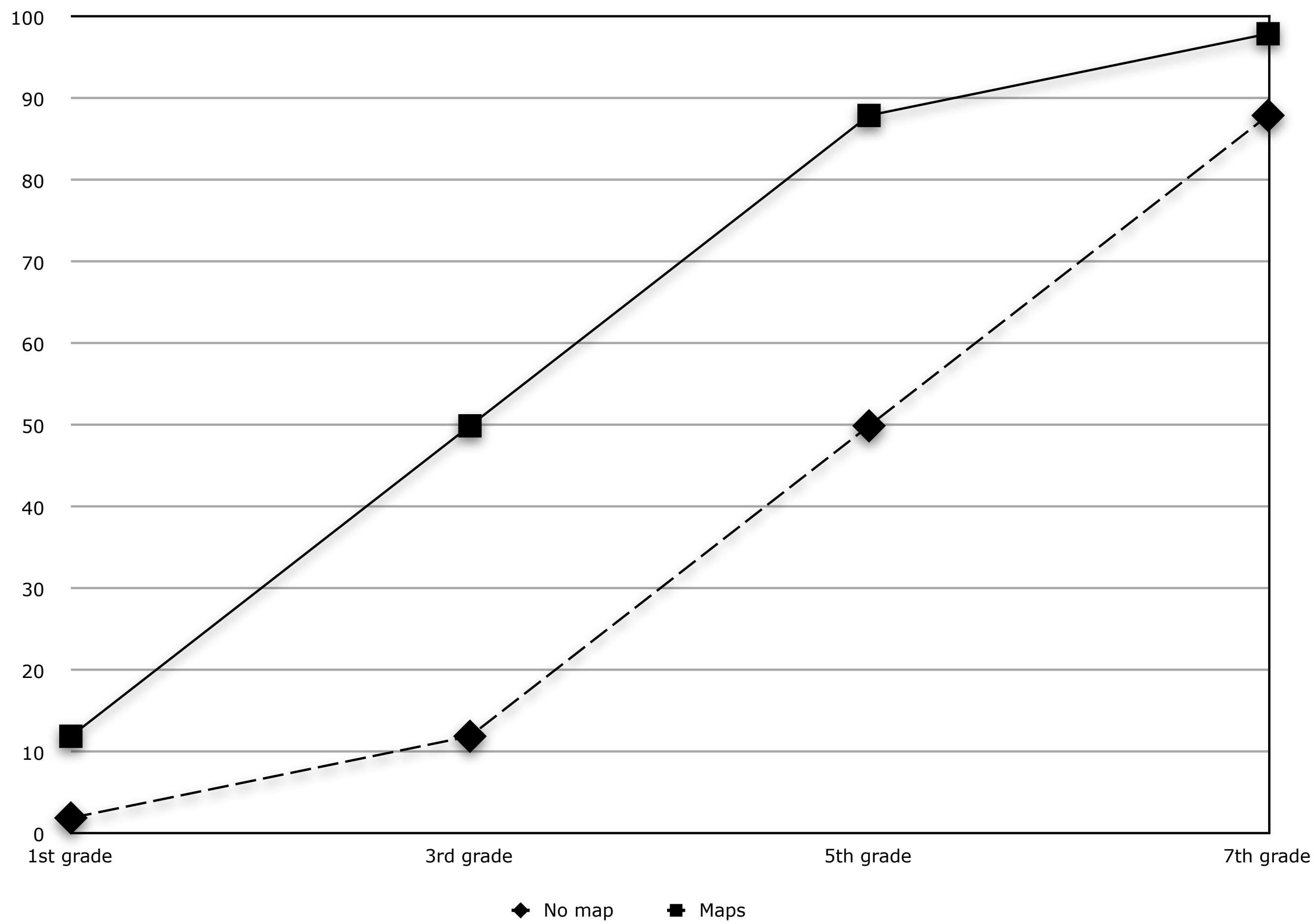
Spatial reasoning facilitated by maps at a critical age



Another cognitive developmentalist believed that there is a critical stage but that it appears earlier than previously thought. Children younger than this stage are not helped by maps, nor are children older than this stage. He randomly assigned 1st, 3rd, 5th, and 7th grade students into two conditions (nested within grade), control and map use. Performance was measured on a task of spatial recall (children were shown toys at particular locations in a set of rooms and then asked to find them again later. Half the children were shown a map of the rooms before doing the task.

	No map	Maps
1st grade	2	12
3rd grade	12	50
5th grade	50	88
7th grade	88	98

Spatial Reasoning is facilitated by map use at a critical age

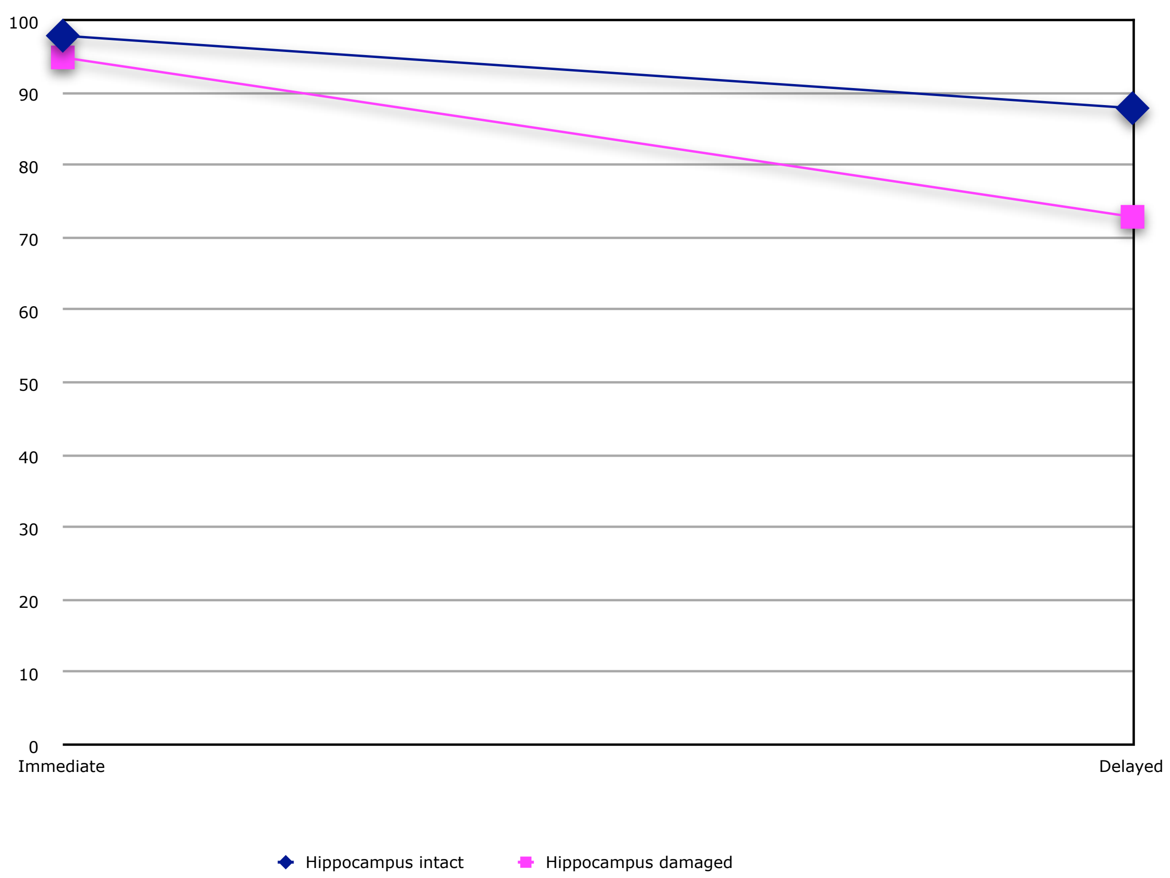


Cognitive-neuro psychologists believe that damage to the hippocampus affects long term but not immediate memory. As a test of this hypothesis, an experiment is done in which subjects with and without hippocampal damage are given an immediate and a delayed memory task. The results are impressive:

	Immediate	Delayed
Hippocampus intact	98	88
Hippocampus damaged	95	73

From these results the investigator concludes that there are much larger deficits for the hippocampal damaged subjects on the delayed rather than the immediate task. The investigator believes these results confirm his hypothesis. Comment on the appropriateness of this conclusion.

Memory = f(hippocampal damange * temporal delay)

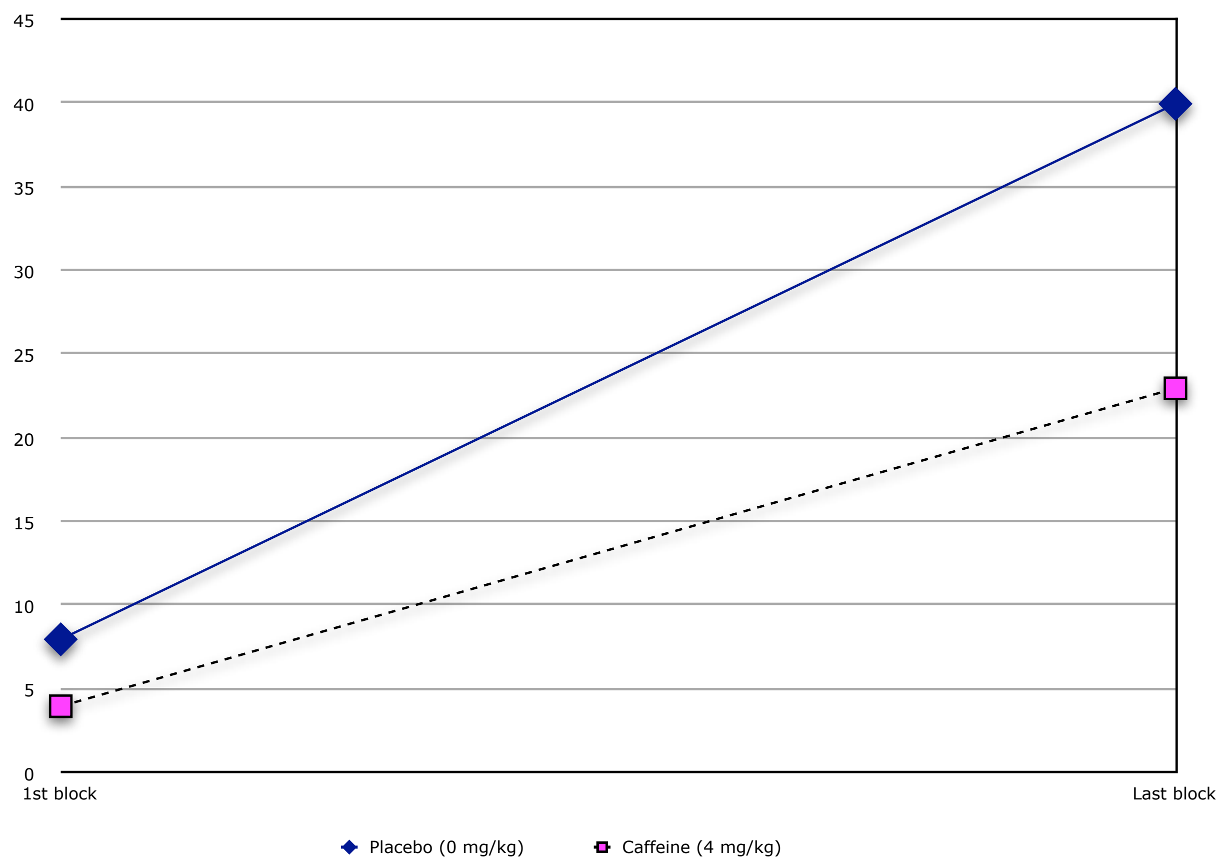


Errors=f(caffeine * time on task)

An investigator believes that caffeine facilitates attentional tasks such that require vigilance. Subjects are randomly assigned to conditions and receive either 0 or 4mg/kg caffeine and then do a vigilance task. Errors are recorded during the first 5 minutes and the last 5 minutes of the 60 minute task. The number of errors increases as the task progresses but this difference is not significant for the caffeine condition and is for the placebo condition.

	1^{st} block	Last block
Placebo (0 mg/kg)	8	40
Caffeine (4 mg/kg)	4	23

Errors=f(caffeine * time on task)



Measuring Arousal

Arousal is a fundamental concept in many psychological theories. It is thought to reflect basic levels of alertness and preparedness. Typical indices of arousal are measures of the amount of palmer sweating. This may be indexed by the amount of electricity that is conducted by the fingertips. Alternatively, it may be indexed (negatively) by the amount of skin resistance of the finger tips. The Galvanic Skin Response (GSR) reflects moment to moment changes, SC and SR reflect longer term, basal levels.

High skin conductance (low skin resistance) is thought to reflect high arousal.

Measuring Arousal

Anxiety is thought to be related to arousal. The following data were collected by two different experimenters. One collected Resistance data, one conductance data.

	Resistance	Conductance
Anxious	2, 2	.5, .5
Low anx	1, 5	1, .2

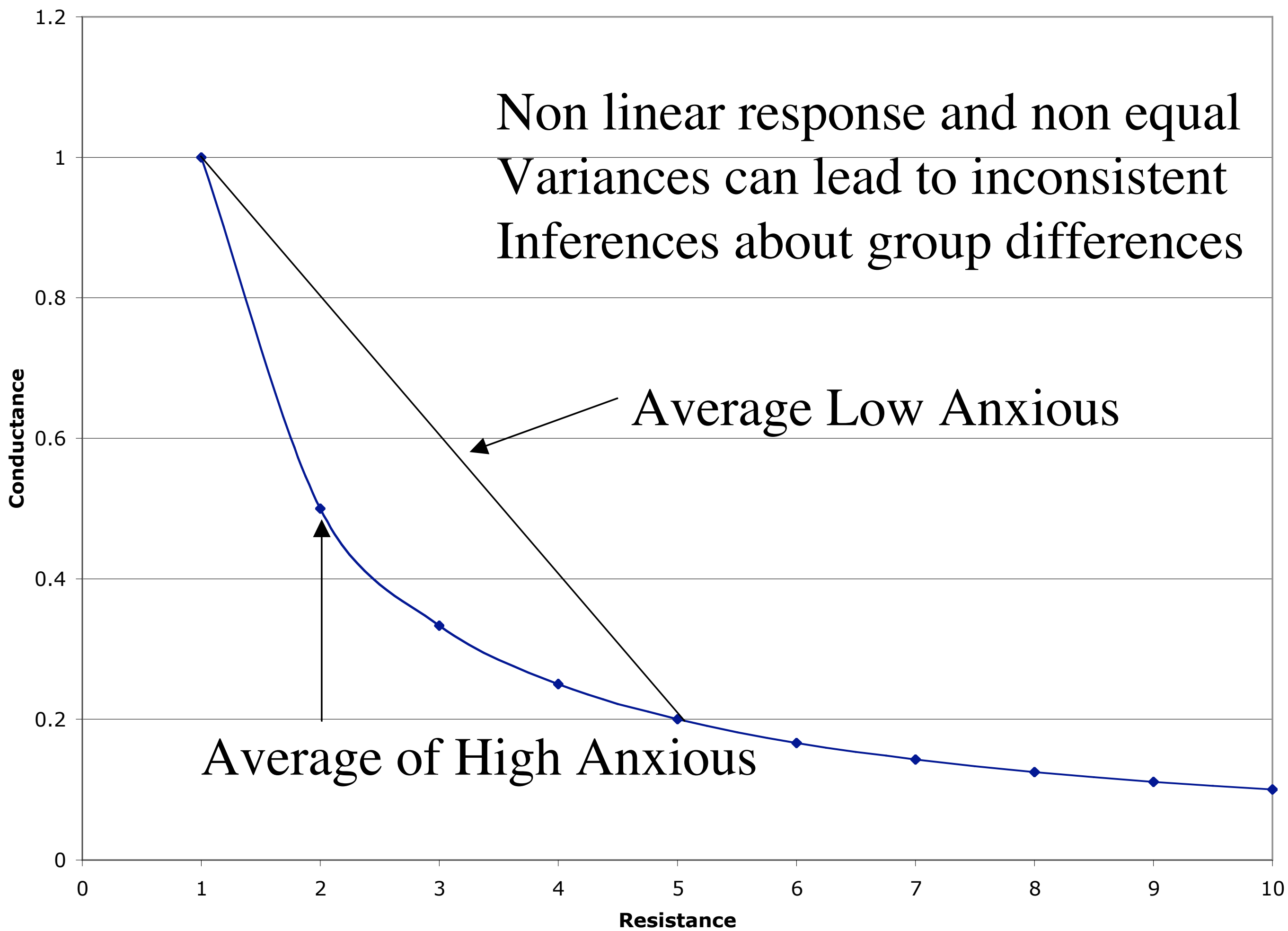
The means were

	Resistance	Conductance
Anxious	2	.5
Low anx	3	.6

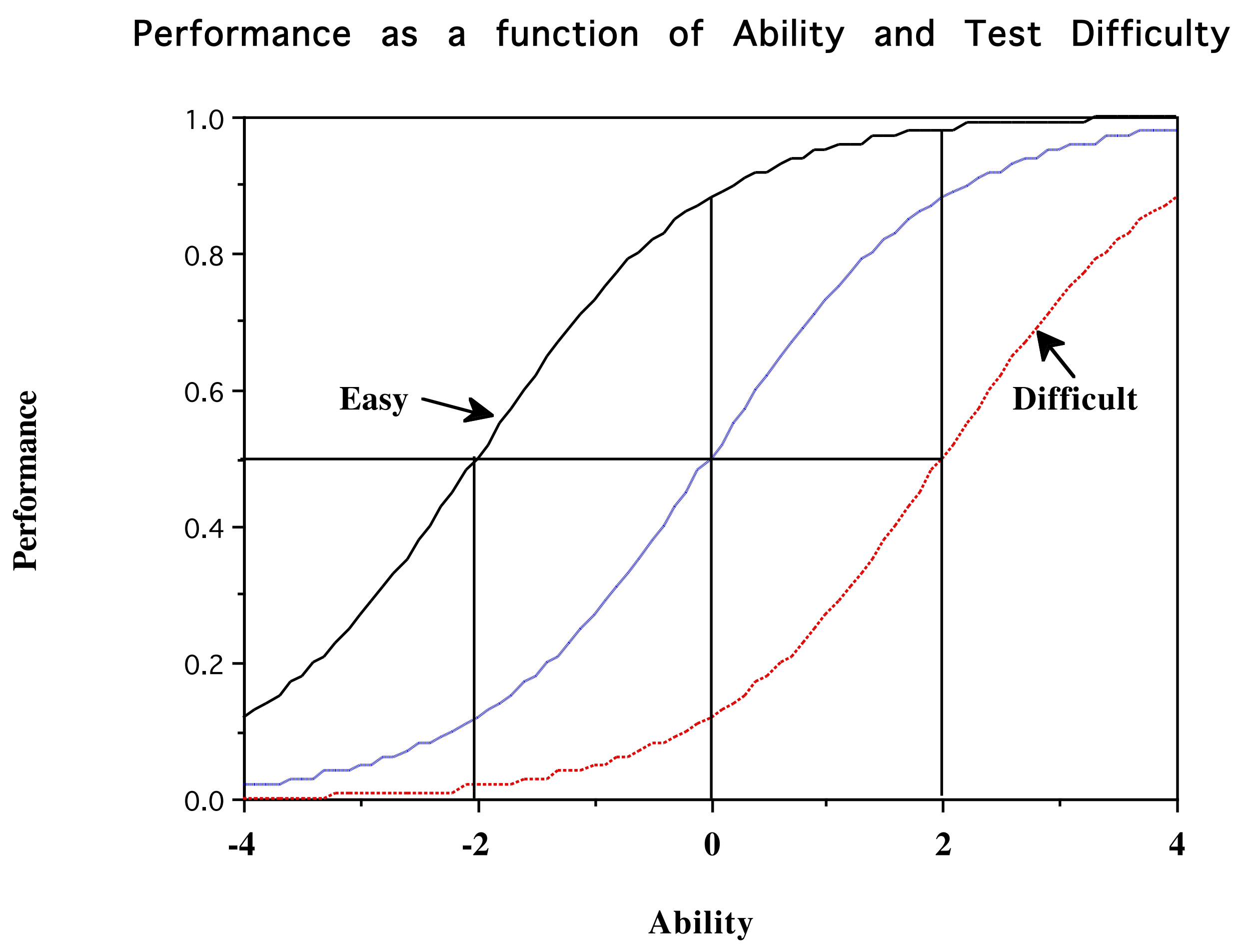
Experimenter 1 concluded that the low anxious had higher resistances, and thus were less aroused. But experimenter 2 noted that the low anxious had higher levels of skin conductance, and were thus more aroused.

How can this be?

Conductance = 1/Resistance



Performance and task difficulty



Performance, ability, and task difficulty

	Difficulty				
	-2	-1	0	1	2
Latent Ability					
-4.00	0.12	0.05	0.02	0.01	0.00
-2.00	0.50	0.27	0.12	0.05	0.02
0.00	0.88	0.73	0.50	0.27	0.12
2.00	0.98	0.95	0.88	0.73	0.50
4.00	1.00	0.99	0.98	0.95	0.88
Change from					
-4 to -2	0.38	0.22	0.10	0.04	0.02
-2 to -0	0.38	0.46	0.38	0.22	0.10
0 to 2	0.10	0.22	0.38	0.46	0.38
2 to 4	0.02	0.04	0.10	0.22	0.38

Performance and Task Difficulty

Note that equal differences along the latent ability dimension result in unequal differences along the observed performance dimension. Compare particularly performance changes resulting from ability changes from -2 to 0 to 2 units.

This is taken from the standard logistic transformation used in Item Response Theory that maps latent ability and latent difficulty into observed scores. IRT attempts to estimate difficulty and ability from the observed patterns of performance.

$$\text{Performance} = 1/(1+\exp^{(\text{difficulty}-\text{ability})})$$

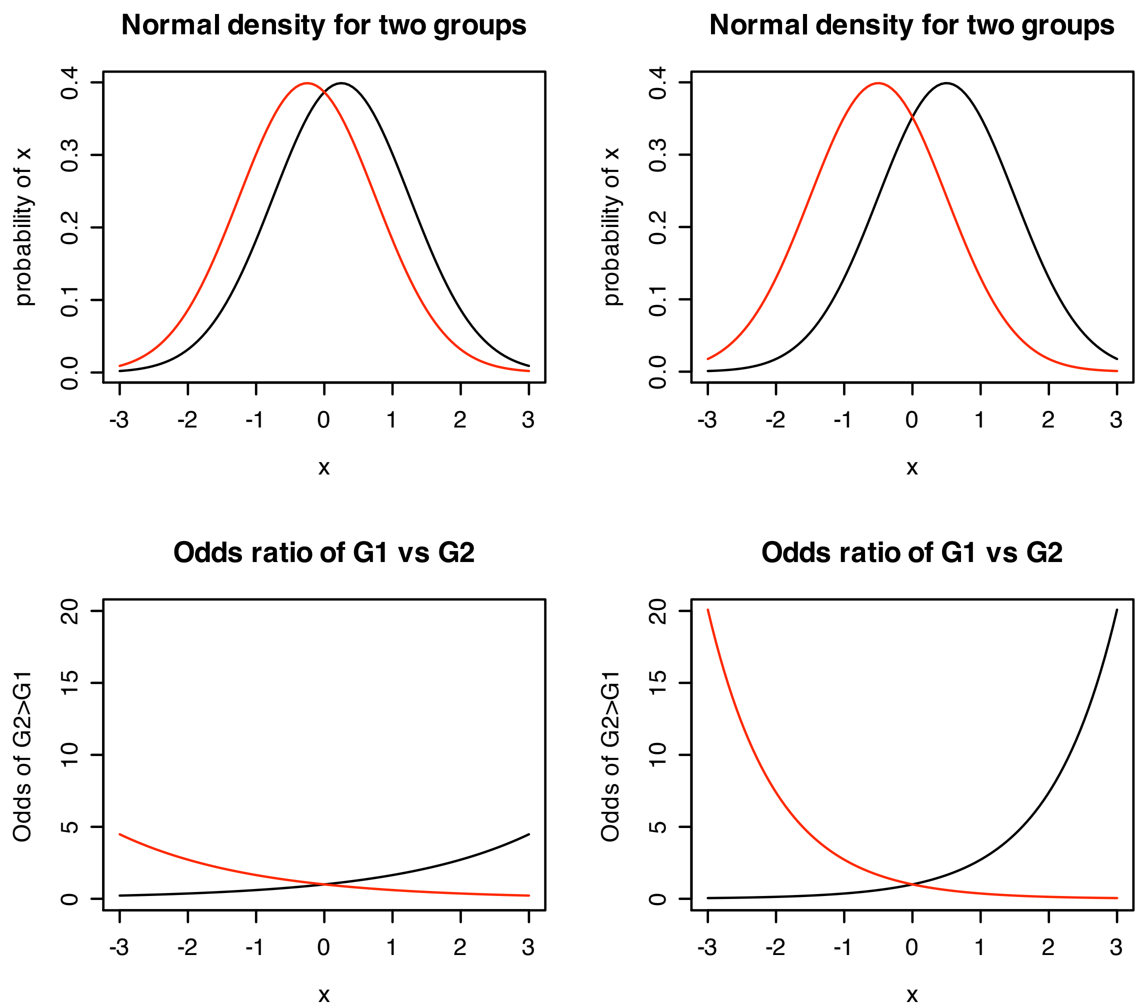
Decision making and the benefit of extreme selection ratios

- Typical traits are approximated by a normal distribution.
- Small differences in means or variances can lead to large differences in relative odds at the tails
- Accuracy of decision/prediction is higher for extreme values.
- Do we infer trait mean differences from observing differences of extreme values?

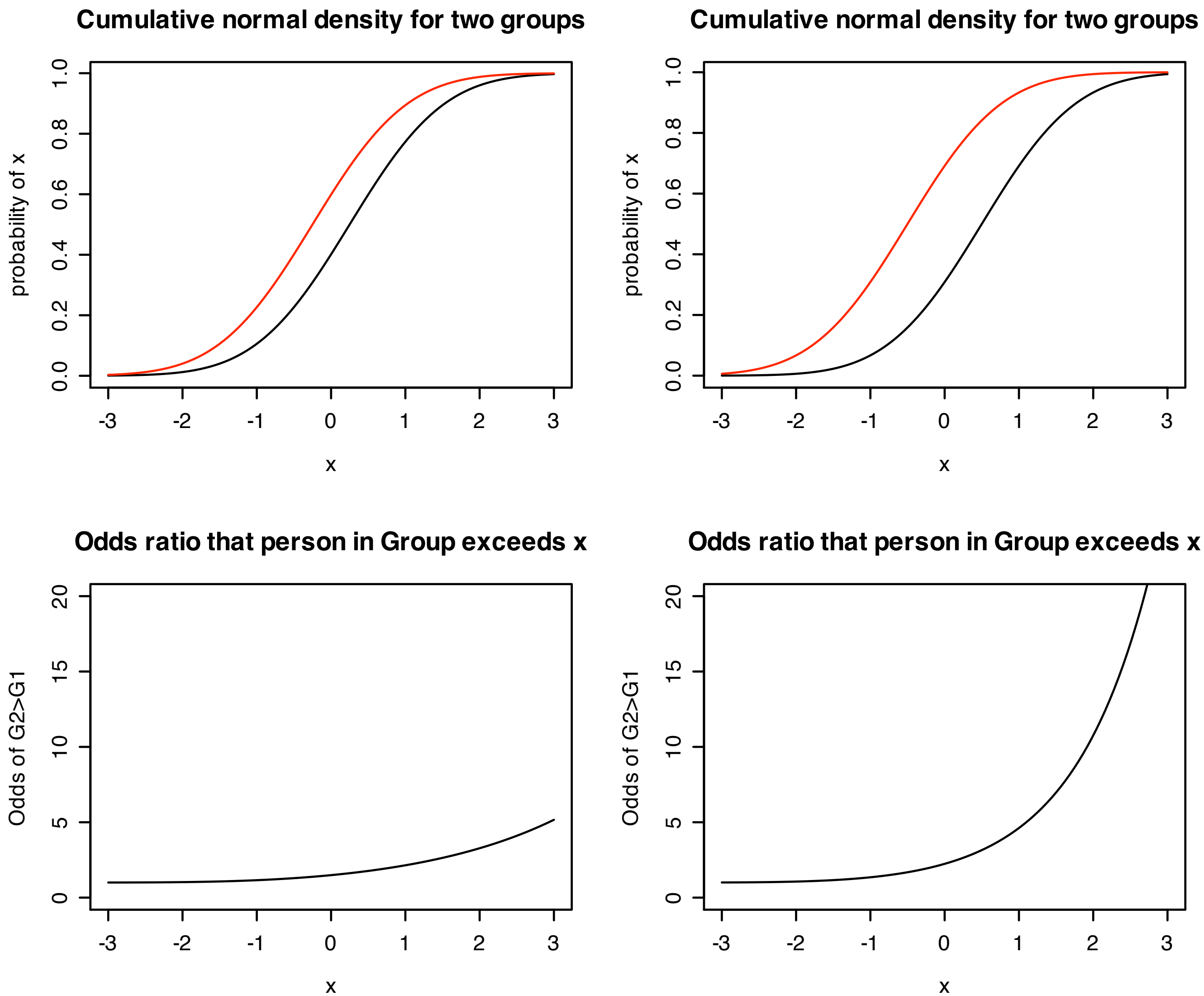
- (code for these graphs at
  - http://personality-project.org/r/extremescores.r)

Odds ratios as f(mean difference, extremity)

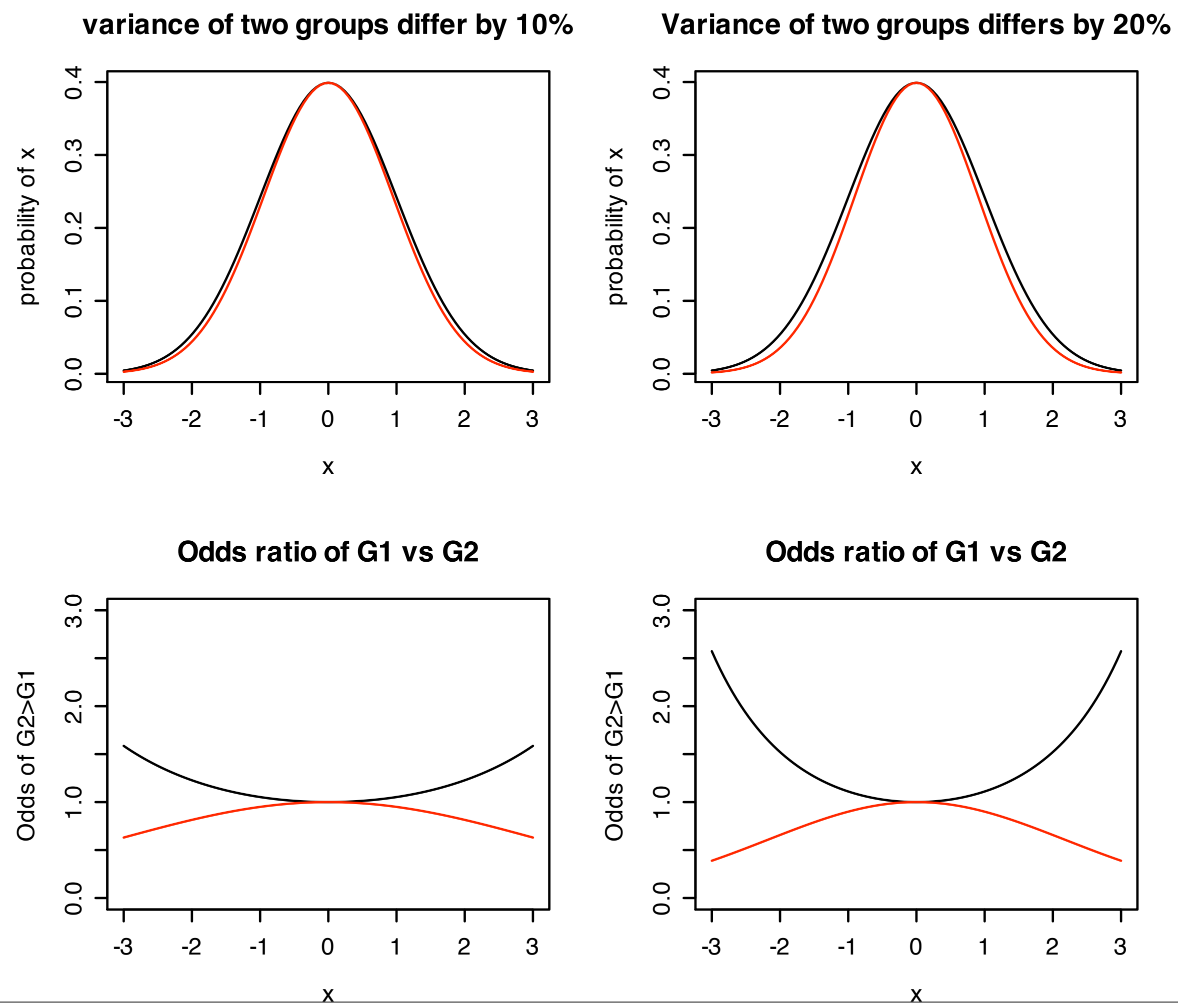
Difference =.5 sigma   Difference =1.0 sigma



The effect of group differences on likelihood of extreme scores



The effect of differences of variance on odds ratios at the tails



Percentiles are not a linear metric and percentile odds are even worse!

- When comparing changes due to interventions or environmental trends, it is tempting to see how many people achieve a certain level (eg., of educational accomplishment, or of obesity), but the magnitude of such changes are sensitive to starting points, particularly when using percentiles or even worse, odds of percentiles.
- Consider the case of obesity:

Obesity gets worse over time

- "Over the last 15 years, obesity in the US has doubled, going from one in 10 to one in five. But the prevalence of *morbid* obesity has quadrupled, meaning that **the number of people 100 pounds overweight has gone from one in 200 to one in 50.** And the number of people roughly 150 pounds overweight has increased by a factor of 5, spiraling from one in 2000 to one in 400."
- "… The fact that super obesity is increasing faster than other categories of overweight suggests a strong environmental component (such as larger portions). If this were a strictly genetic predisposition, the numbers would rise only in proportion to the increase in other weight categories." (Tufts Health Newsletter, Dec. 2003, p 2)

Is obesity getter worse for the super obese? - Seemingly

Label	Definition	Odds	Change in Odds
Obese	BMI = 30 40 lb for 5'5"	1/10 to 1/5	2
Morbid Obese	BMI = 40 100 lb	1/200 to 1/50	4
Super Obese	BMI = 50 150 lb	1/2000 to 1/400	5

Is obesity getter worse for the super obese? -- No

Label	Definition	Odds	Change in Odds	z score	Change in z
Obese	BMI = 30	1/10 to 1/5	2	-1.28 -.84	0.44
Morbid Obese	BMI = 40	1/200 to 1/50	4	-2.58 -2.05	0.53
Super Obese	BMI = 50	1/2000 to 1/400	5	-3.29 -2.81	0.48

Psychometric Theory: A conceptual Syllabus