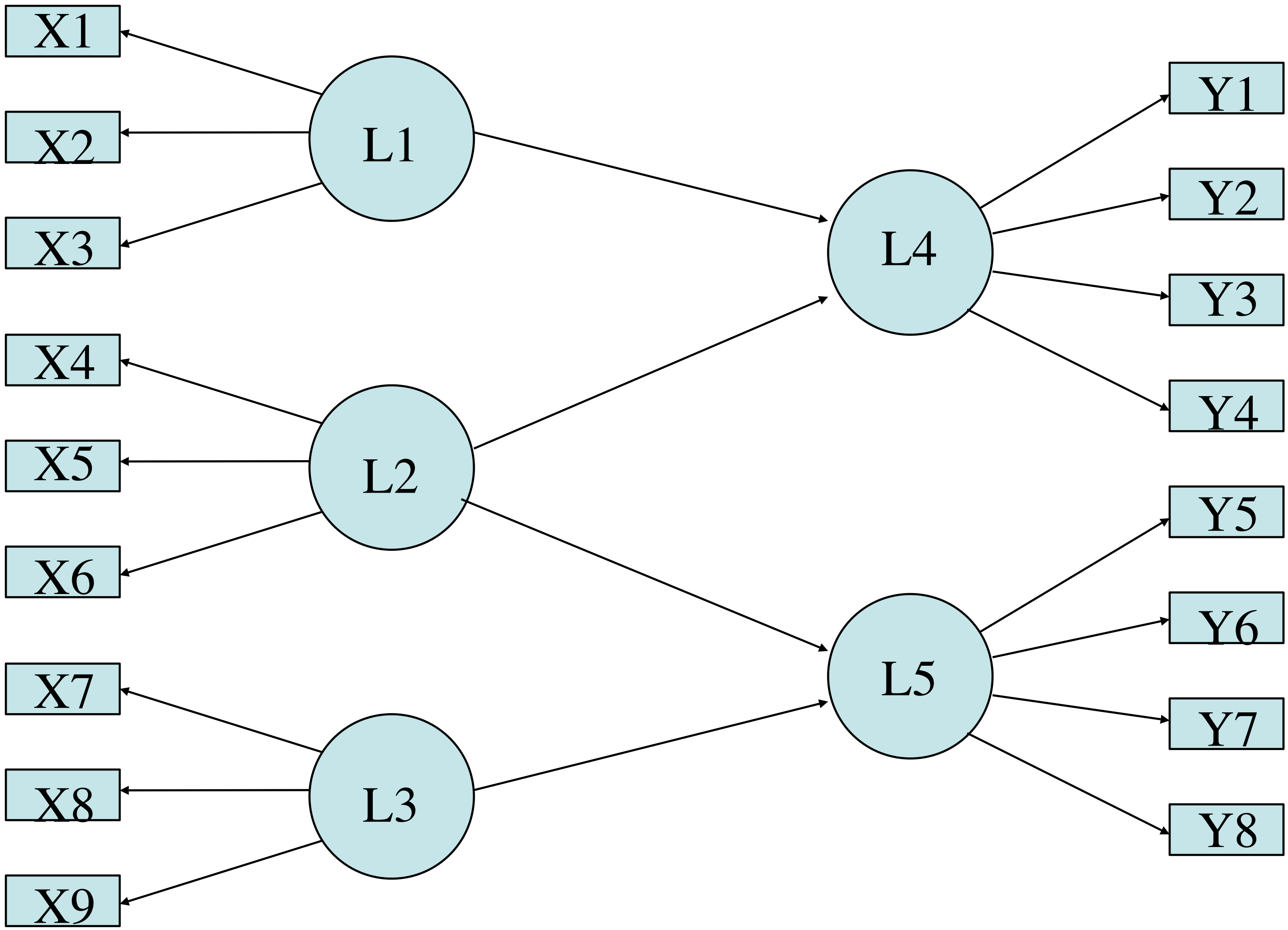
Psychometric Theory
Data = Model + Error

A summary of models of measurement

Psychometric Theory: A conceptual Syllabus



Data = Model + Error

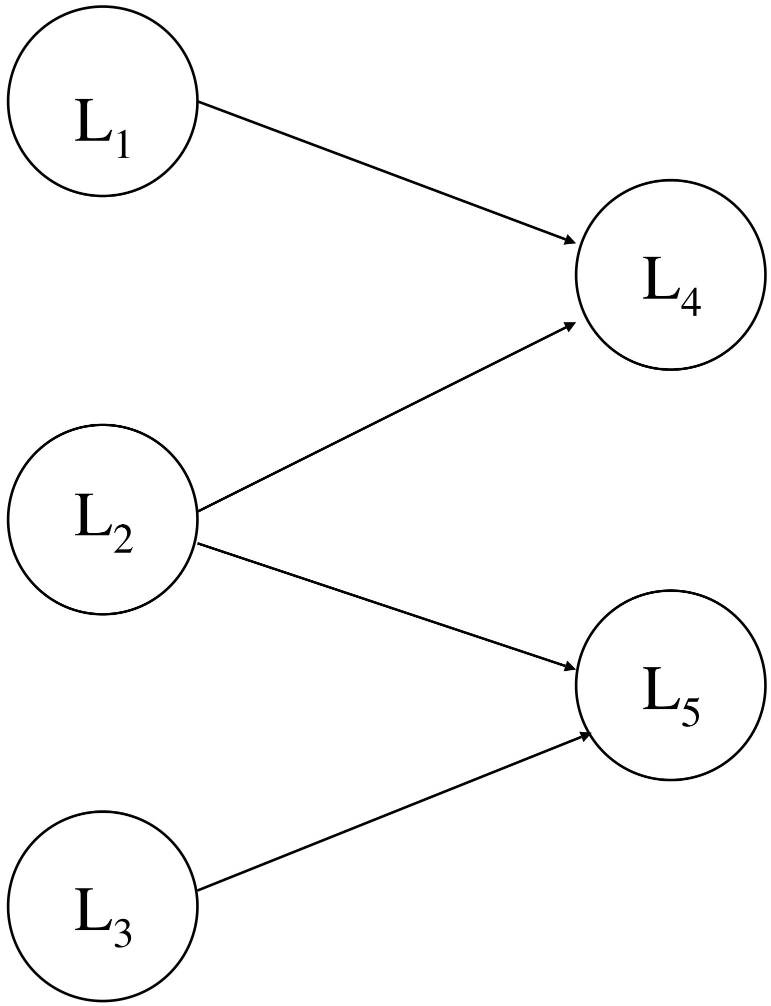
- Statistics are smooths (models) of data
- Models are idealized representations of data.
- Models are projections of data from a higher order space into a lower order space.
- Lack of model fit (Data-Model) is Error (for that model).
- Models differ in complexity and fit
- Review the models presented

Models as functions of data

- Data = Model + Error
- Error = Data - Model
- Observations = f(Model) + error
- Model = f^{-1}(Observations) + error

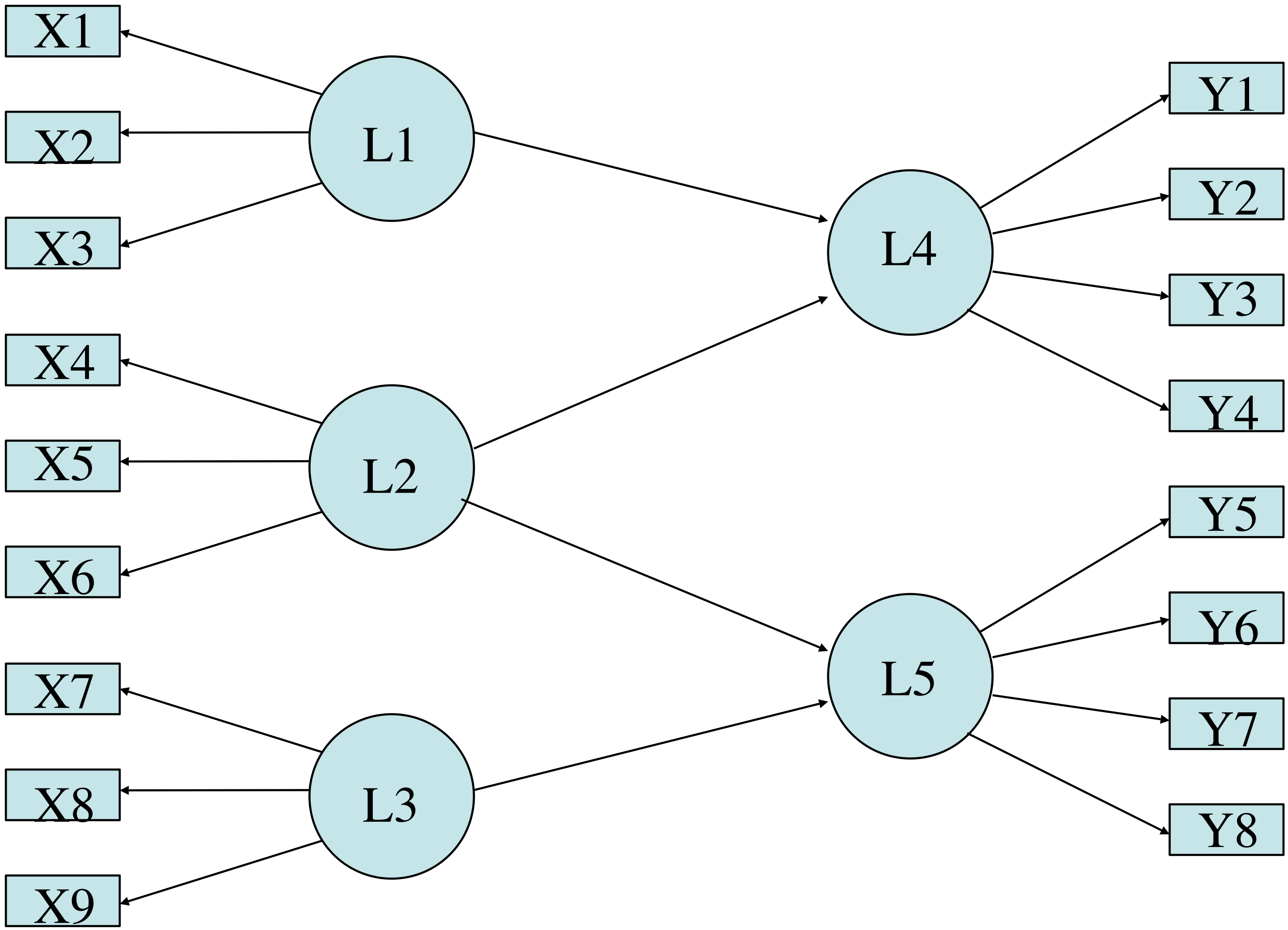
Observed Variables

Latent Variables



Theory as network of constructs

Psychometric Theory: A conceptual Syllabus



Three models of Psychological Research

	B=f(Person)	B=f(Environment)
Method/ Model	Correlational Observational Biological/field	Experimental Causal Physical/lab
Statistics	Variance Dispersion Correlation/ Covariance	Mean Central Tendency t-test, F test
Effects	Individuals Individual Differences	Situations General Laws
	B=f(P,E) Effect of individual in an environment Multivariate Experimental Psychology	

Kinds of measures: Theory of Data

X_4	Ordering of individuals
X_5	Ordering of stimuli
X_6	Attitudes and preferential choice

Coombs typology of data

	Single Dyads	Pairs of Dyads
Single Stimuli	Measurement (S*O) $s_i < o_j$ (Abilities) $\lvert s_i - o_j \rvert < d$ (attitudes)	$\lvert s_i - o_j \rvert <$ $\lvert s_k - o_l \rvert$ Unfolding (S*O)*(S*O) Preferential choice
Pairs of Stimuli	Scaling of stimuli (O * O) $o_i < o_j$	MDS (O*O)*(O*O) $\lvert o_i - o_j \rvert < \lvert o_k - o_l \rvert$
		Individual differences in Multidimensional Scaling S * (O*O) * (O*O)

Thurstonian Scaling of Stimuli

- What is scale location of objects I and J on an attribute dimension D?
  - Assume that object I has mean value m_i with some variability.
  - Assume that object J has a mean value m_j
  - Assume equal and normal variability (Thurstone case 5)
    - Less restrictive assumptions are cases 1-4)
- Observe frequency of ($o_i < o_j$)
  - Convert relative frequencies to normal equivalents
- Result is an interval scale with arbitrary 0 point

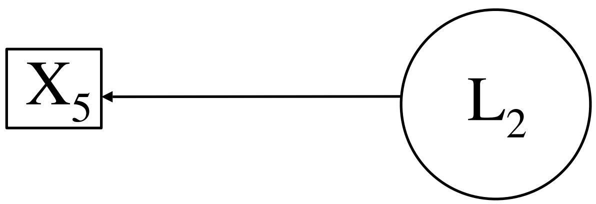
Multidimensional scaling

- Given a n * n distance matrix, is it possible to represent the data in a k dimensional space?
  - $d_{ij} = \mathrm{sqrt}\{\sum(o_{ik} - o_{jk})^2\}$
- How well does that model fit?
- How sensitive is the model to transformations of the original distances?
- Need to find distances
  - absolute distance between pairs
  - ranks of distances between pairs of pairs

Models of data

- Probability of judging A > B = f(A-B)
  - Probability of endorsement given ability and item difficulty = f(difficulty-ability)
  - Strength of desire = f(| self - item |)
- Person P prefers A over B if |P -A| < |P-B|
  - Probability that Person P prefers A over B = f(|P-A| - |P-B|)

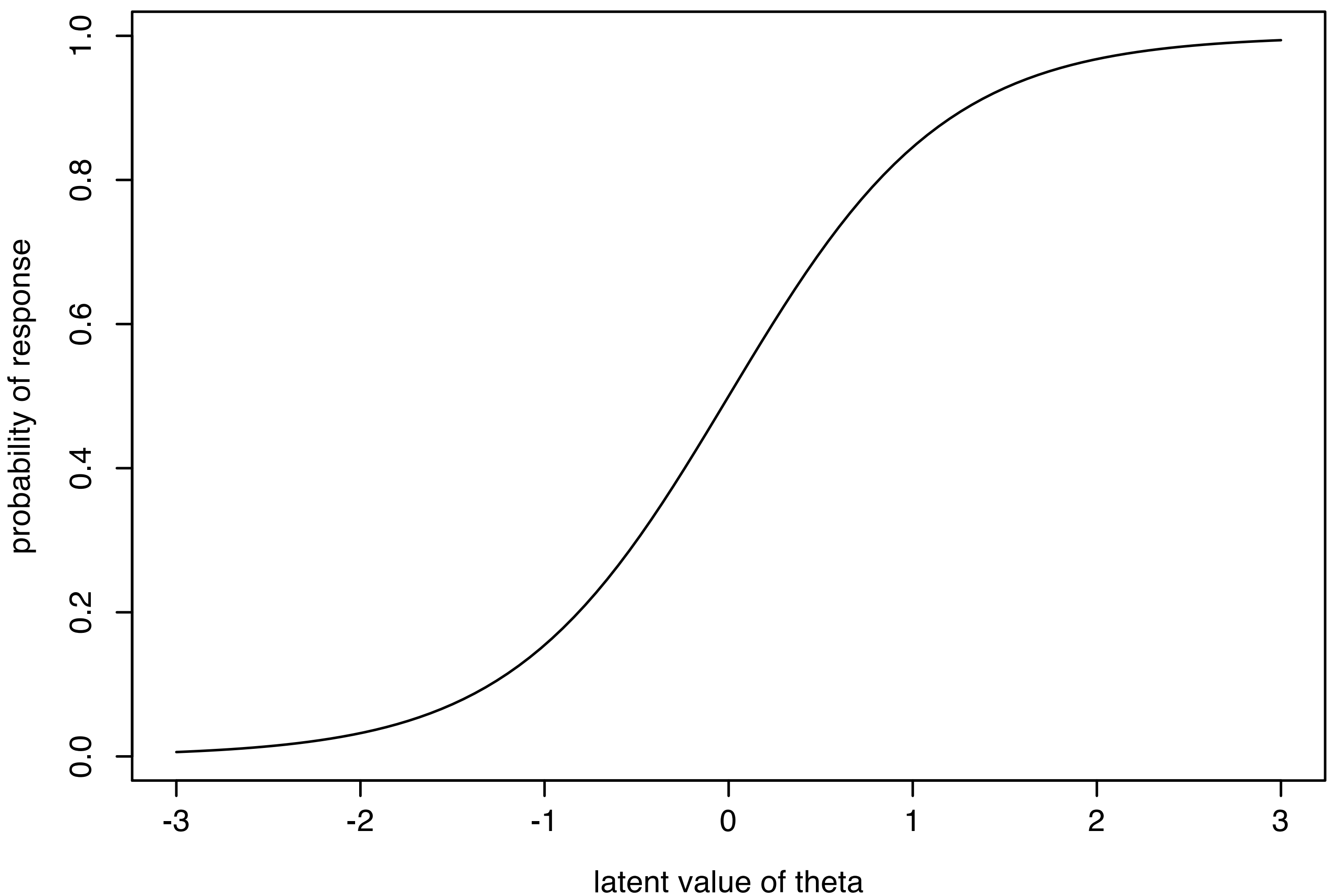
Observed variable -Latent variable relationships: Problems of scaling



Types of relationship

Levels of measurement

$$X = 1/(1+\exp(-1.7*\text{theta}))$$



Models of scaling

- Ordinal $f(A)>f(B) <=> A > B$
- Interval $f(A-B) > f(C-D) <=> (A-B)>(C-D)$
- Ratio $f(A/B)>f(C/D) <=> A/B > C/D$

Models of scaling

- Ordinal   f(A)>f(B) <=> A > B
- Interval f(A-B) > f(C-D) <=> (A-B)>(C-D)
- Ratio f(A/B)>f(C/D) <=> A/B > C/D
- Why is this important? Examples from ANOVA
- Do interactions of observed scores <=> Interactions of latent scores?   That is,
  – does f(A-B) > f(C-D) <=> (A-B)>(C-D)
  – Only when we have interval metrics or better

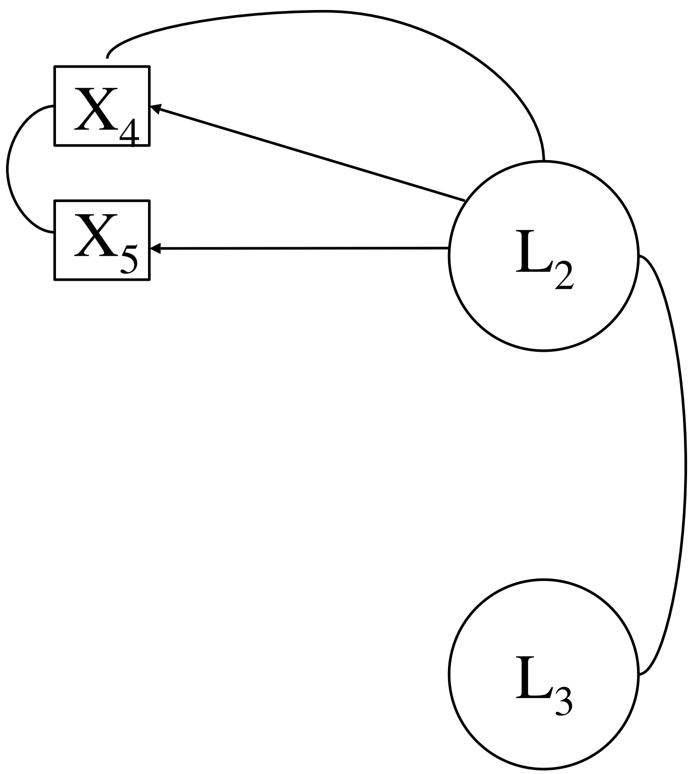
Measurement and Interactions

Data			Model		
A	B		α	β	
C	D		χ	δ	
Interaction = (A-B) - (C-D)			Interaction = $(\alpha-\beta) - (\chi-\delta)$		

Models of aggregation

- Central tendency of a set of scores
- Mean = $\sum x_i/N$
- Other forms of mean are anti-transformed means of transformed data
  - Geometric Mean = $\exp(\sum \ln(x_i)/N)$
  - Harmonic Mean = $1/\sum(1/x_i)/N$
- Mean as expected score when aggregating across observations, but expectation across which observations?
- Consider more robust estimates: median, trimmed mean

Correlation and Regression



Observed - Observed

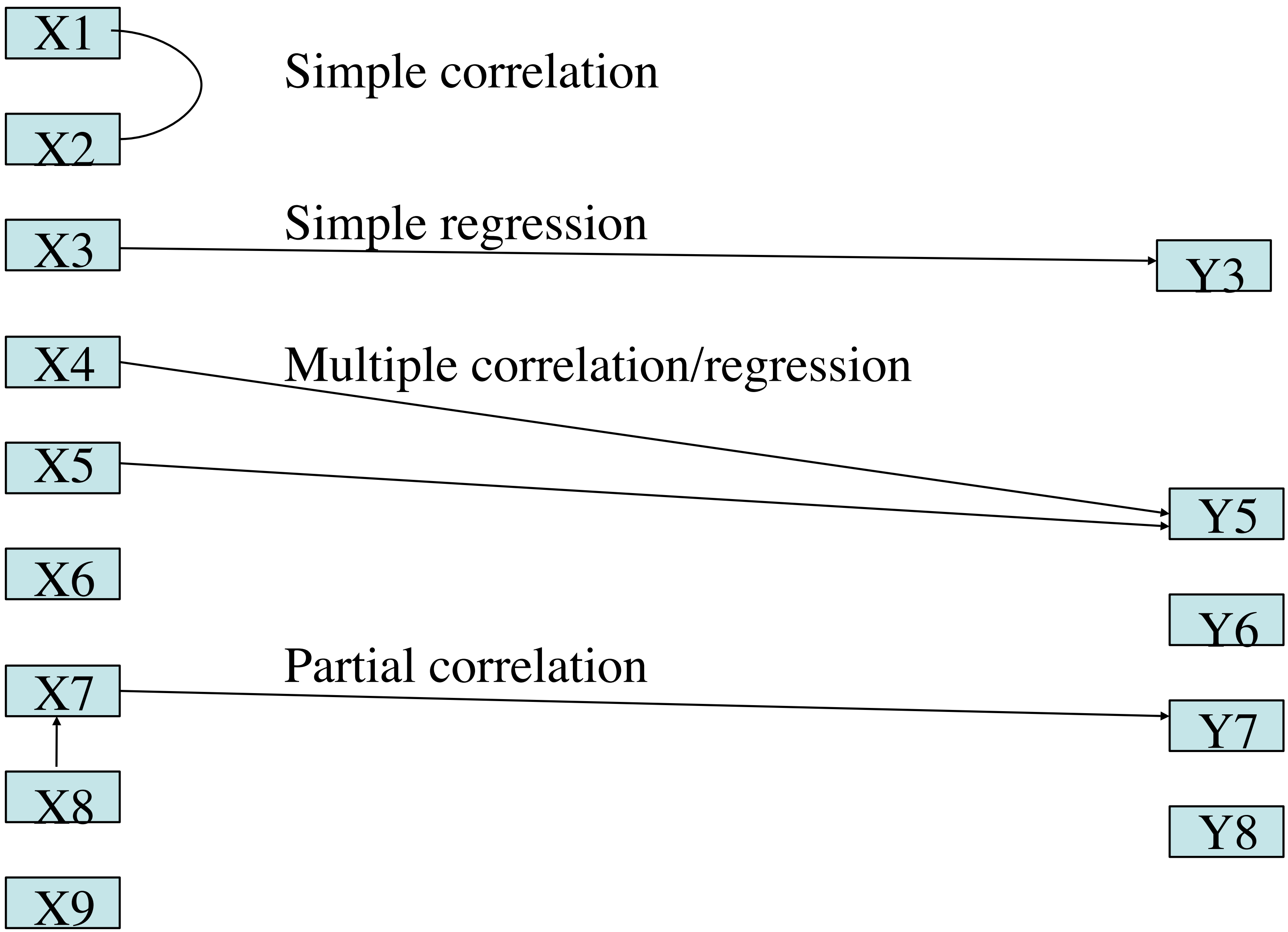
Observed - Latent

Latent Latent

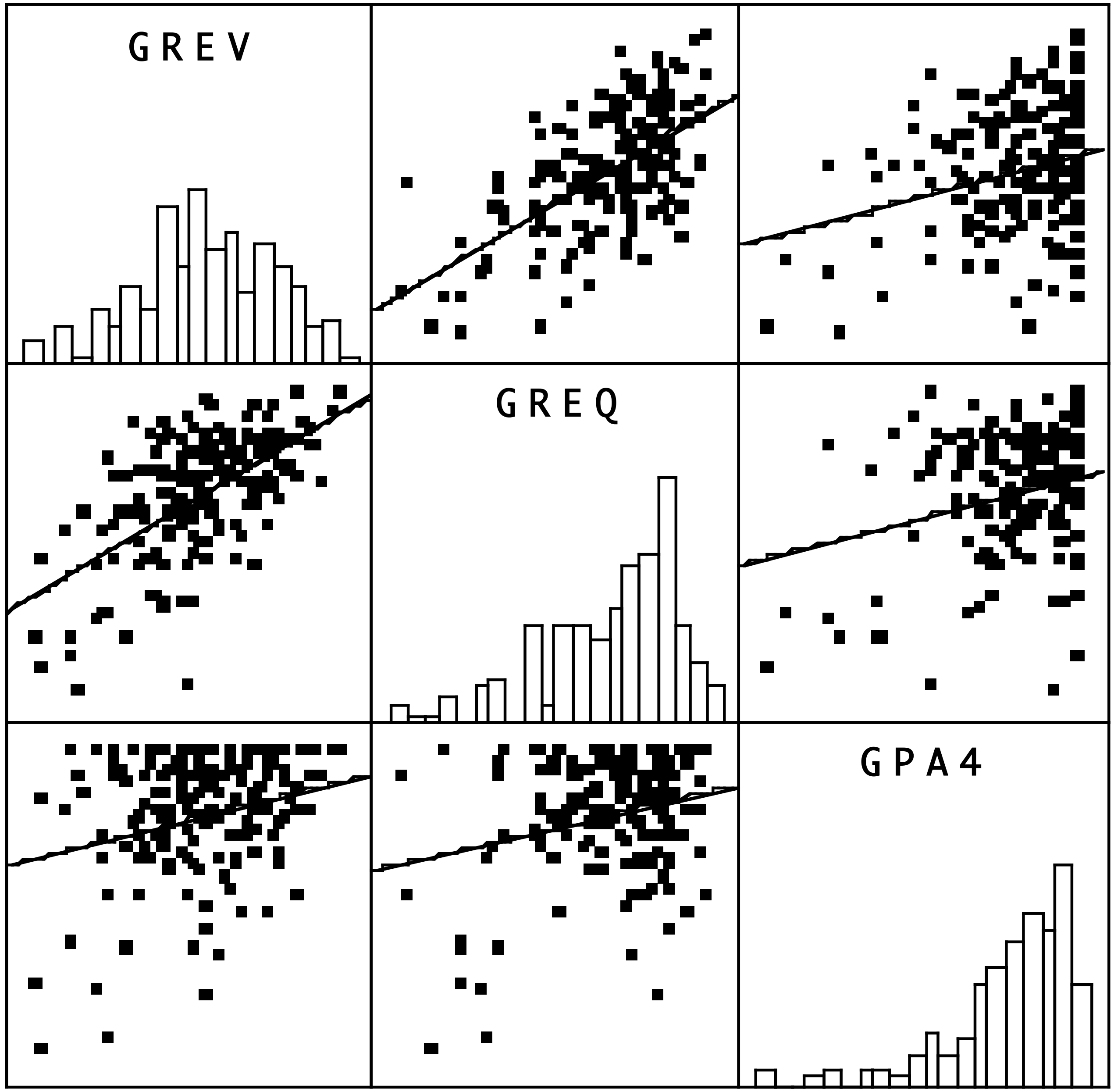
Correlation, regression, and predictive models

- Predicted Y = b(X-X.) + E
- Predicted y = bx + e   data level
- Error = data - model = y - bx
- Optimal linear model (structure level) to minimize squared error
  - $b_{y.x}$ = Cov $_{xy}$/Var $_{X}$
  - r_{xy} = Cov $_{xy}$/sqrt(V_x*V_y)

Variance, Covariance, and Correlation



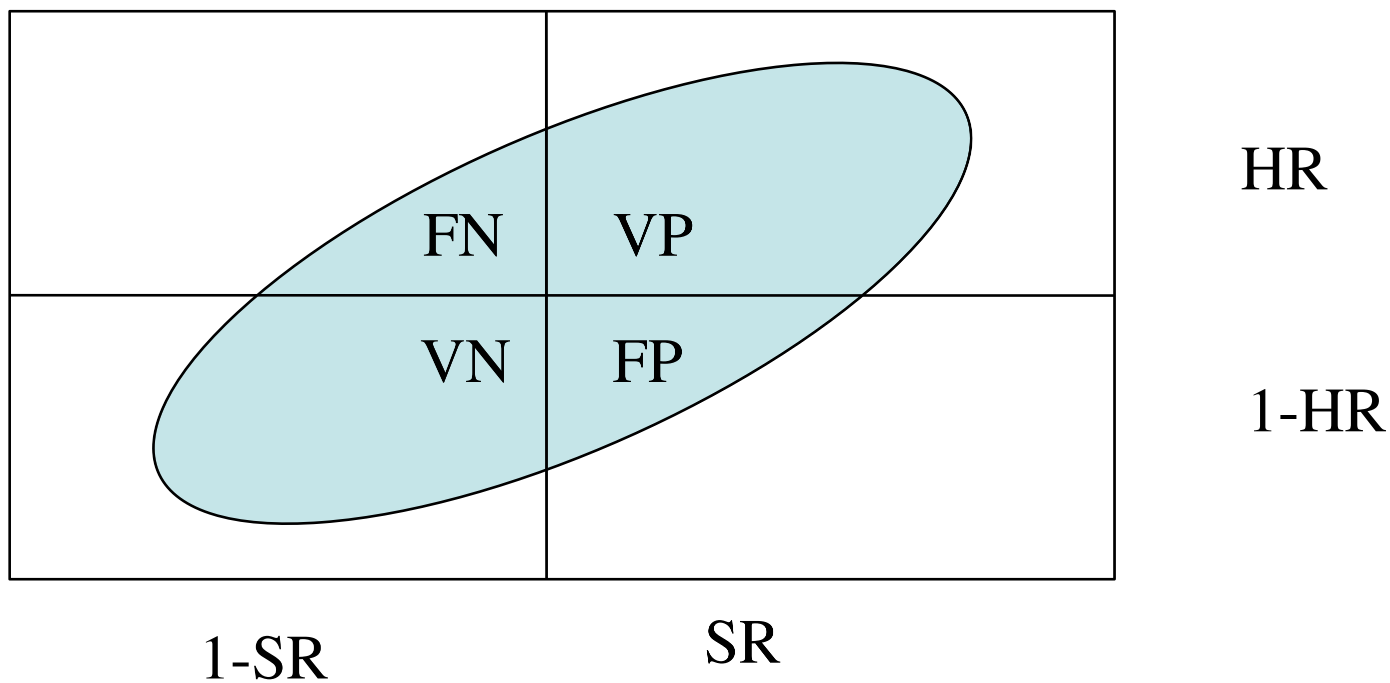
SPLOM of GRE V, Q, GPA



Phi coefficient of correlation

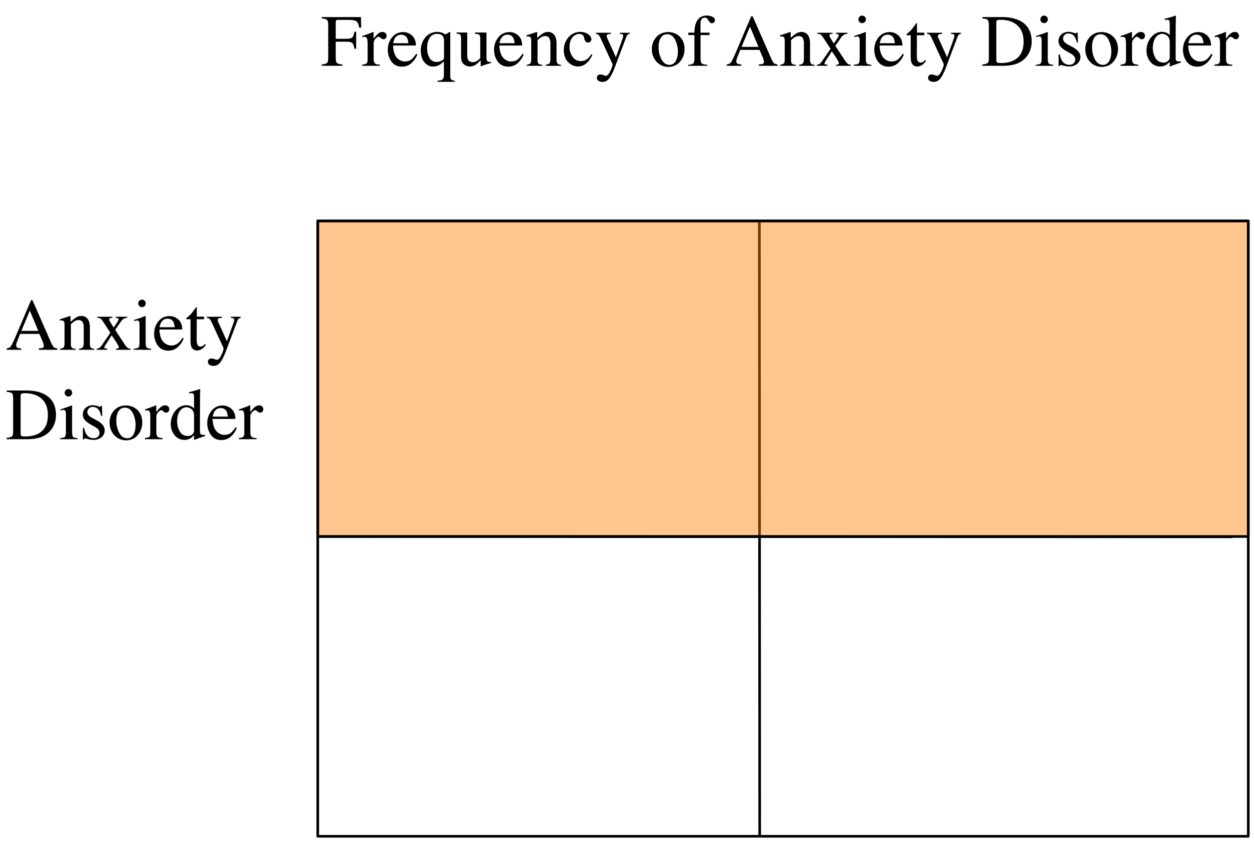
Hit Rate = Valid Positive + False Negative

Selection Ratio = Valid Positive + False Positive



Phi =(VP - HR*SR) /sqrt{HR*(1-HR)*(SR)*(1-SR)}

Correlation and “comorbidity”



Correlation and "comorbidity"

Frequency of Major Depression

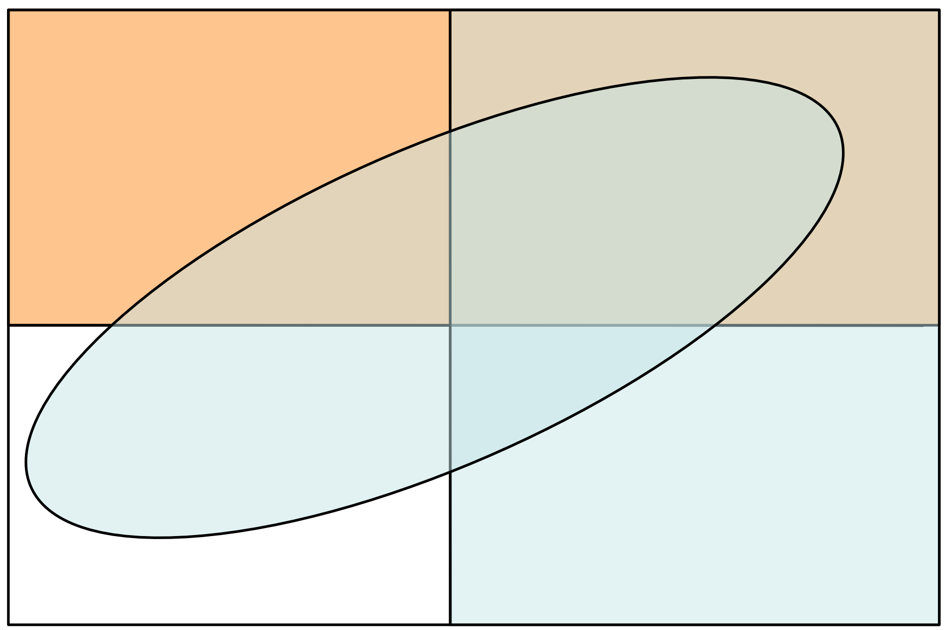
Major Depression

Correlation and “comorbidity”

Comorbidity = joint frequency

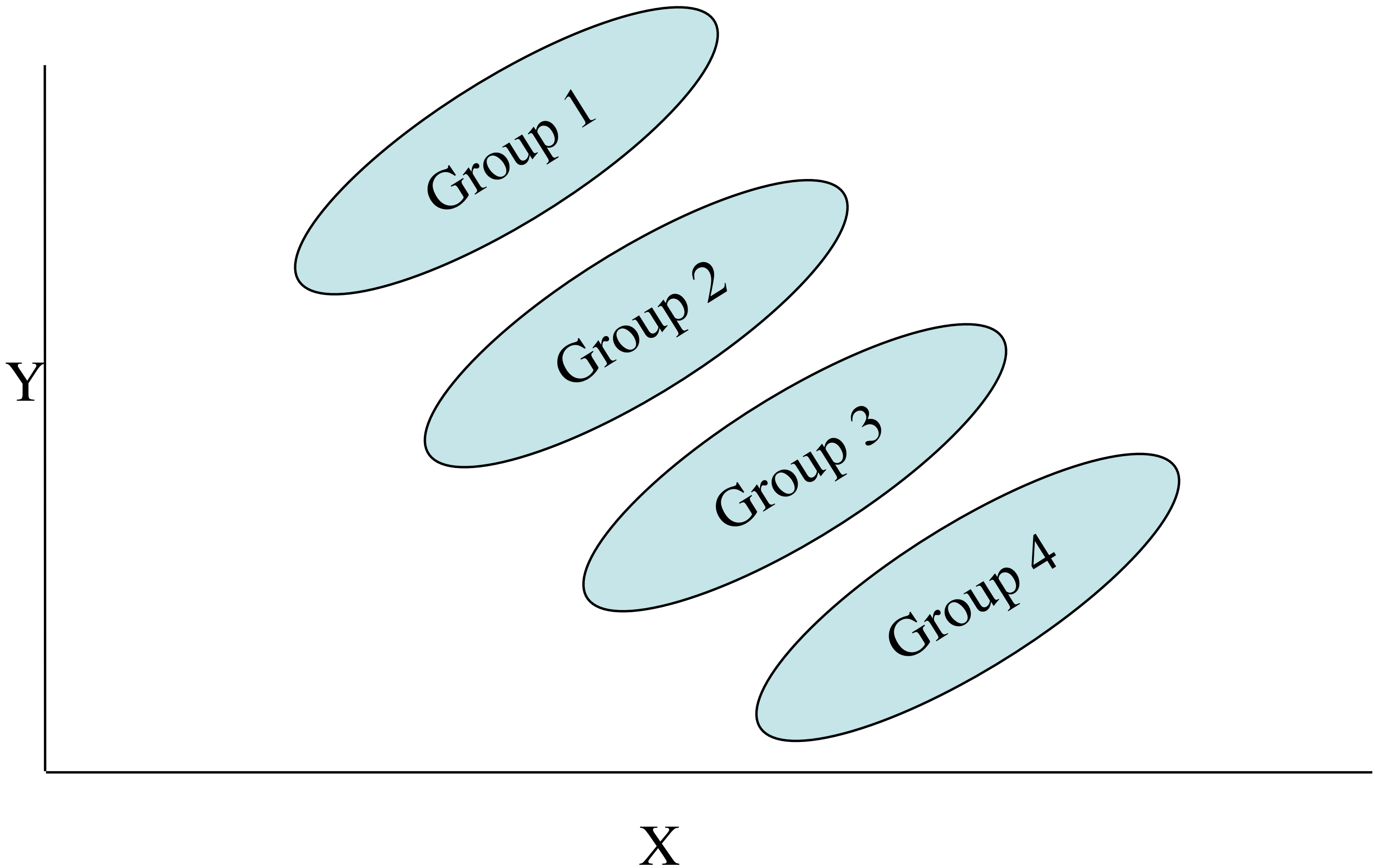
Correlation of Anxiety and Depression

Anxiety
Disorder

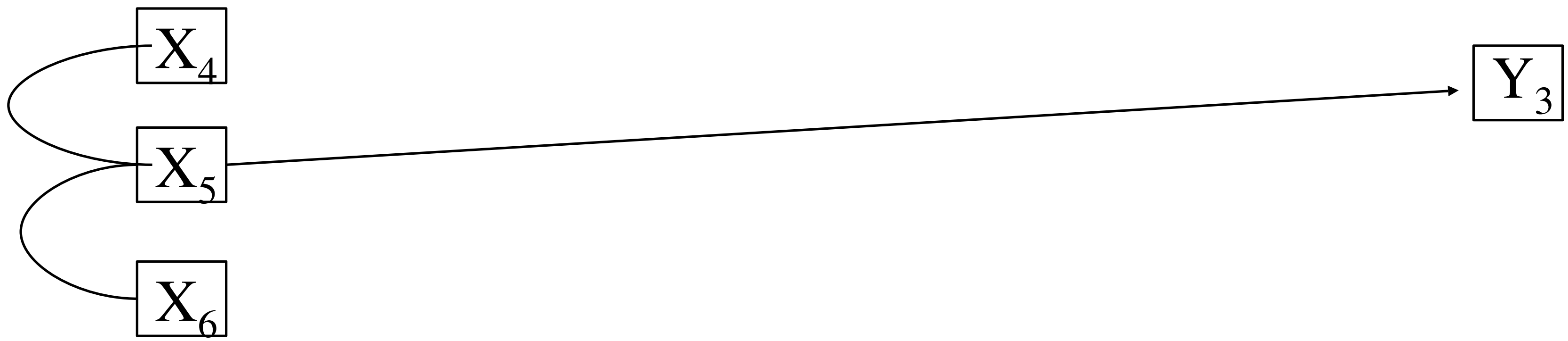


Major Depression

Within Group ≠ Between Group correlation (Simpson's paradox)

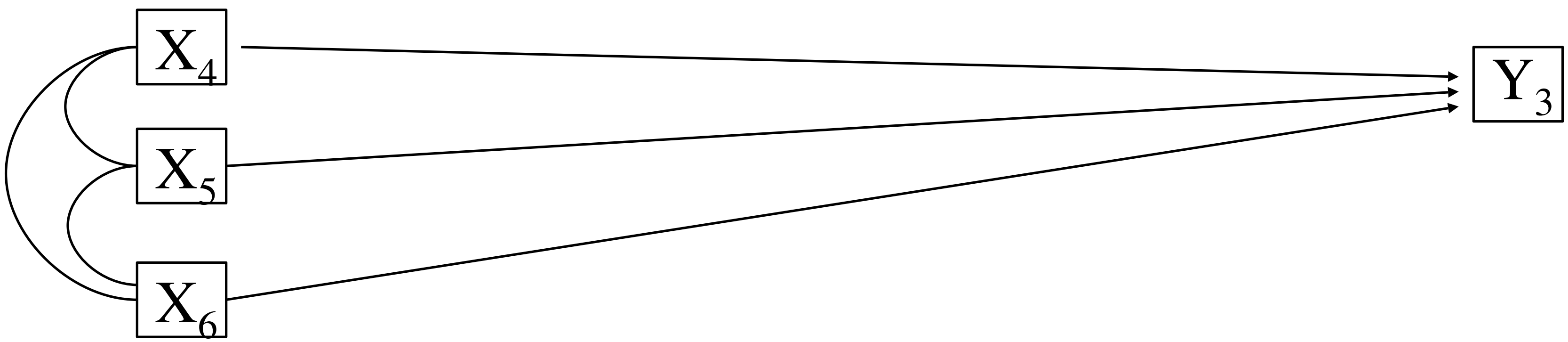


Partial Correlation



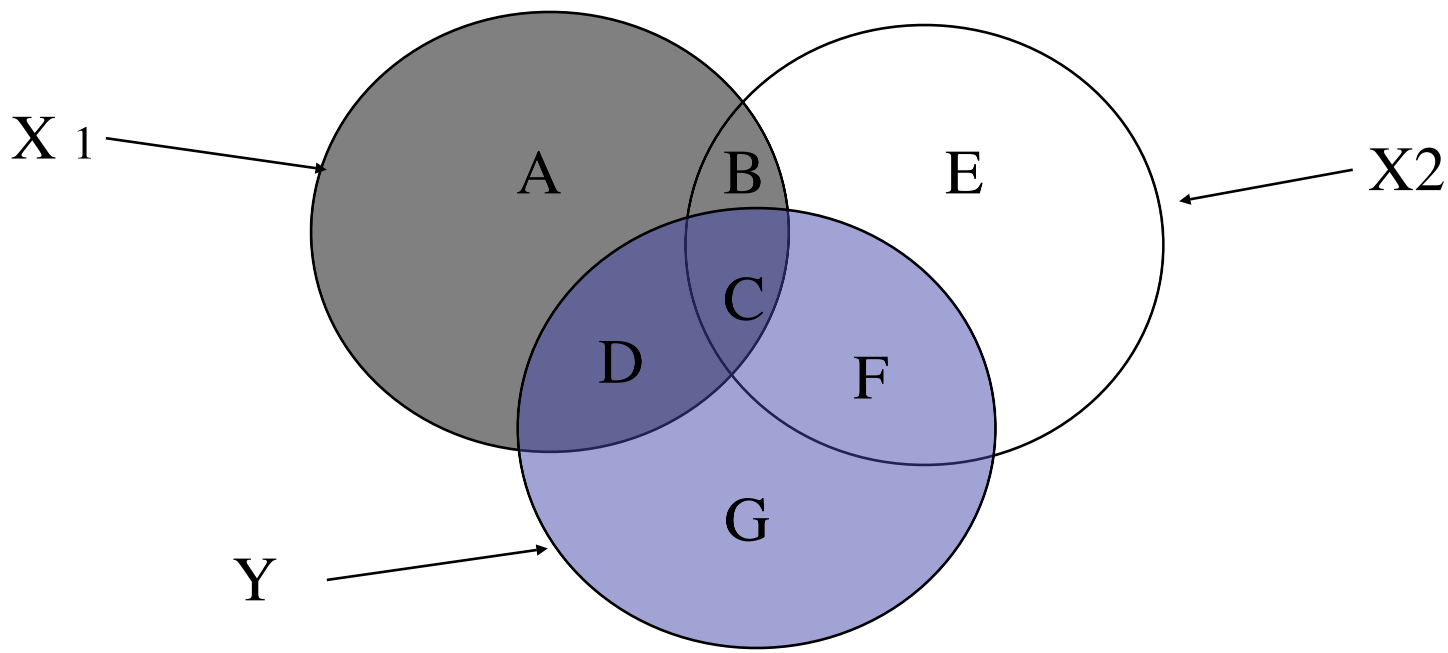
Multiple Correlation

$R_{(x4x5x6)y}$



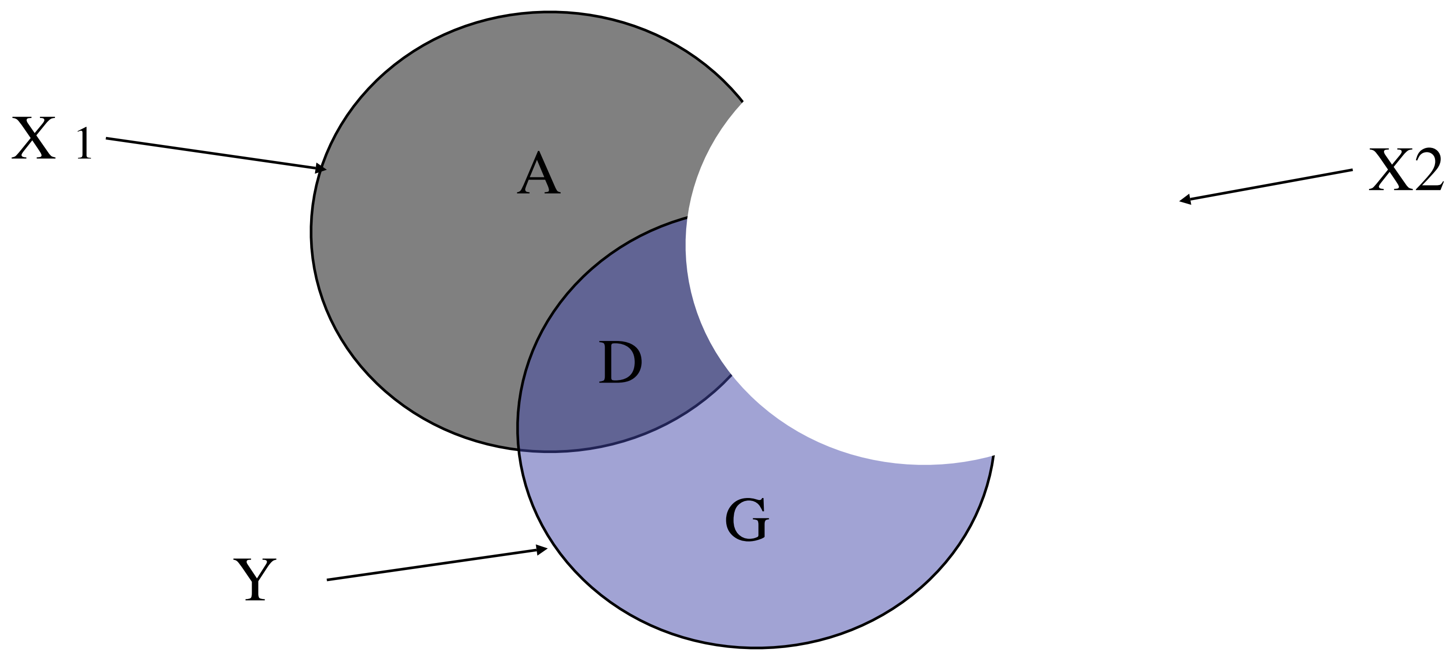
Partial and Multiple Correlation

The conceptual problem



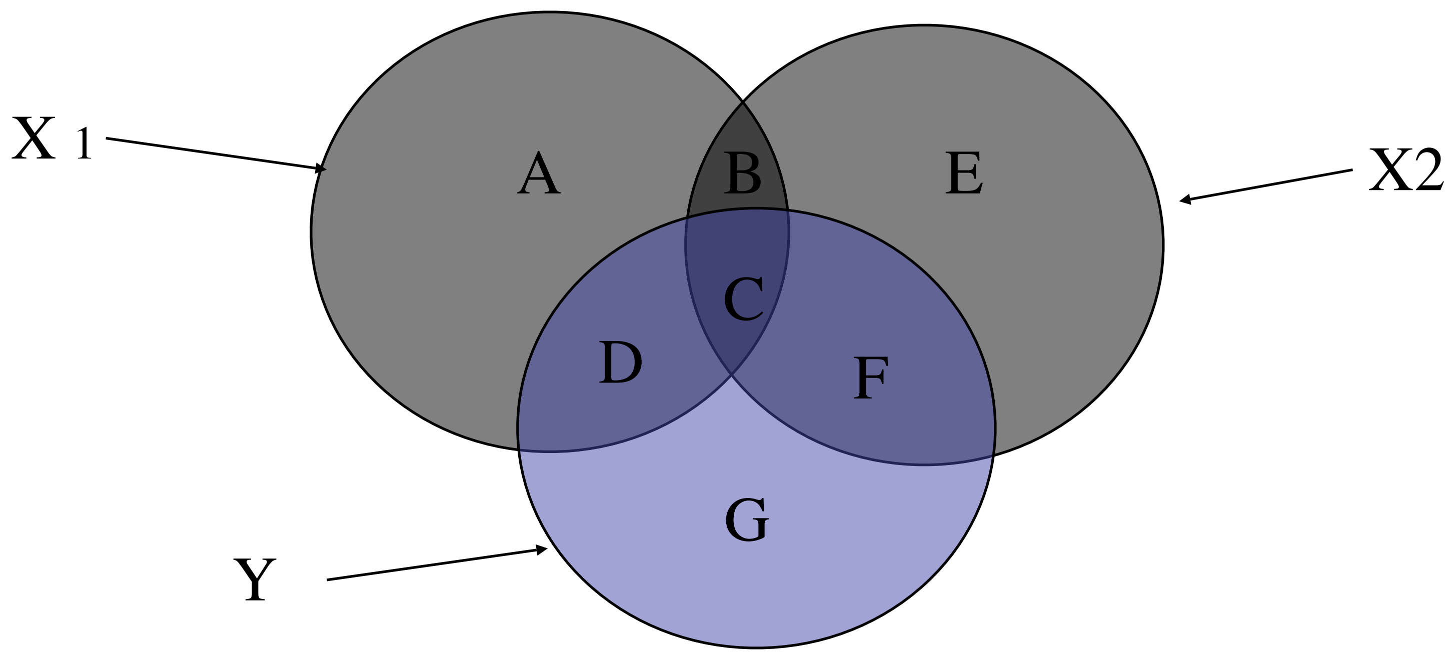
Partial Correlation

The conceptual problem

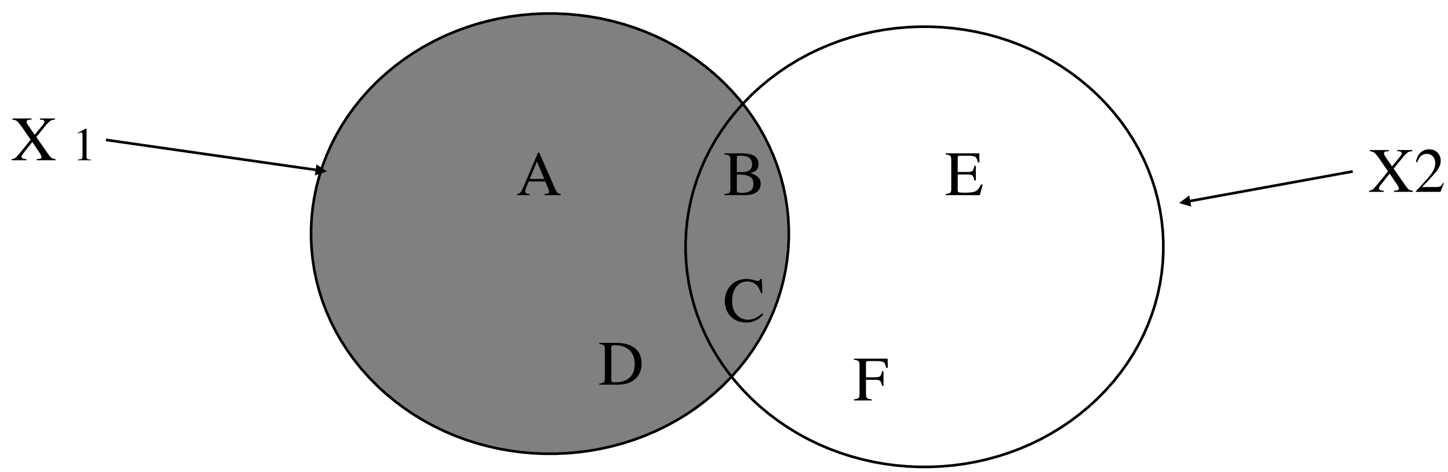


Partial and Multiple Correlation

The conceptual problem



Variance, Covariance and Correlation



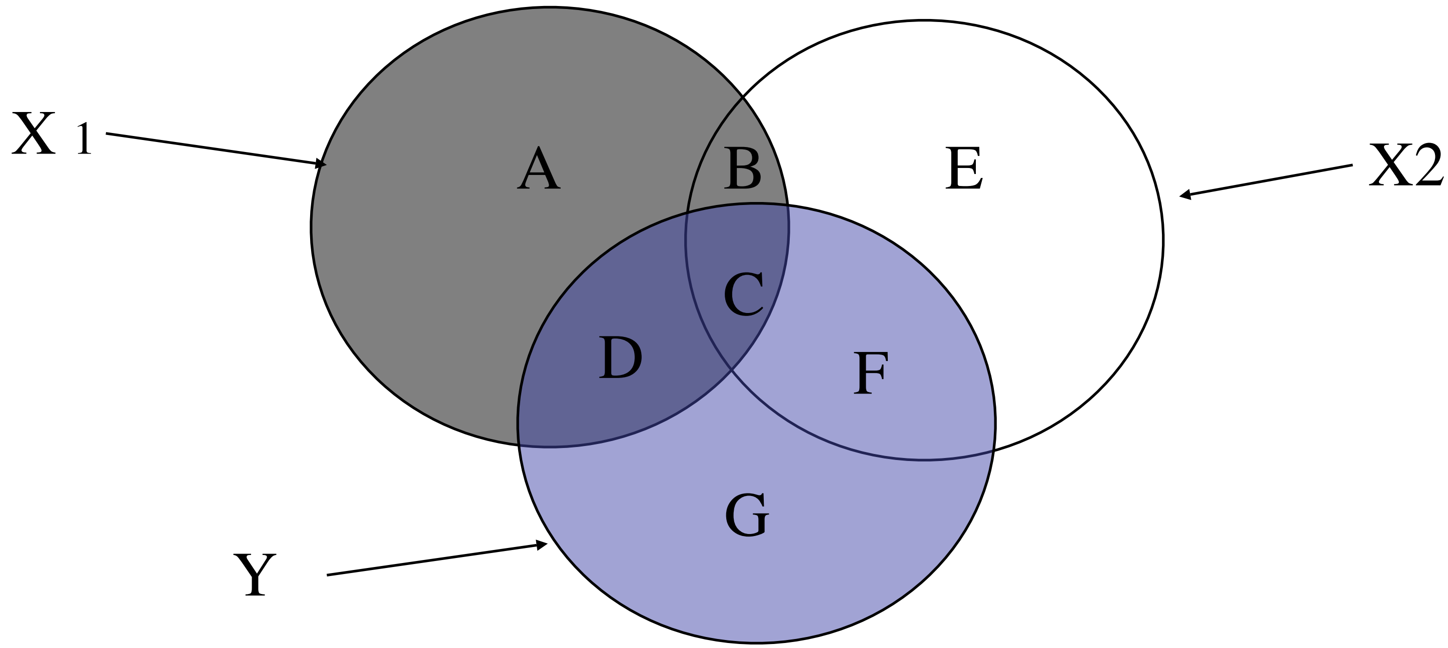
$V1 = A + B + C + D$ $\qquad$ $C12 = B + C$ $\qquad$ $V2 = E + B + C + F$

$r = C12/\mathrm{sqrt}(V1V2)$

$V1.2 = A + D$ $\qquad$ $V2.1 = E + F$

$V1.2 = V1(1-r^2)$ $\qquad$ $V2.1 = V2(1-r^2)$

Partial and Multiple Correlation



V1 = A + B + C + D

V2 = E + B + C + F

VY = D + C + F + G

V1.2 = A + D

C12 = B + C

C1Y = C + D

C2Y = C + F

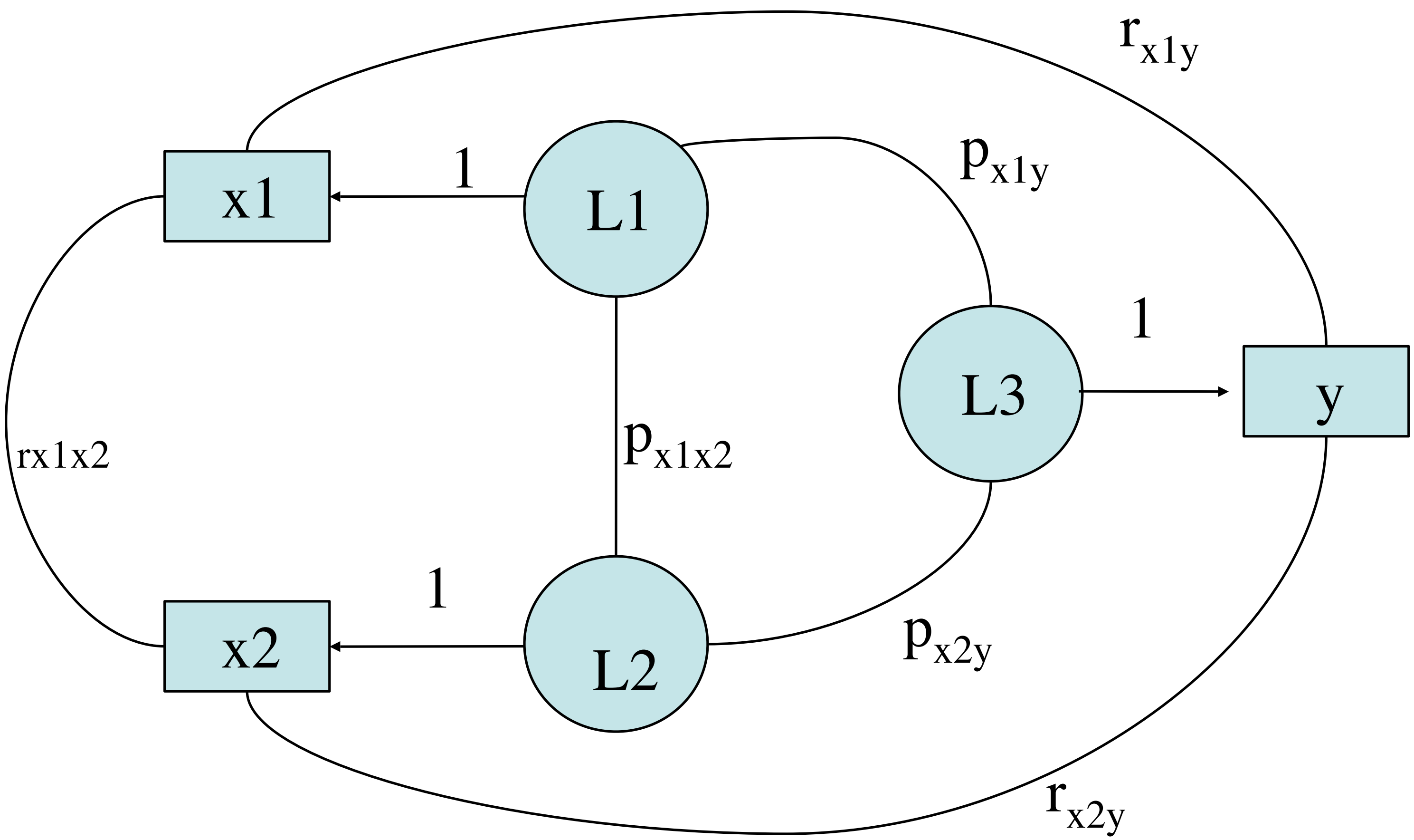
V2.1 = E + F

C1Y.2 = D

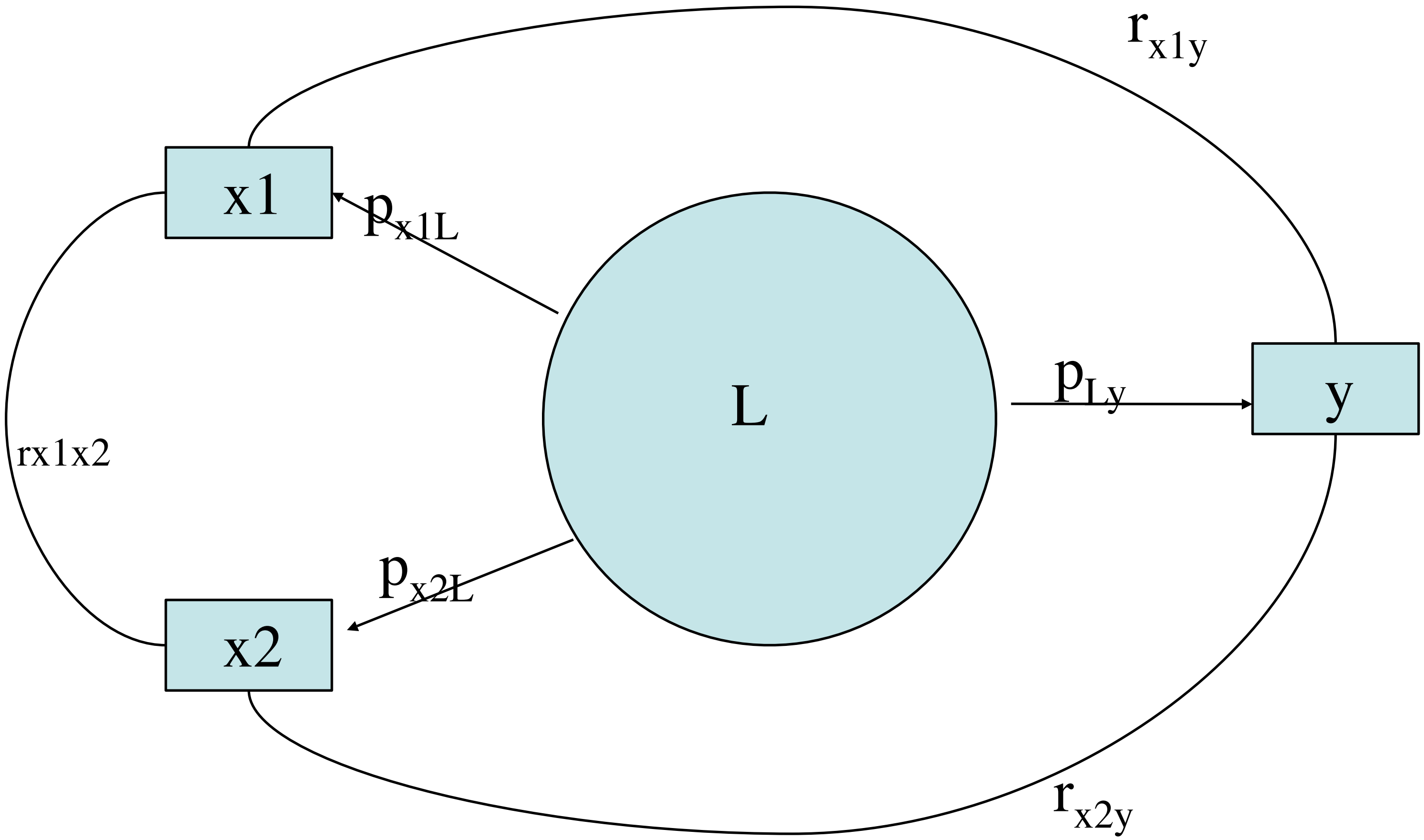
C2Y.1 = F

C(12)Y3 = D + C + F

Partial correlation: conventional model



Partial correlation: Alternative model



Alternative forms of r

$r = cov_{xy}/Sqrt(V_x * V_y) =$

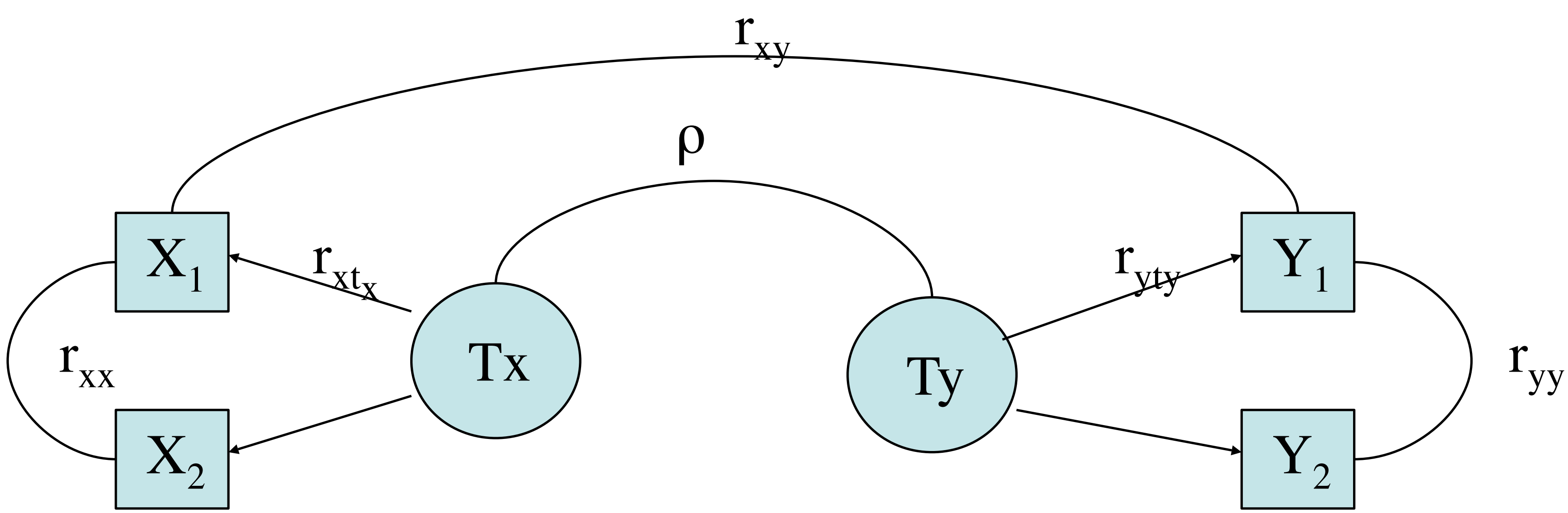
$(\sum xy/N)(sqrt(\sum x^2/N*\sum y^2/N) = (\sum xy)(sqrt(\sum x^2*\sum y^2)$

Correlation	X	Y
Pearson	Continuous	Continuous
Spearman	Ranks	ranks
Point biserial	Dichotomous	Continuous
Phi	Dichotomous	Dichotomous
Biserial	Dichotomous (assumed normal)	Continuous
Tetrachoric	Dichotomous (assumed normal)	Dichotomous (assumed normal
Polychoric	categorical (assumed normal)	categorical (assumed normal)

Other forms of multiple correlation

- Continuous criterion -- multiple regression
- Dichotomous criterion - logistic regression
  - $P(y|x1,x2, \ldots) = 1/(1+\exp(b_1x_1+b_2x_2 \ldots))$
  - Odds ratio of $P(y)/P(\text{not } y) = \exp(b_1x_1+b_2x_2$
  - $\text{Ln(Odds)} = b_1x_1+b_2x_2$
- Choose the appropriate model to best represent the data

Reliability- Correction for attenuation

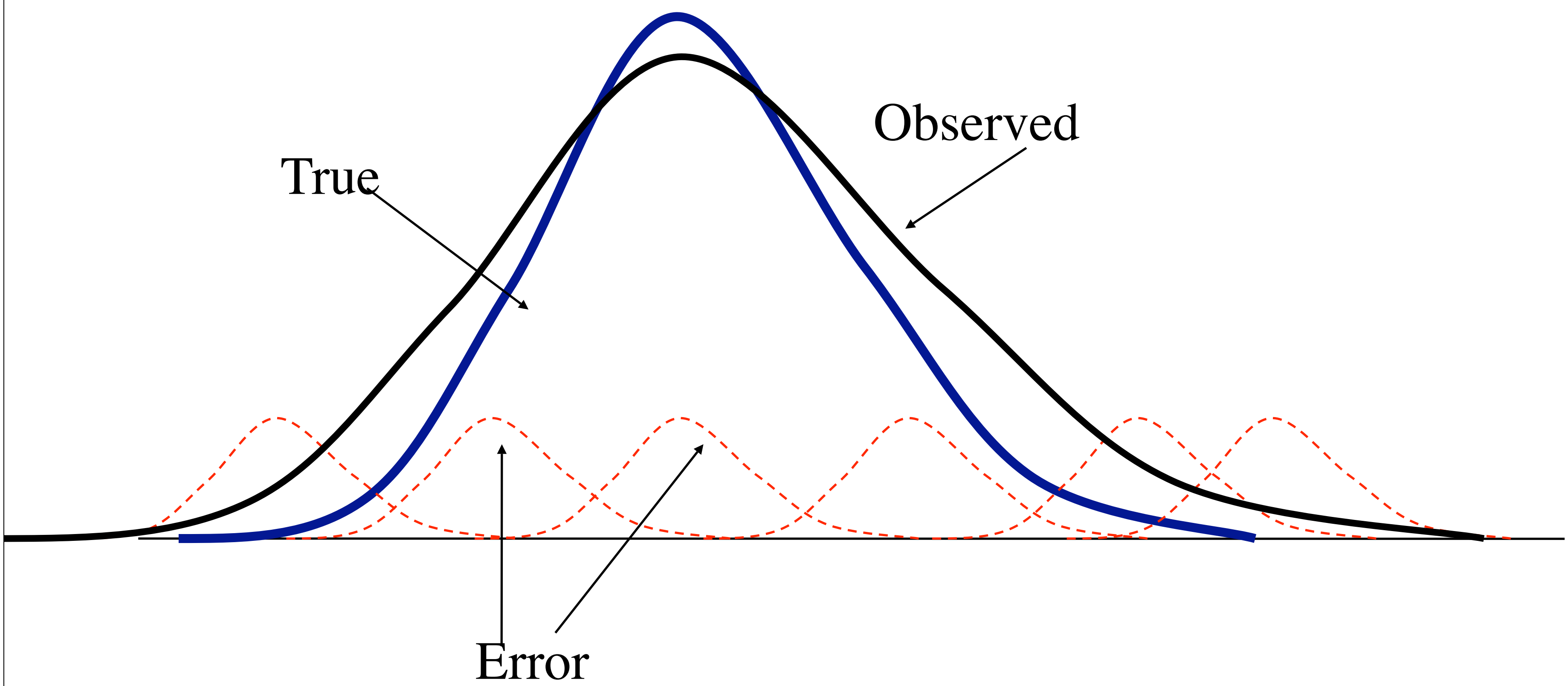


r_{xt_x}= sqrt(r_{xx})

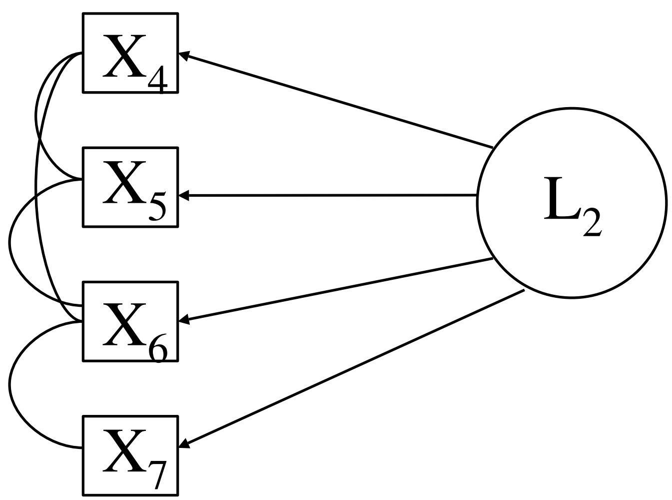
r_{yty}= sqrt(r_{yy})

$\rho = r_{xy}/\mathrm{sqrt}(r_{xx}*r_{yy})$

Observed= True + Error



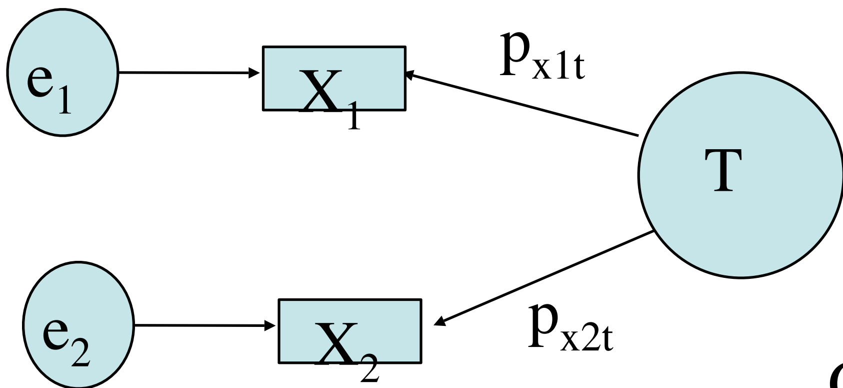
Reliability and Classical Test Theory



Parallel tests

Congeneric measurement

Parallel Tests



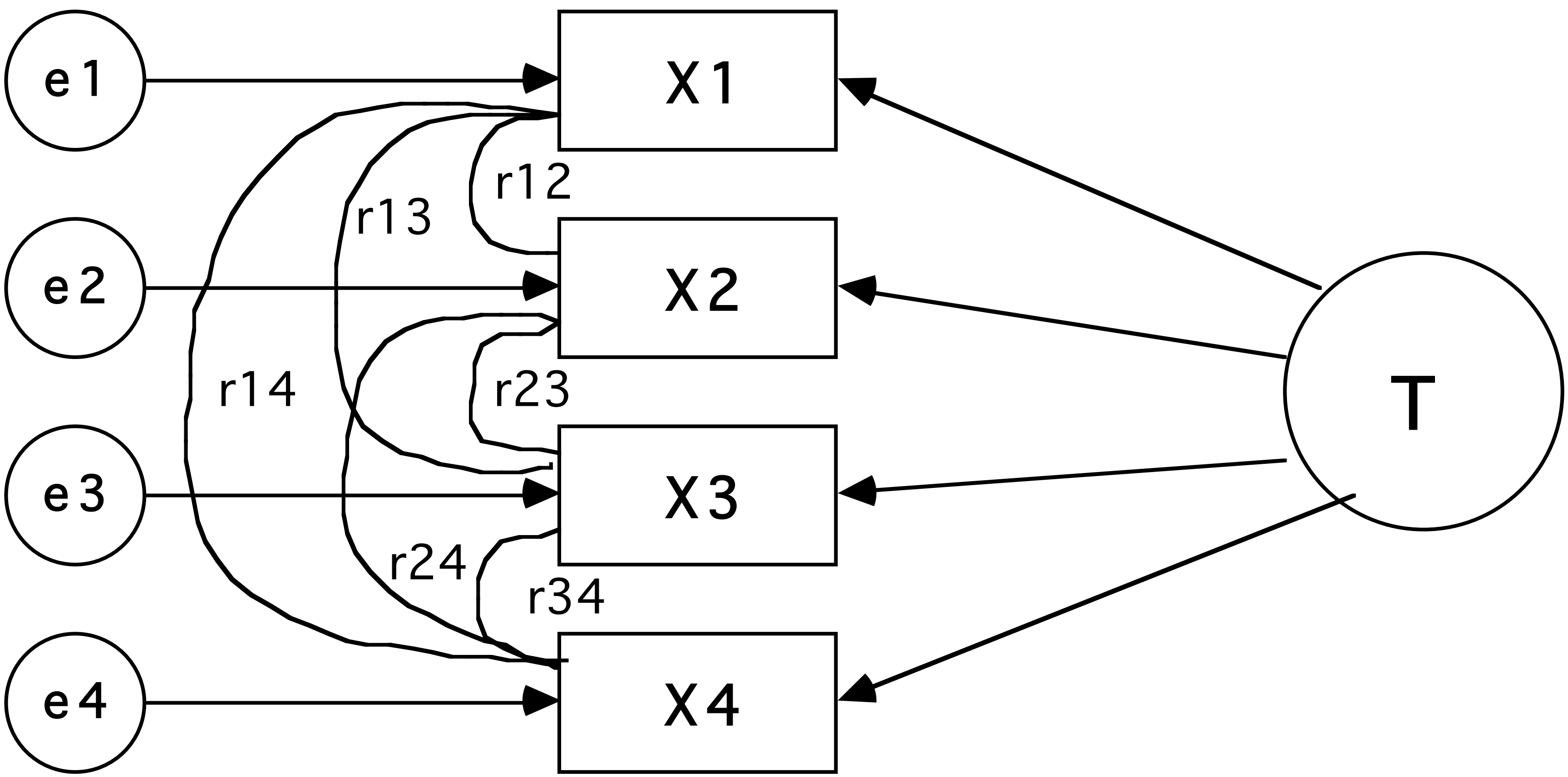
$$V_{x1}=V_t+V_{e1}$$
$$V_{x2}=V_t+V_{e2}$$

$$C_{x1x2}=V_t+C_{te1}+C_{te2}+C_{e1e2}= V_t$$

$$r_{xx}=C_{x1x2}/Sqrt(V_{x1}*V_{x2}) = V_t/V_x$$

The reliability of a test is the ratio of the true score variance to the observed variance = the correlation of a test with a test “just like it”

Congeneric Measurement



Reliability and components of variance

- Components of variance associated with a test score include
  - General test variance
  - Group variance
  - Specific item variance
  - Error variance (note that this is typically confounded with specific)
- Reliability coefficients are estimates of various sources of variance

Coefficients Alpha, Beta, Omega

Test	General	Group	Specific	Error
Reliable	General	Group	Specific	
Common Shared	General	Group		
Alpha	General	< group		
Beta	≈general			
Omega	general			

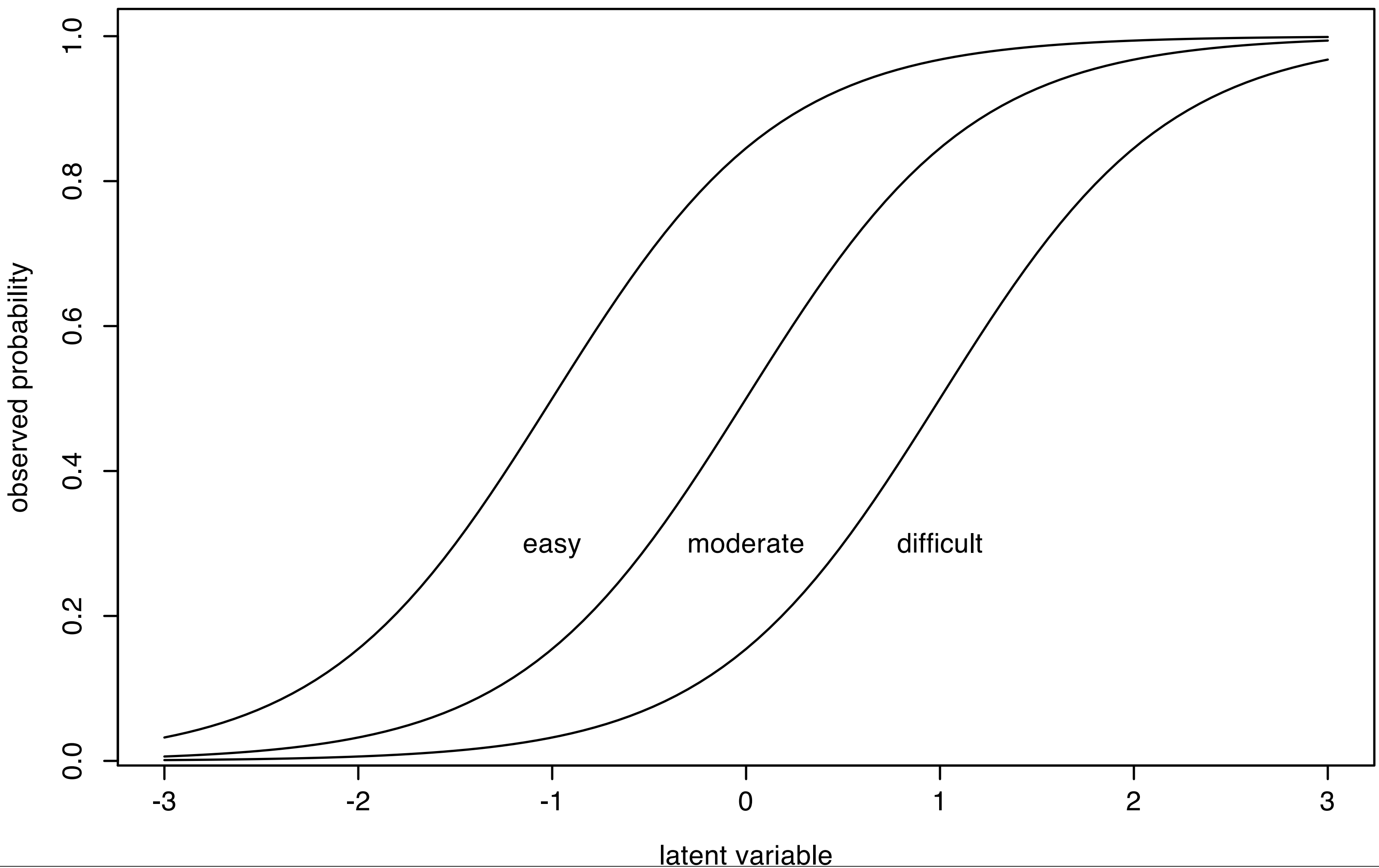
Facets of reliability

Across items	Domain sampling Internal consistency
Across time	Temporal stability
Across forms	Alternate form reliability
Across raters	Inter-rater agreement
Across situations	Situational stability
Across "tests" (facets unspecified)	Parallel test reliability

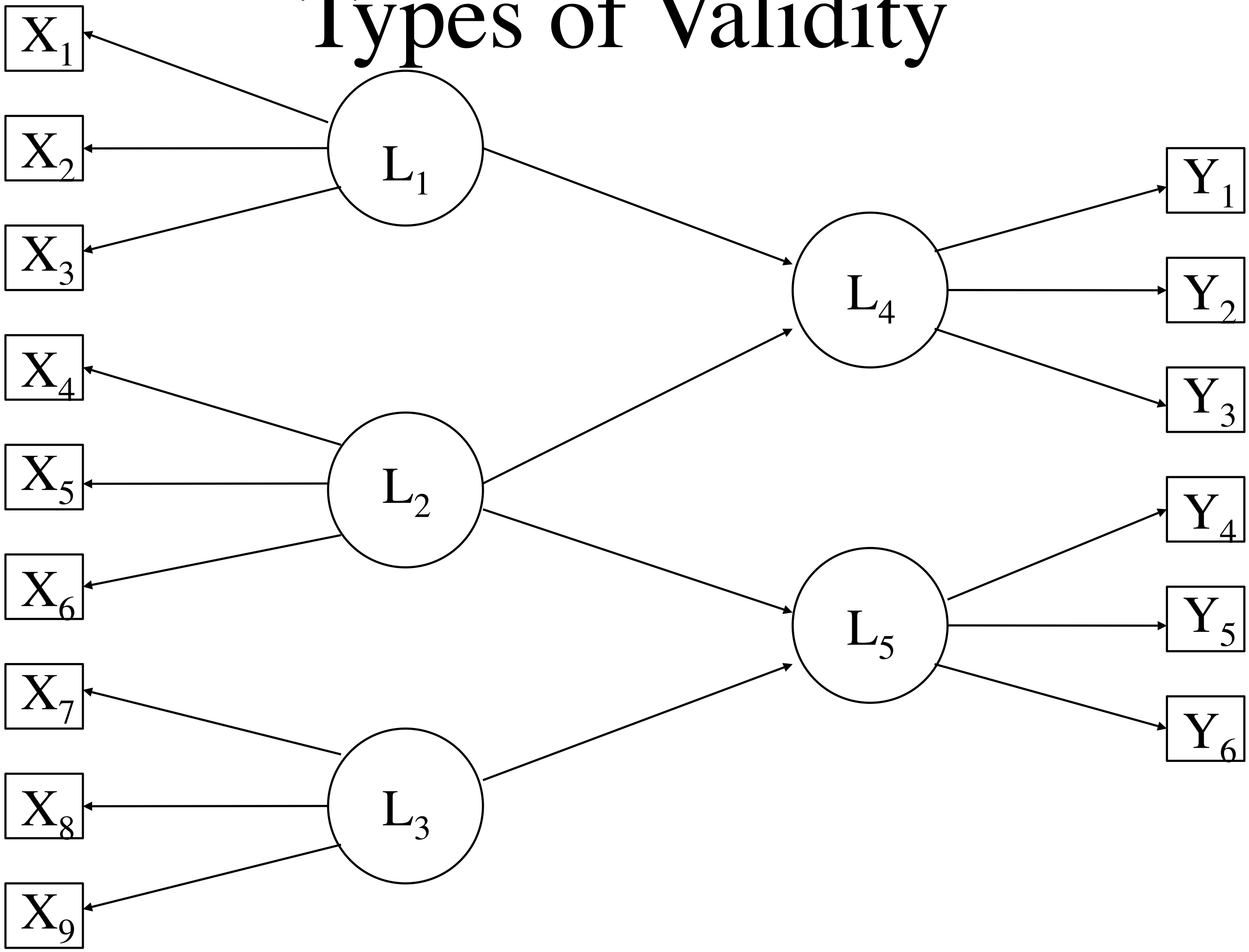
Classic versus "new" test models

- Classic
  - X = T + E
  - Observed score = True score + Error score
  - Estimated t = r_{xx} * x
  - Goodness of fit is r_{xx}
- New
  - $Pr(X \mid \phi,\delta) = 1/(1+\exp(1.7*(\delta-\phi))$
  - Find ϕ_i, δ_j to best fit observed scores
  - Goodness of fit = $Pr(X_{ij}) - 1/(1+\exp(1.7*(\delta-\phi))$

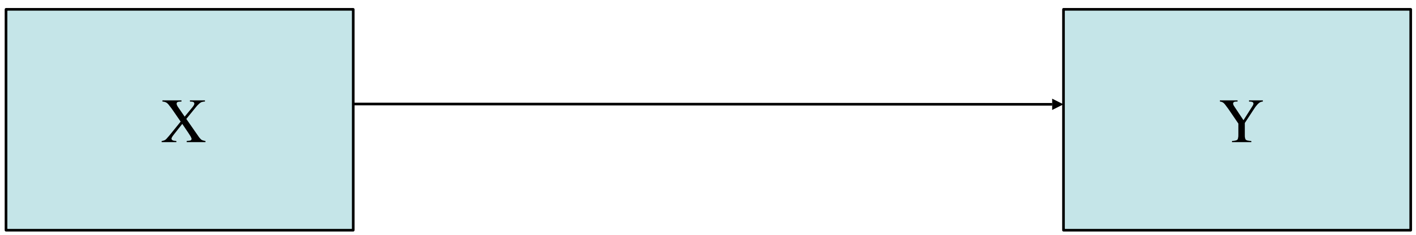
IRT of three item difficulties



Types of Validity



Predictive Validity

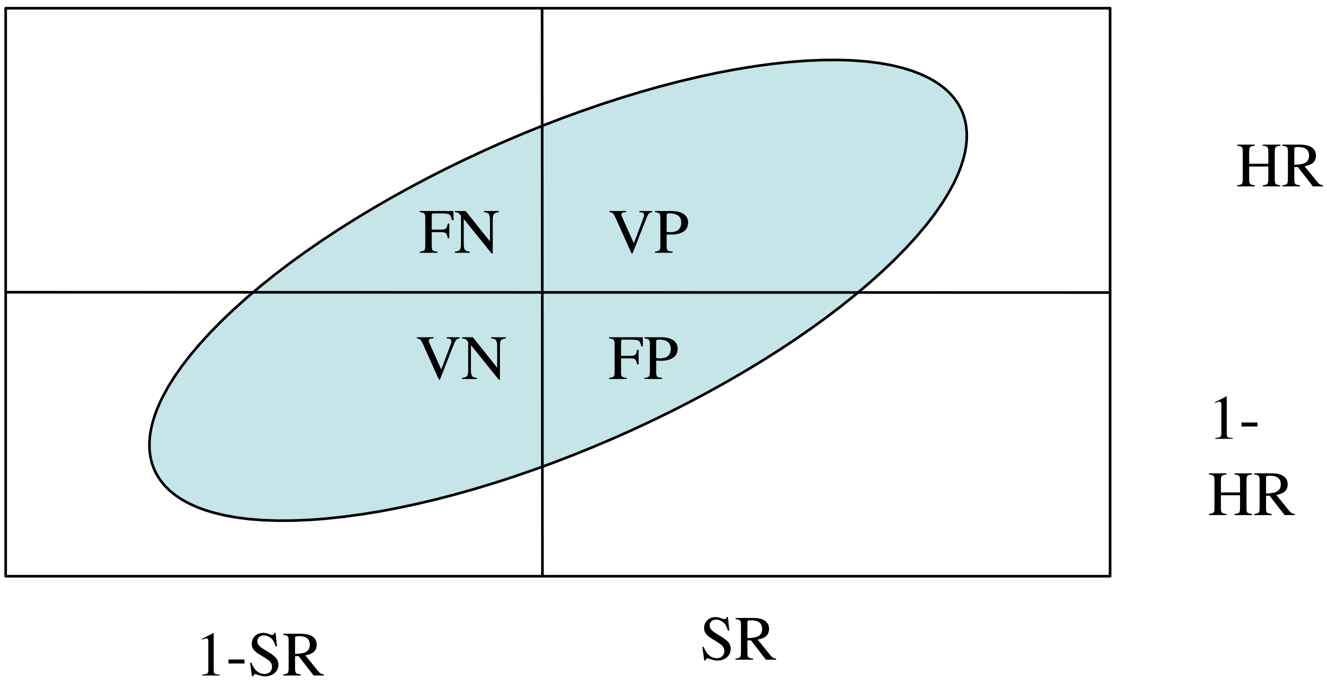


- Does a measure correlate with the criterion?
- Need to define the criterion.
- Requires waiting for time to pass.

Predictive and Concurrent Validity and Decision Making

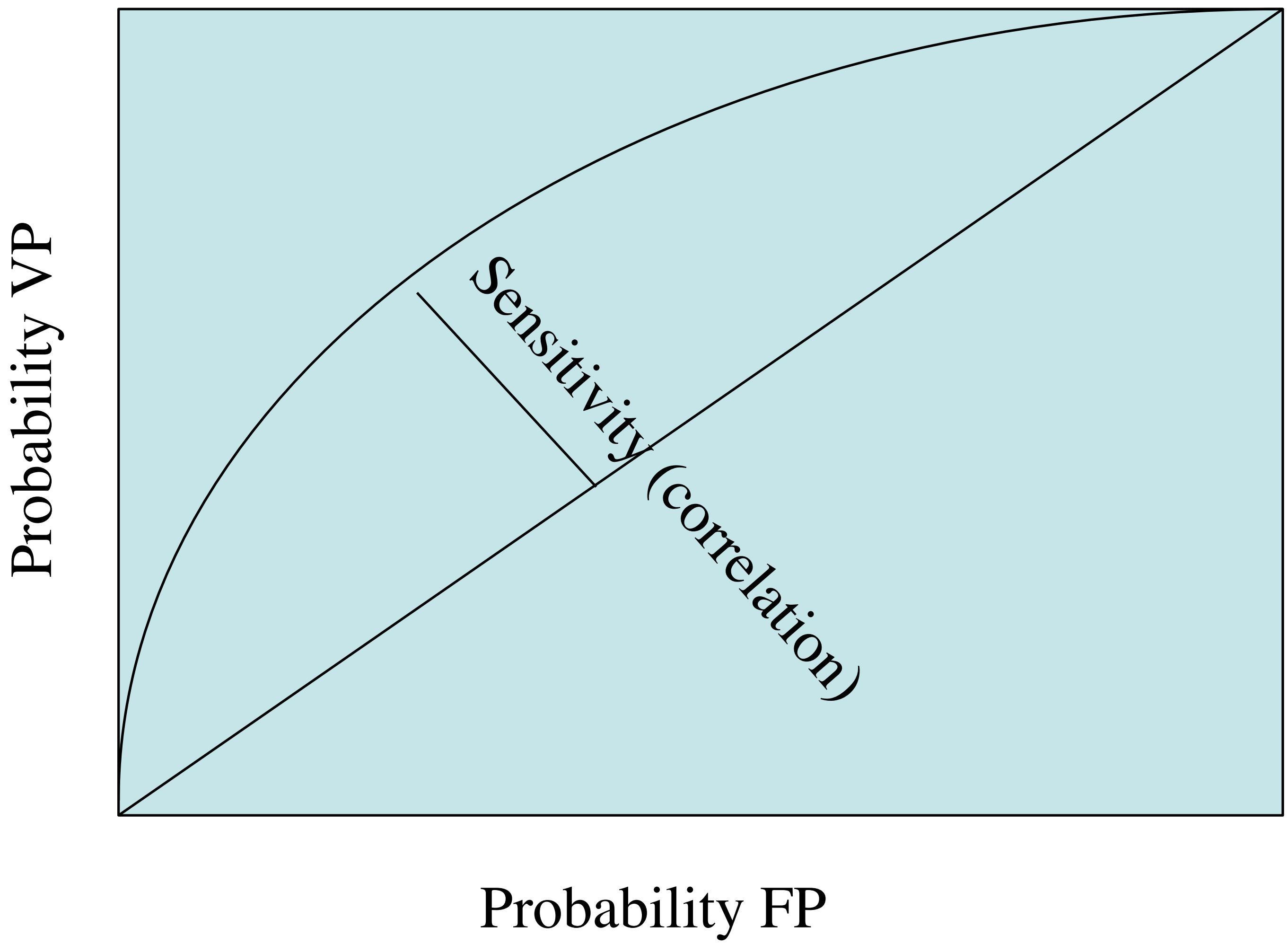
Hit Rate = Valid Positive + False Negative

Selection Ratio = Valid Positive + False Positive

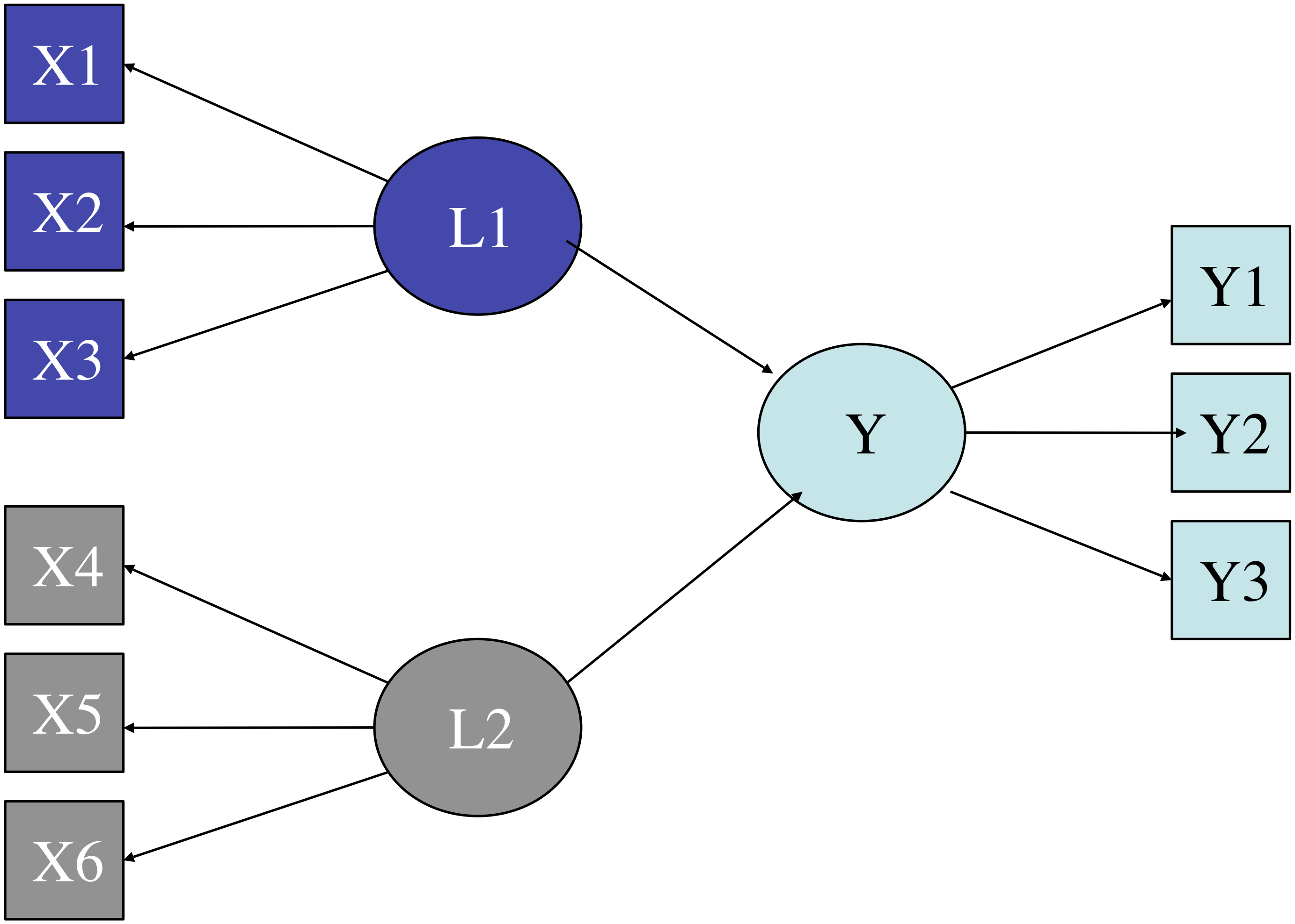


Phi =(VP - HR*SR) /sqrt(HR*(1-HR)*(SR)*(1-SR)

Decision Theory and Signal Detection



Construct Validity: Convergent, Discriminant, Incremental



Multi-Trait, Multi-Method Matrix

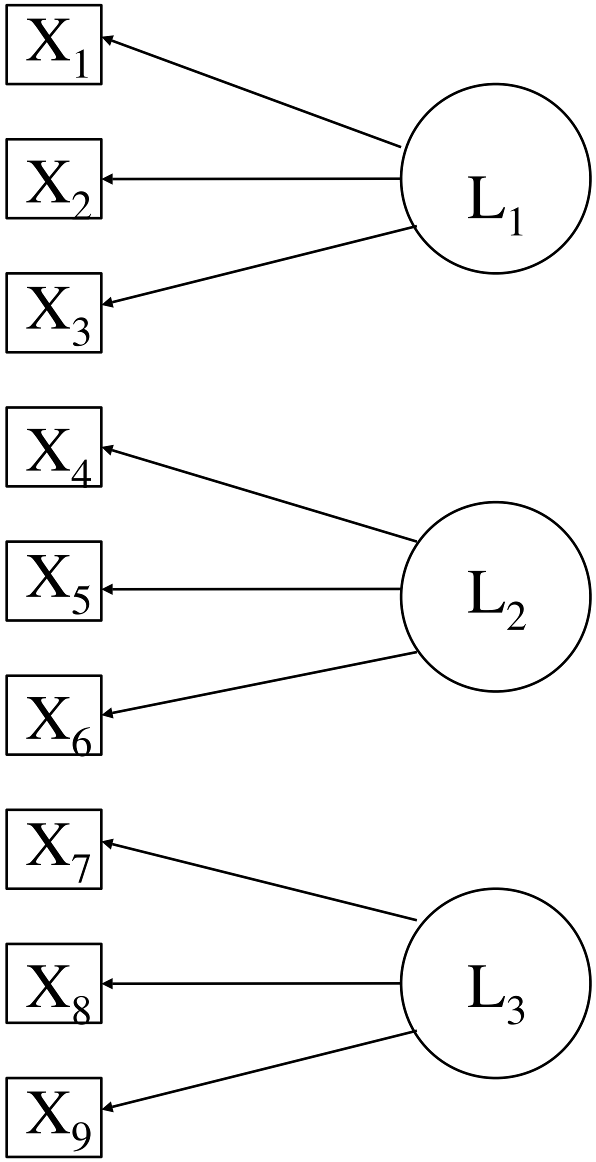
	T1M1	T2M1	T3M1	T1M2	T2M2	T3M2	T1M3	T2M3	T3M3
T1M1	T1M1								
T2M1	M1	T2M1							
T3M1	M1	M1	T3M1						
T1M2	T1			T1M2					
T2M2		T2		M2	T2M2				
T3M2			T3	M2	M2	T3M2			
T1M3	T1			T1			T1M3		
T2M3		T2			T2		M3	T2M3	
T3M3			T3			T3	M3	M3	T3M3

Mono-Method, Mono trait = reliability

Hetero Method, Mono Trait = convergent validity

Hetero Method, Hetero Trait = discriminant validity

Factor and principal component analysis



Data simplification and description

Number of factors/PC

Interpretation of factors/PC

Factor Analysis: the model

- $R \approx FF' + U^2$
  - Residual Matrix $R^* = R - (FF' + U2)$
  - Try to minimize the (squared) residual
- Variables are linear composites of unknown (latent) factors. $X_{ij} = \sum f_{ik} s_{kj} = \sum$ loading*scores
- Covariance structures of observables in terms of covariance of unobservables $r_{xy} = \sum f_{xi} f_{yi}$

Principal components: the model

- $R \approx CC'$
  - Residual Matrix $R^* = R - (CC')$
  - Try to minimize the (squared) residual
- Components are linear composites of known variables.
- Covariance structures of observables in terms of covariance of observables
- Components account for observed variance

Rotations and transformations

- Orthogonal rotations
  - Factors are orthogonal, rotated to reduce (or maximize) particular definition of simple structure
- Oblique transformations and higher order factors
  - Allows factors to be correlated (and thus have higher order factors)

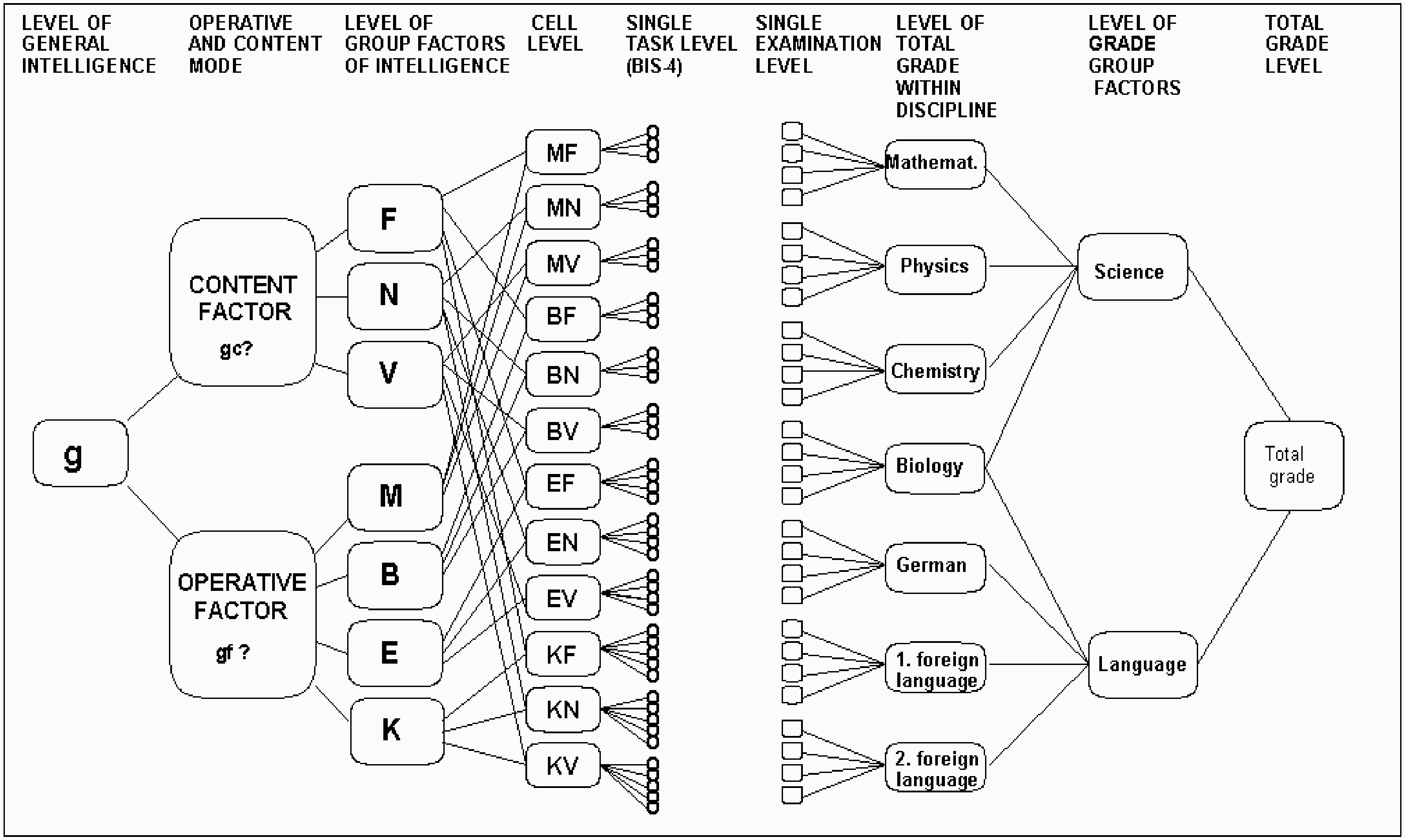
${}_nF_{k}\,{}_kF'_n \approx {}_nR_n + U^2$

${}_nF_{k}\,{}_kT_{k}\,{}_kF'_n \approx {}_nR_n + U^2$ where ${}_kT_k$ is a transformation matrix. If T is orthogonal, then this is called rotation

Factors, Components and models

- data level
  - $d_{ij} = \sum f_{ik}\lambda_{jk} + e_{ij}$
- structure level
  - $r_{xy} = \sum f_{ix} * f_{iy}$
  - $R \approx FF' + U^2$ or $R = CC'$

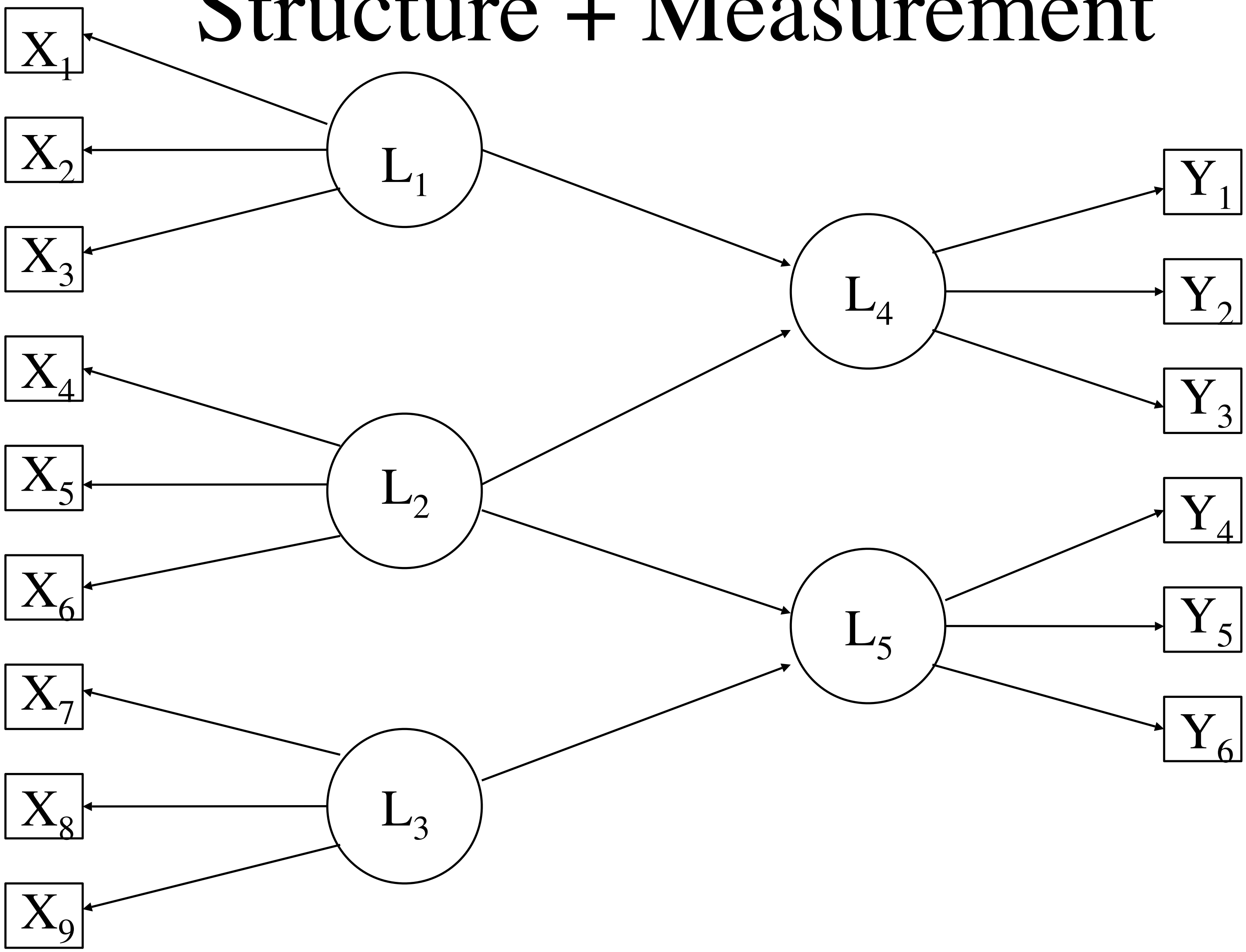
Fig. 9: **Hierarchical version of the Berlin model of intelligence and a grade hierarchy model**



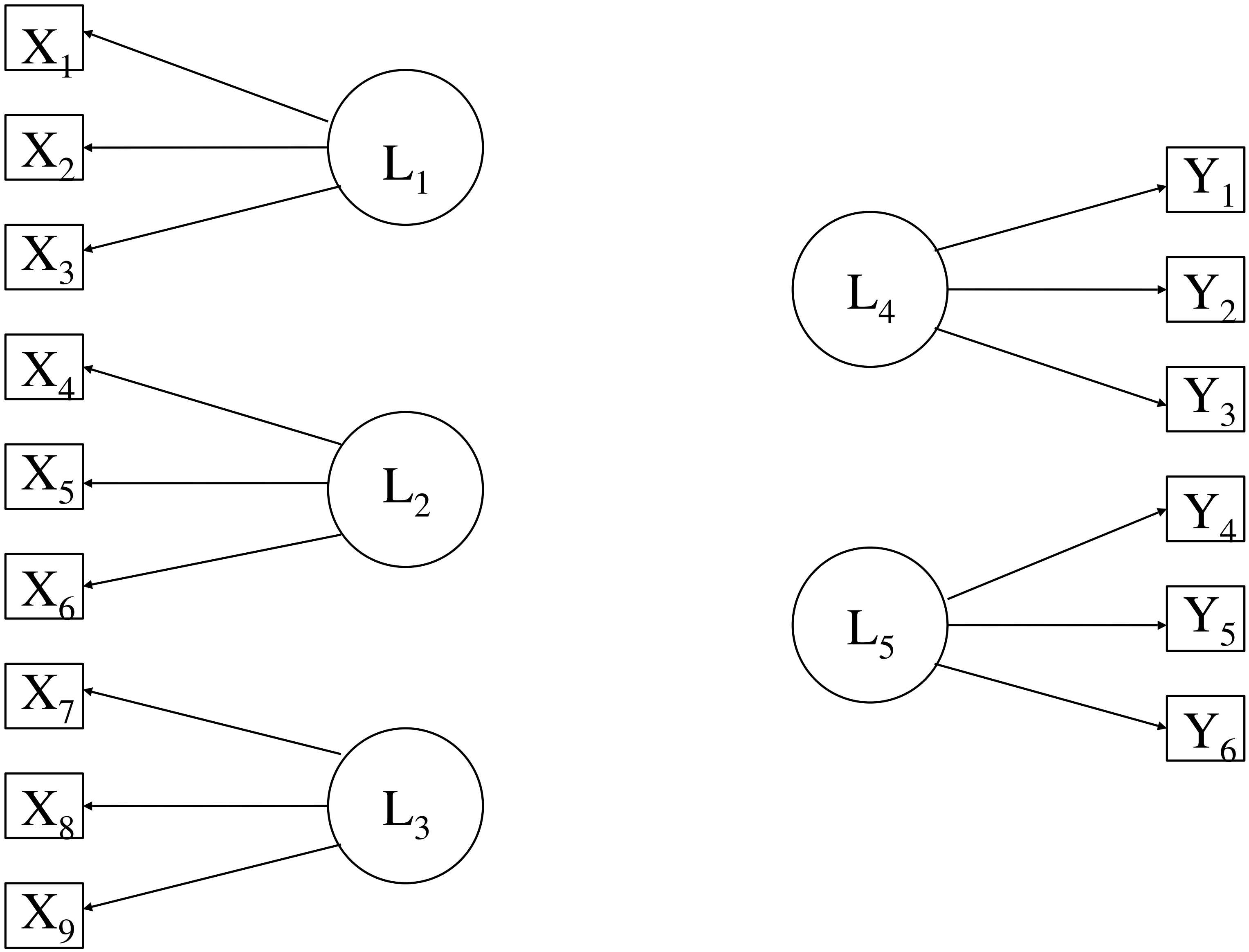
K: Processing capacity for complex information, i.e. reasoning
E: Creativity
B: Speed on relatively simple tasks
M: Memory, i.e. storage capacity for information

F: figural Intelligence
N: numerical Intelligence
V: verbal intelligence

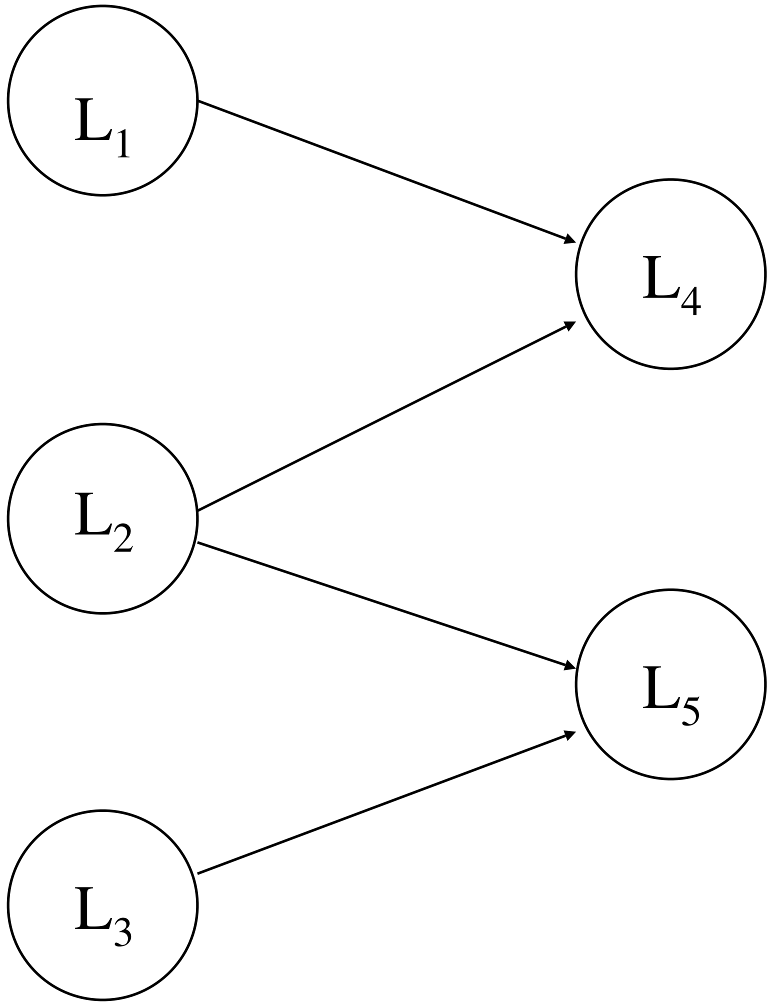
Structural Equation Modeling = Structure + Measurement



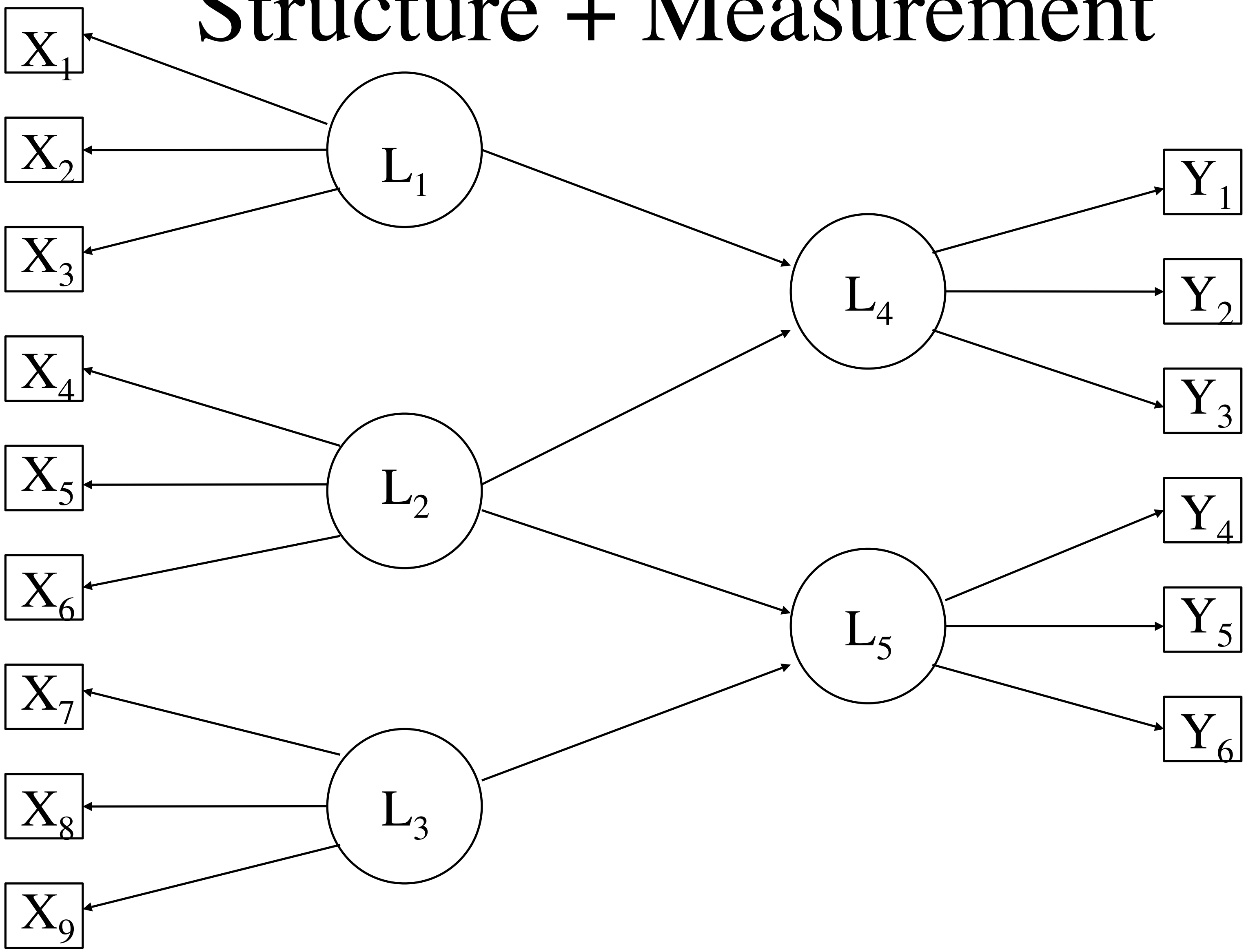
Measurement Model: Reliability



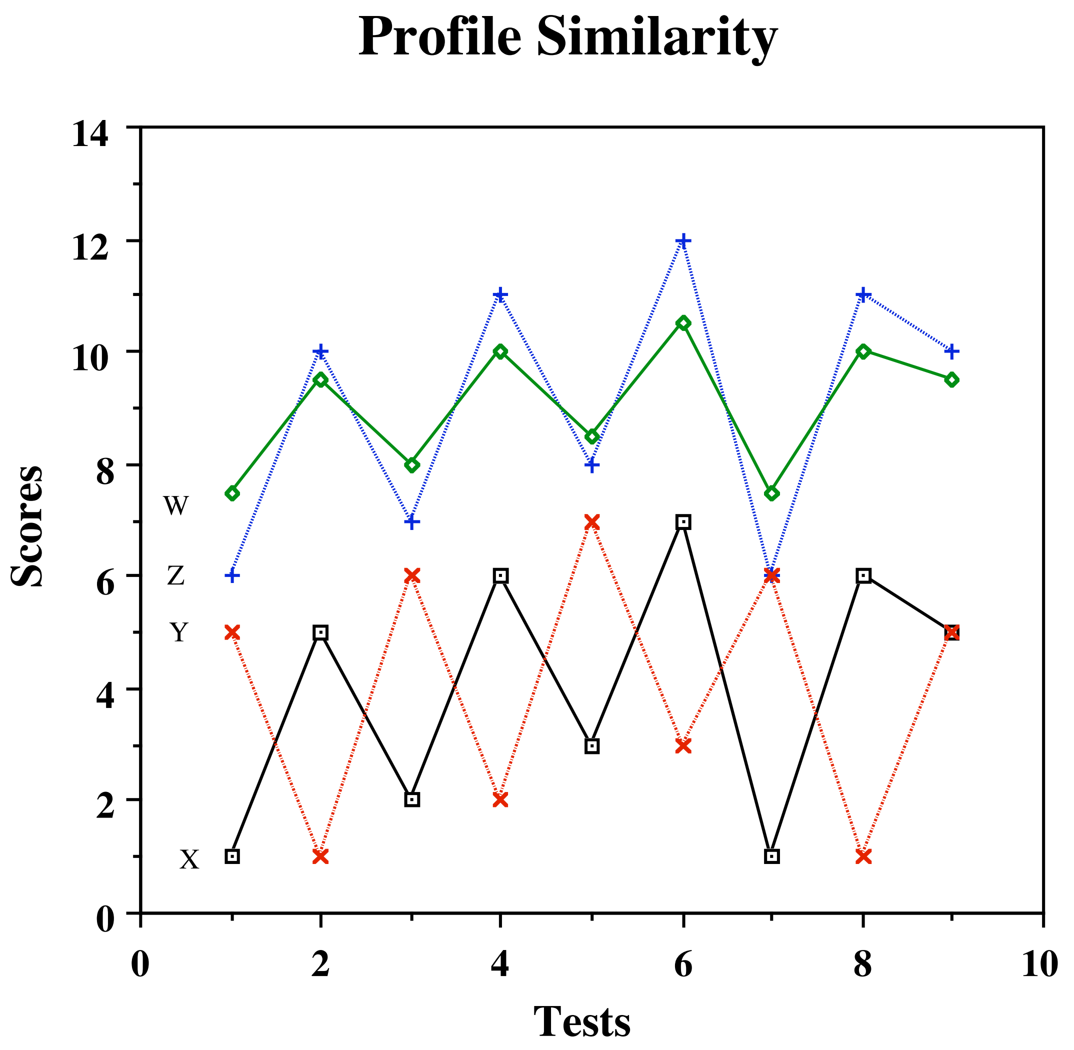
Structural Model: Regressions of Constructs



Structural Equation Modeling = Structure + Measurement



Measuring similarity



Similarity

Questions:

Given a set of scores on multiple tests (a subject profile), how should we measure the similarity between different profiles? What does it mean to have a similar profile?

What metric to use?

Minkowski Distances = $\sqrt[r]{\sum (X_i - Y_i)^r}$

r=1 city block metric ==> all distances equally important (no diagonals)
r=2 Euclidean metric ==> diagonals are shorter than sums
r>2 non-Euclidean ==> emphasizes biggest differences
r=∞ non-Euclidean ==> distance = biggest difference

Sources of Data

Structured interviews (e.g., SCID)

Other ratings

- Peer ratings
- supervisory ratings
- subordinate ratings

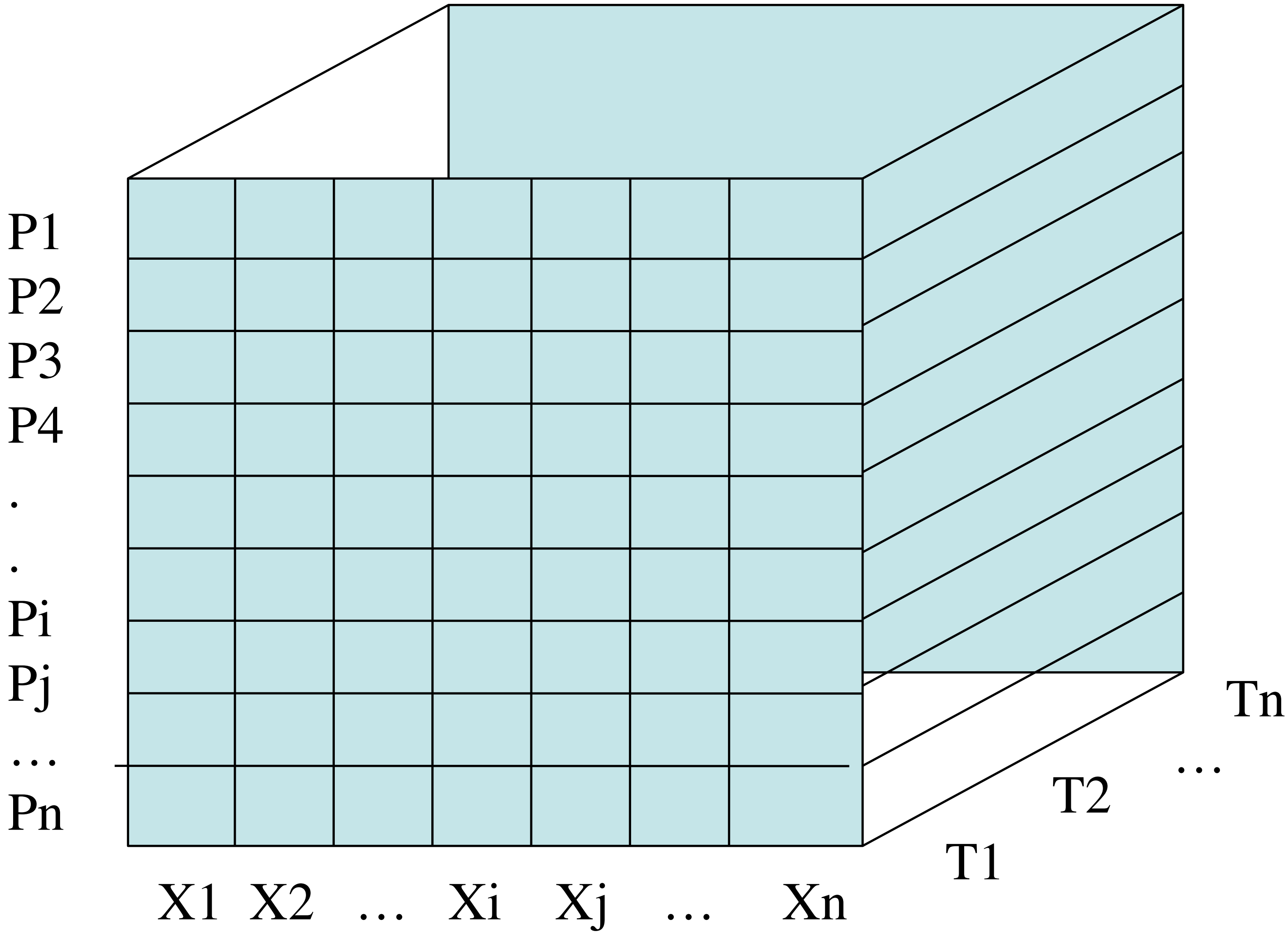
archival/unobtrusive measures

- unobtrusive measures
- historical record
  - GPA
  - Publications
  - Citations
- Neuropsychological
  - a) neurometrics
  - b) "lie detection”

Sources of Data

- Performance tests
  - OSS stress tests
  - New faculty job talks
  - Clinical graduate applicant interviews
  - Internships
  - Probationary Periods
- Web based instrumentation
  - self report
  - indirect (IAT)

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



Data = Model + Error

- Scaling: $d = f(ø) + e$
- Correlation and regression

  data level $d = ßx + c + e$. $ß = Cov/V$

  structure level $r_{xy} = Cov_{xy}/sqrt(V_x V_y)$

  goodness of fit of model = r^2
- Reliability

  classical $r_{xx} = r_{xø}^2$ domain sampling/ congeneric

Data = Model + Error

- Structural models

  Data = Structure + Error

  Structure = Validity * Reliability
- Analysis of Variance/Regression:

  data = αA + βB + χAB + ... + error
- Prediction and Selection

  performance = prediction + error
- Individual Differences in Scaling
  - Individual Data = Group Space*Individual Weight

Psychometric Theory: A conceptual Syllabus