Psychology 405: Psychometric Theory
Validity

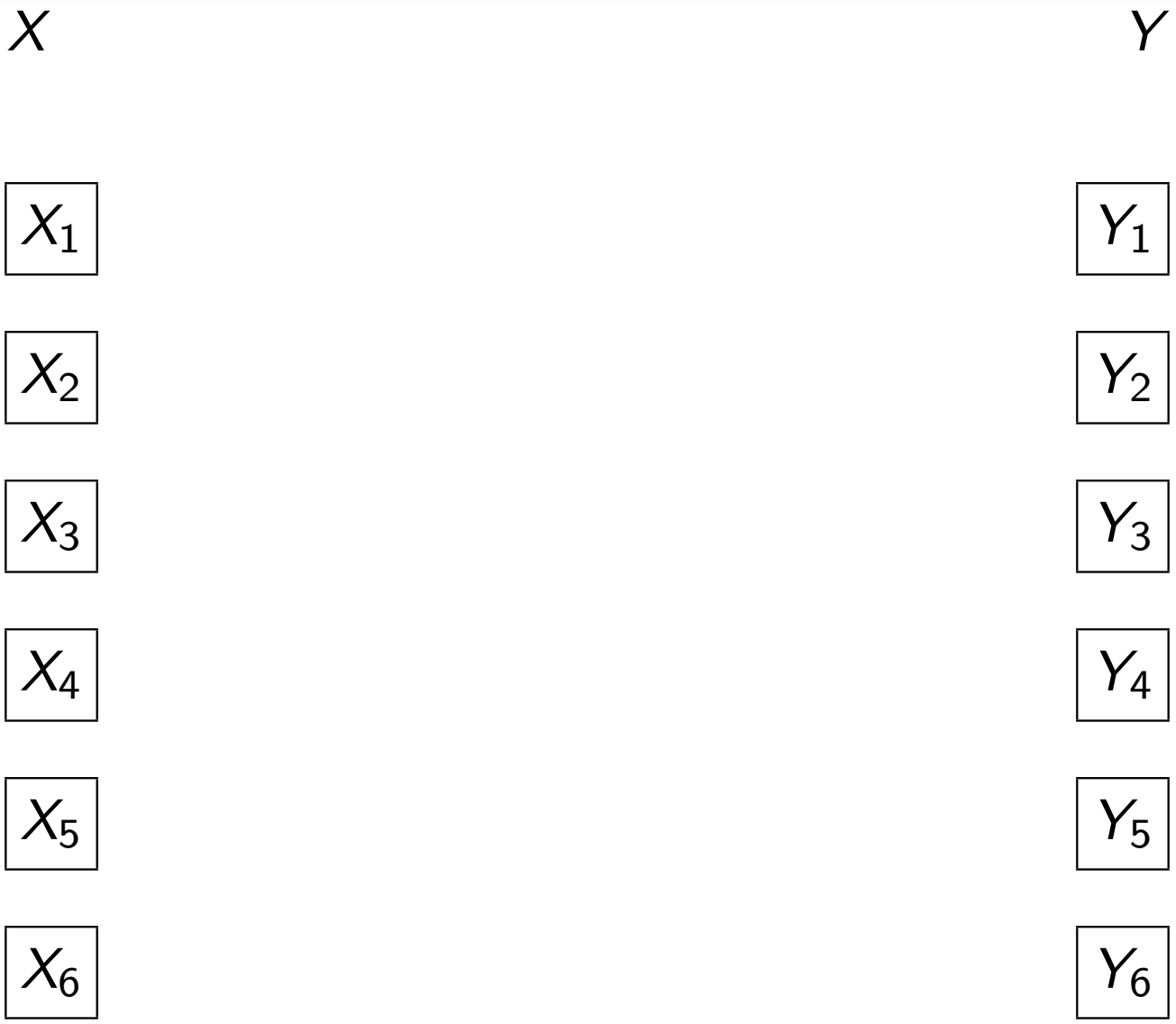
William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

May, 2016

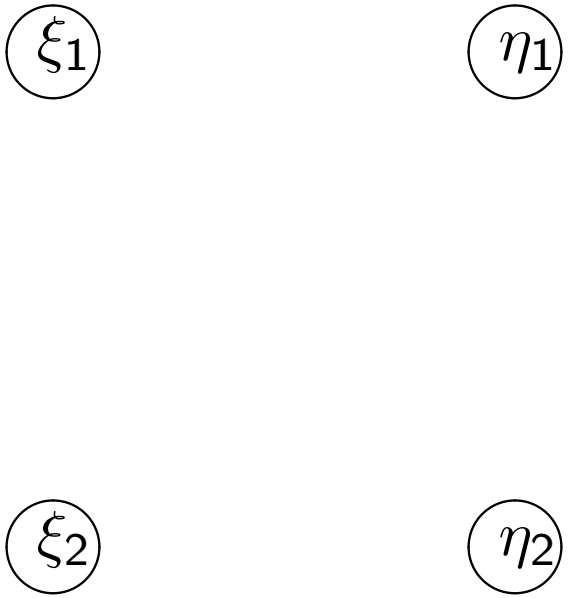
Outline

Observed Variables



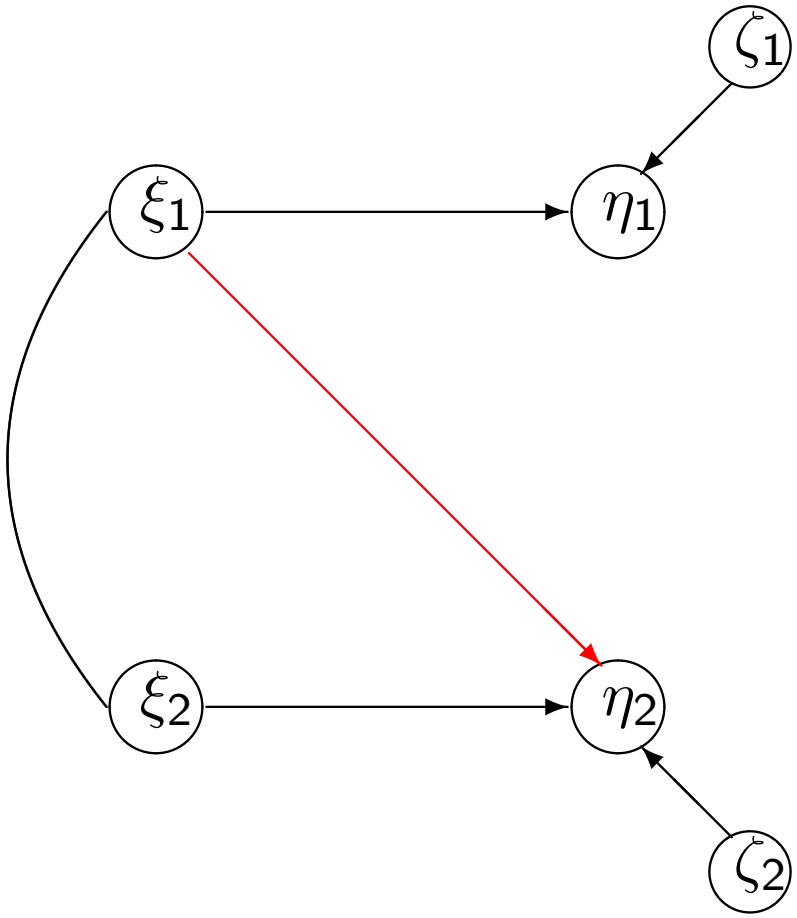
Latent Variables

ξ η



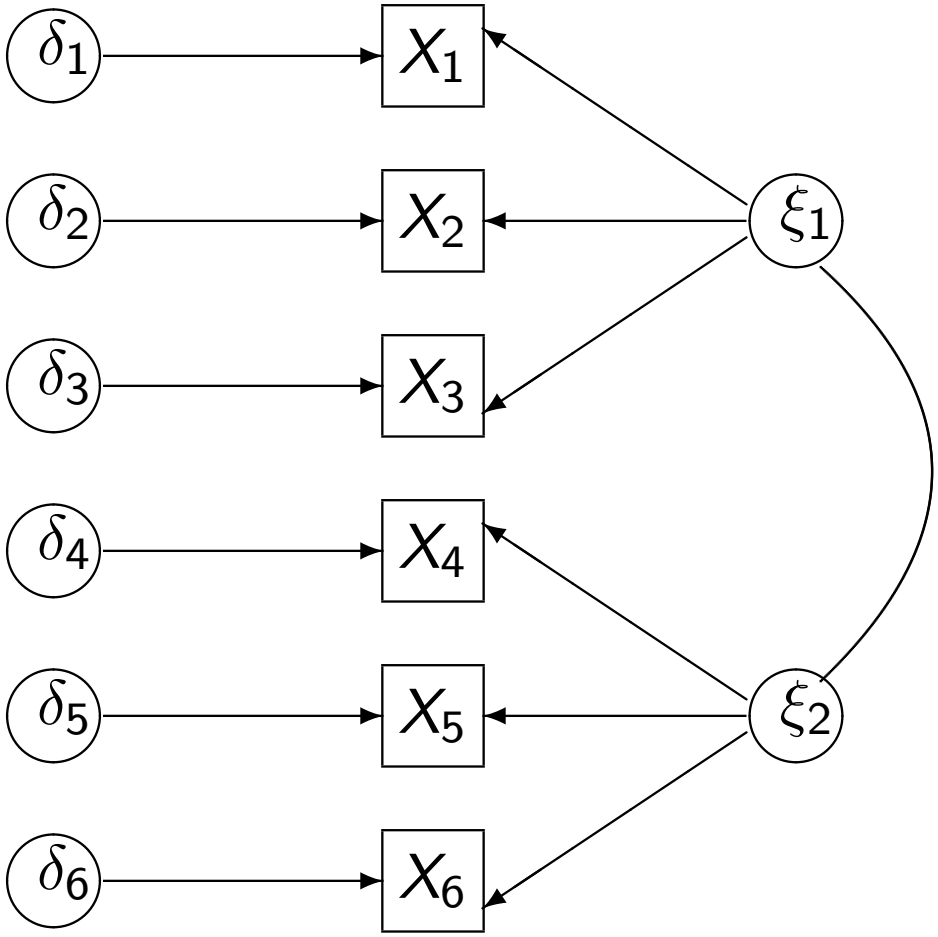
Theory: A regression model of latent variables

ξ η

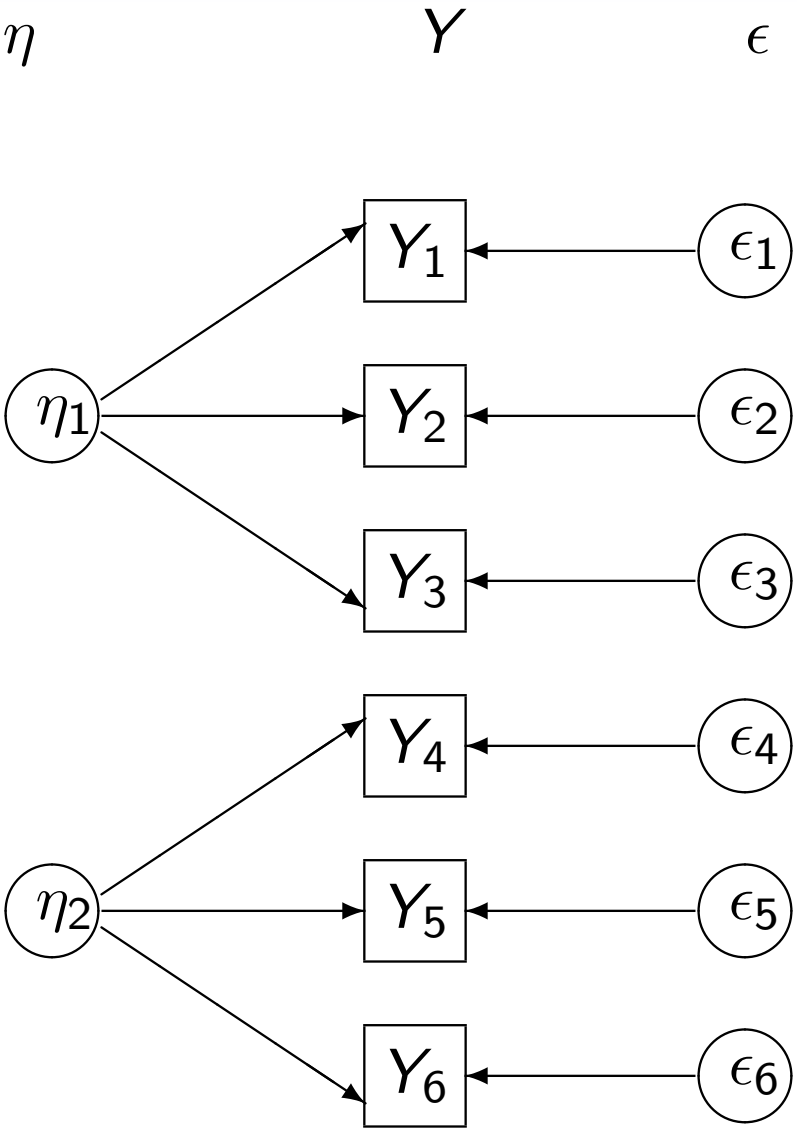


A measurement model for X – Correlated factors

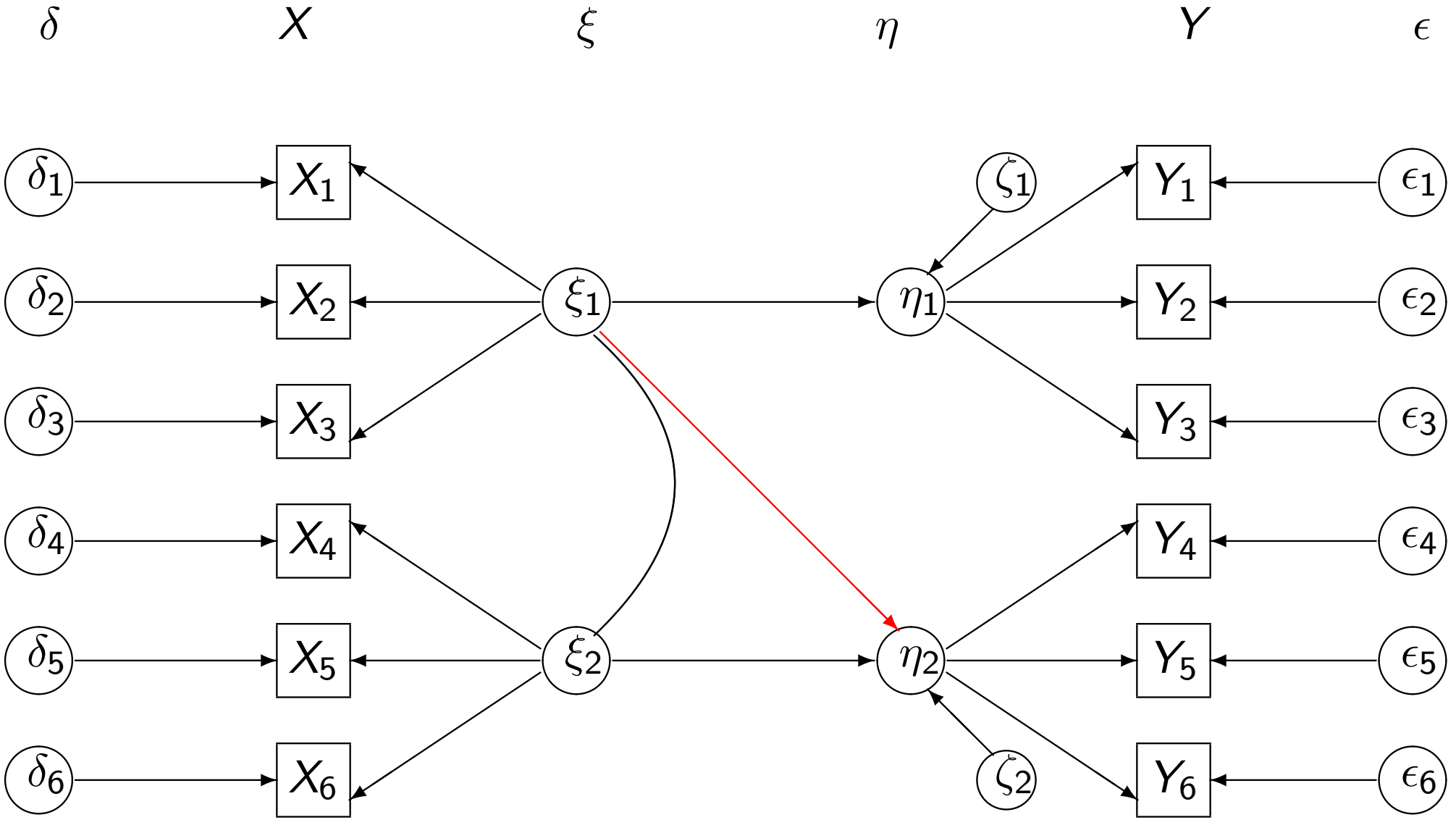
δ X ξ



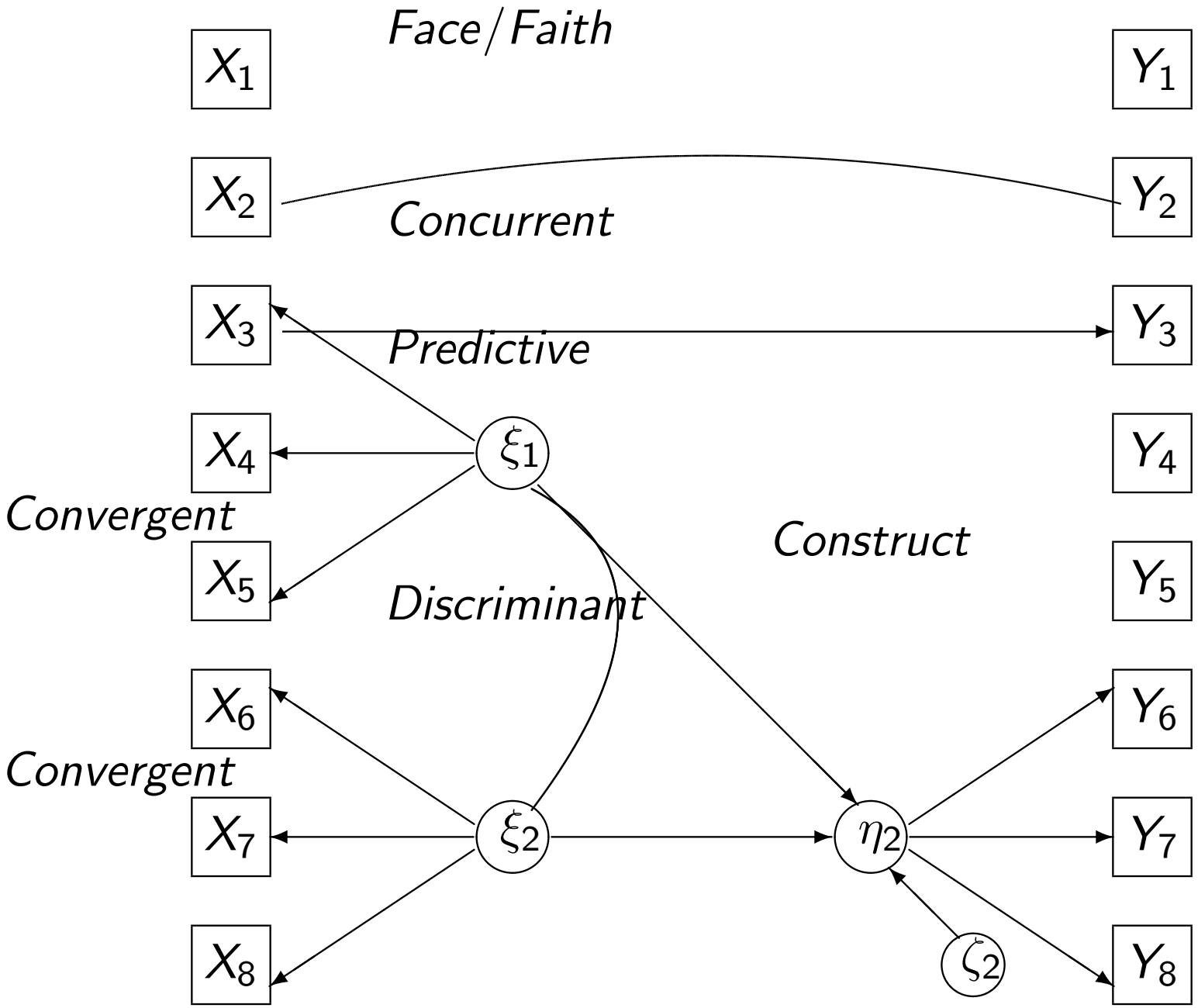
A measurement model for Y - uncorrelated factors



A complete structural model



Types of Validity



Face Validity

X_1

Face/Faith

Representative Content

Seeming relevance

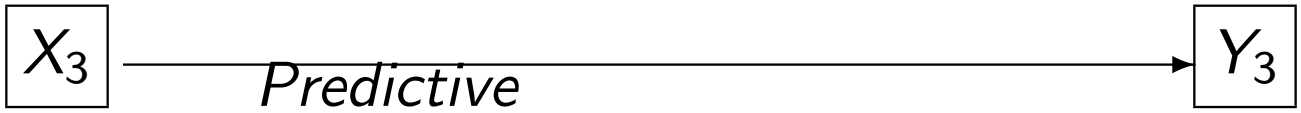
Concurrent Validity

Does a measure correlate with the criterion?

Need to define the criterion.

Assumes that what correlates now will have predictive value.

Predictive Validity



Does a measure correlate with the criterion?

Need to define the criterion.

Allow time to pass

Prediction

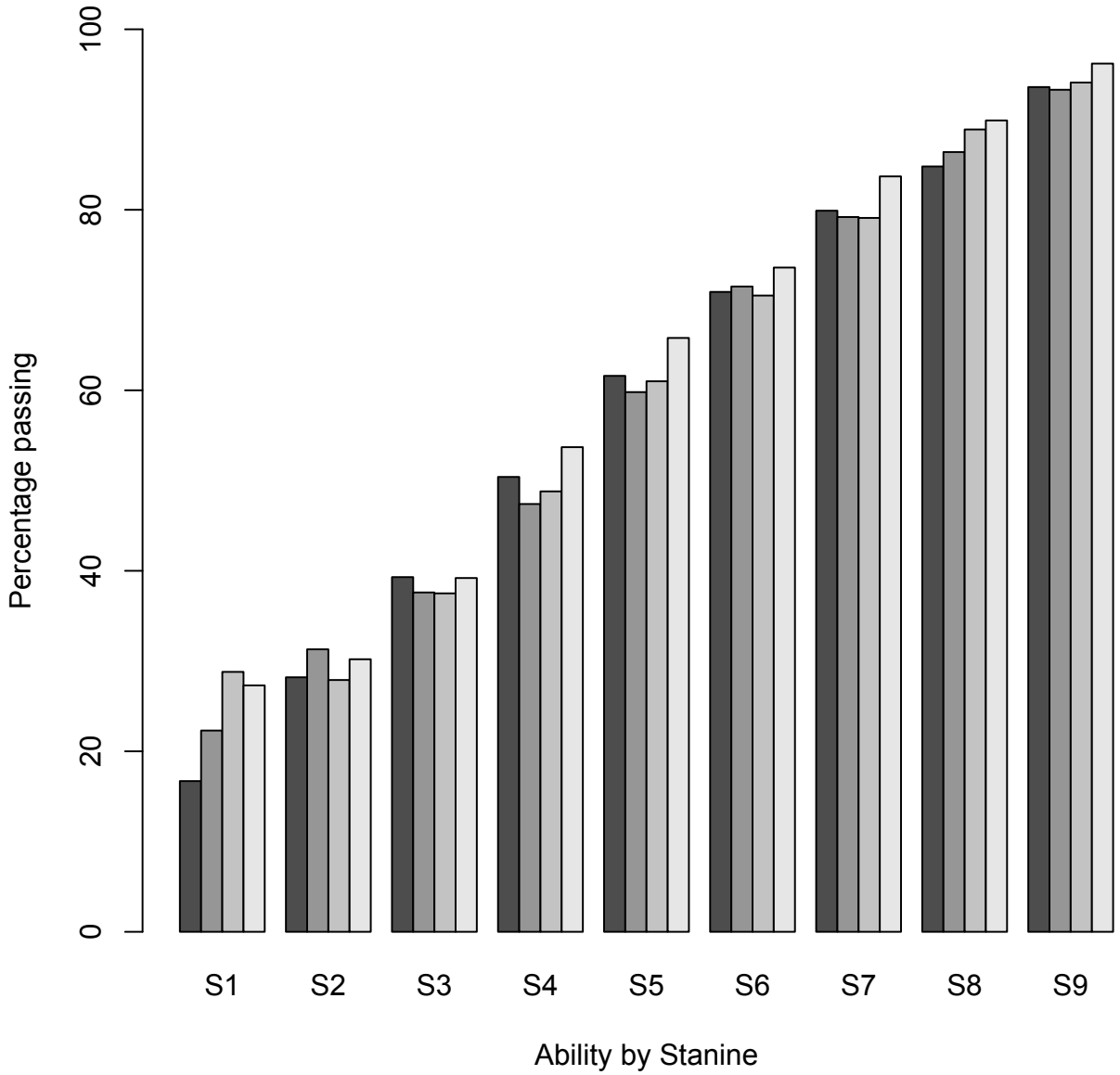
1. Continuous predictor, continuous criterion
   - Regression, multiple regression, correlation
   - Slope of regression implies how much change for unit change in predictor
2. Continuous predictor, dichotomous criterion
   - point bi-serial correlation
3. Dichotomous predictor, dichotomous outcome
   - Phi

Classics in Prediction and selection

1. Gideon's selection of soldiers
2. OSS and Army Air Corps selection studies
3. Kelly and Fiske (1950) selection of psychology students
4. Astronaut selection
5. Peace Corps selection

Gideon's assessment

The assessment of pilots – how to show a .45 correlation makes a difference



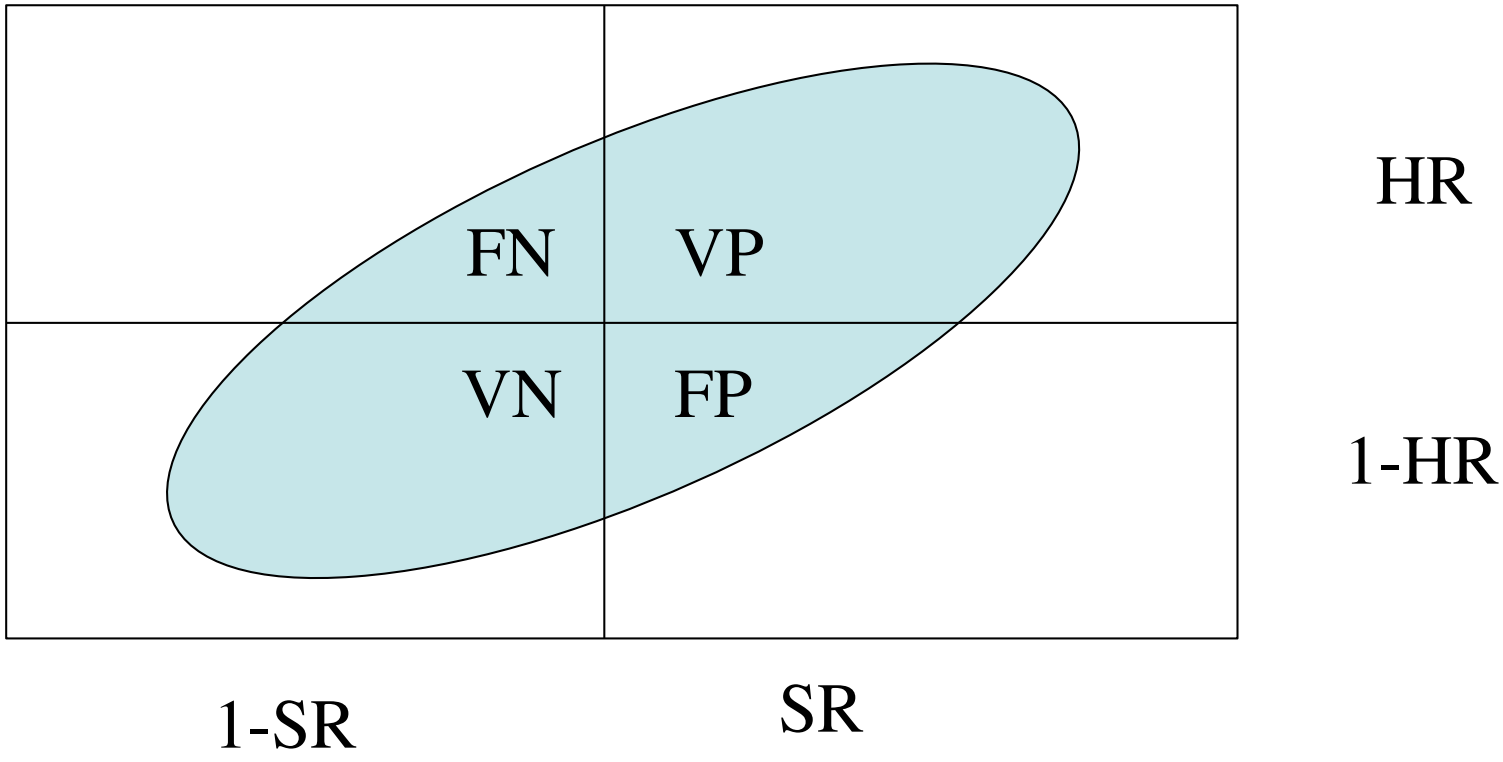
Predicting clinical psychologists – Kelly and Fiske

1. Multiple predictors of graduate school performance: Kelly and Fiske (1950), Multiple predictors
2. Ability, Interests, temperament (each with r $\approx$ .2 -.25) have multiple R of .4-.5
3. Are they able, interested and stable?
4. Kuncel et al. (2001)

Predictive and Concurrent Validity and Decision Making

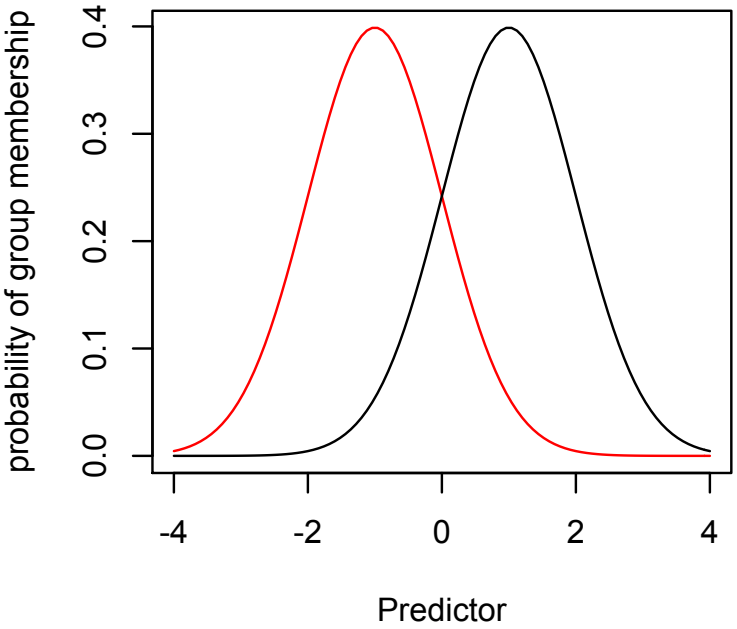
Hit Rate = Valid Positive + False Negative

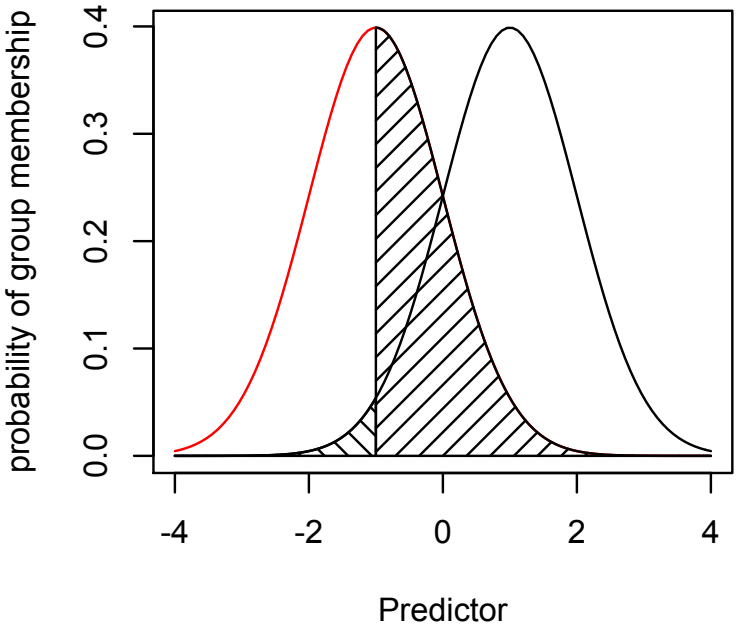
Selection Ratio = Valid Positive + False Positive



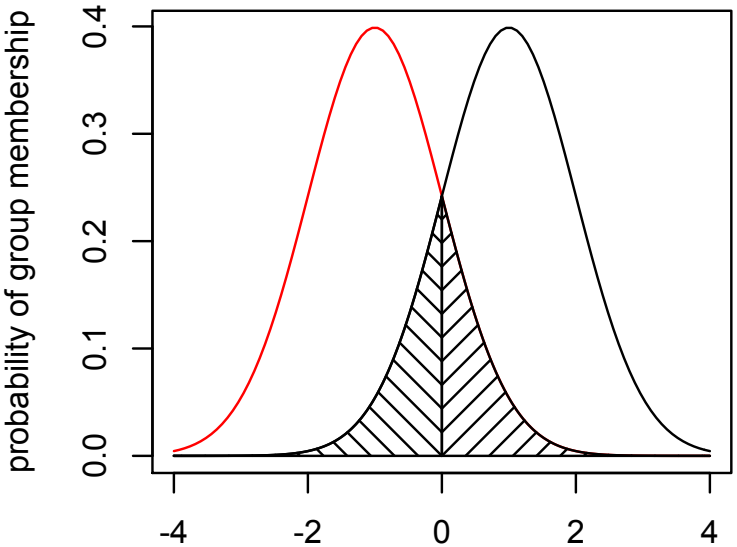
Phi =(VP - HR*SR) /sqrt(HR*(1-HR)*(SR)*(1-SR)

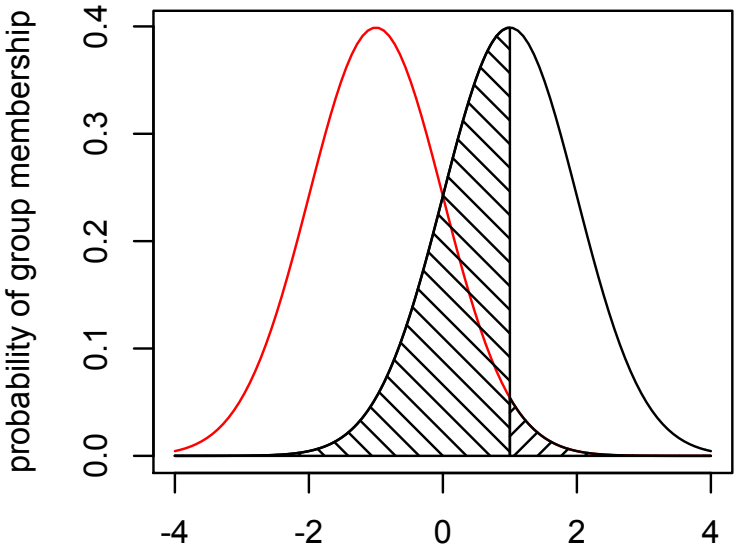
Validity as decision making



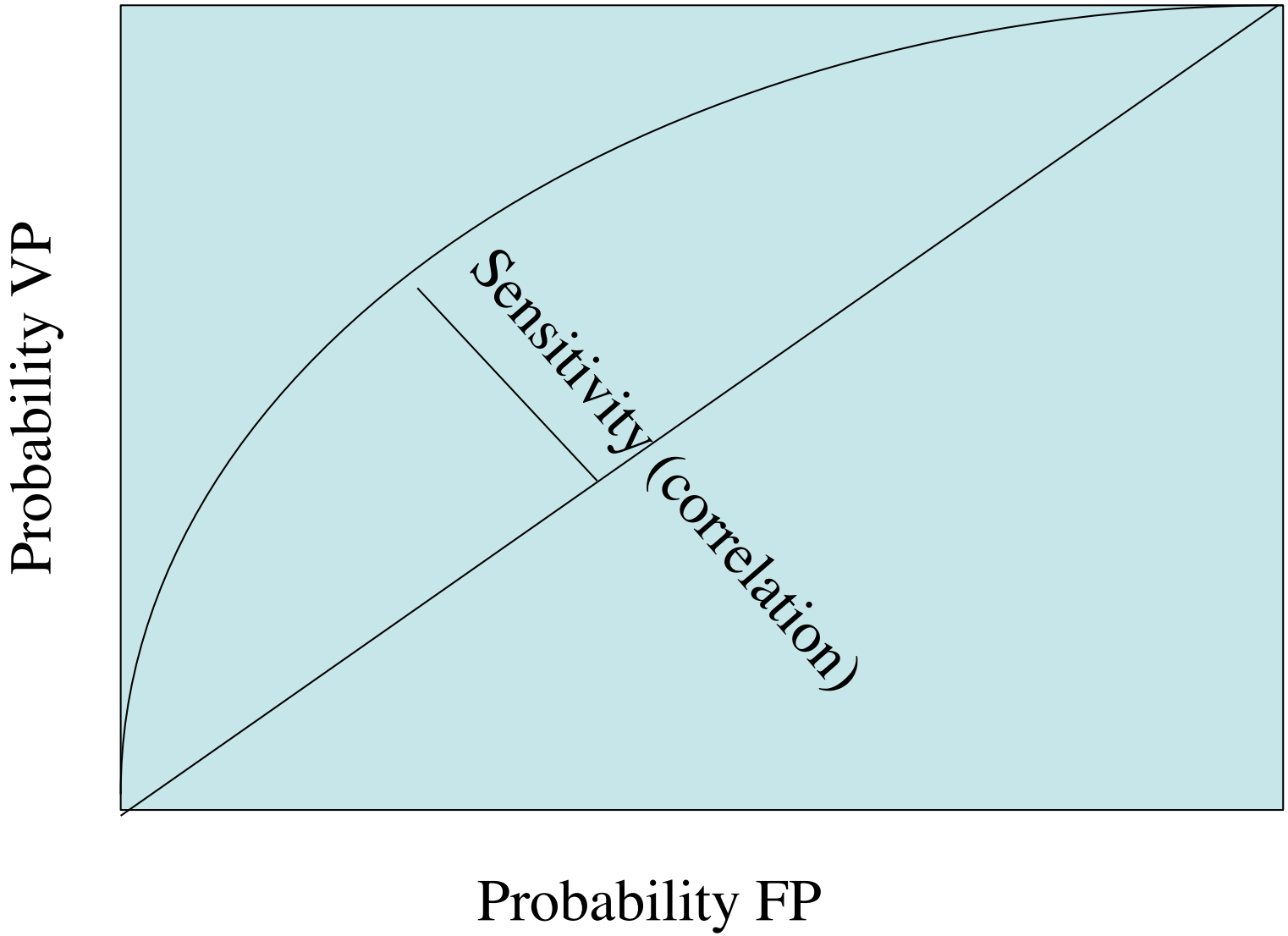


Trading off Valid positives for False Positives





Decision Theory and Signal Detection



Signal detection theory

- d prime and beta
  - d prime maps to the correlation
  - beta maps to selection ratio
- type I and type II error
  - Need to consider utility of types of error

Predictive Validity and Decision Theory

			State of world
	FN	VP	Hit rate
	VN	FP	1-HR
Decision	1-SR	Selection Ratio	

Predictive Validity, Utility and Decision Theory

			State of world
	FN $*U_{FN}$	VP $*U_{VP}$	Hit rate
	VN $*U_{VN}$	FP$*$ U_{FP}	1-HR
Decision	1-SR	Selection Ratio	

Utility of test = VP $*U_{VP}$ + VN $*U_{VN}$ + FN $*U_{FN}$ + FP$*$ U_{FP} - Cost of test

Decisions for institutions, advice for individuals

			State of world
	FN $*U_{FN}$	VP $*U_{VP}$	Hit rate
	VN $*U_{VN}$	FP$*$ U_{FP}	1-HR
Decision	1-SR	Selection Ratio	

Utility of test = VP $*U_{VP}$ + VN $*U_{VN}$ + FN $*U_{FN}$ + FP$*$ U_{FP} - Cost of test

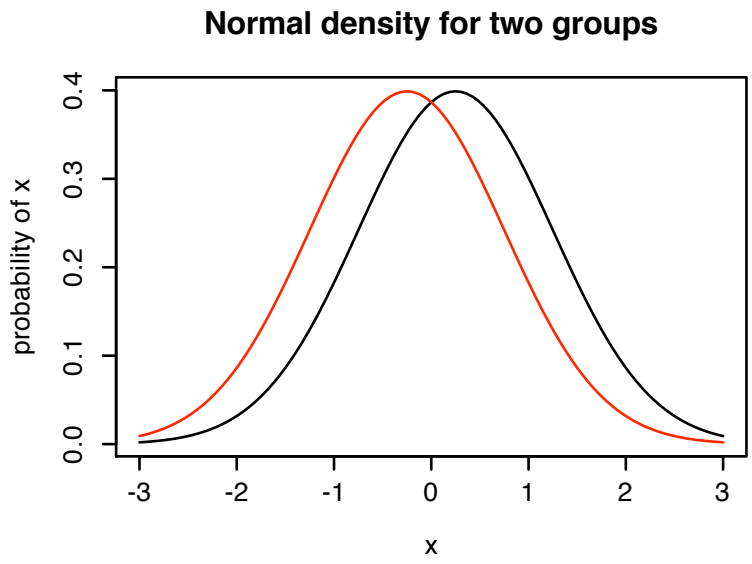
Decision making and the benefit of extreme selection ratios

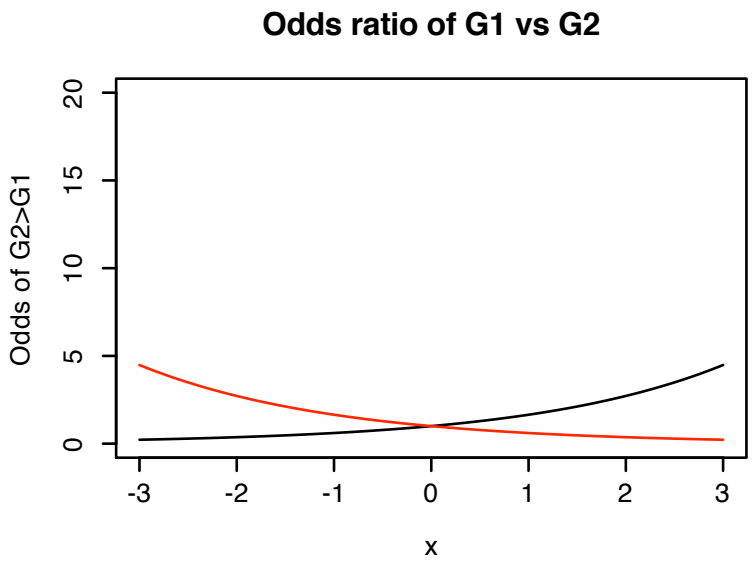
- Typical traits are approximated by a normal distribution.
- Small differences in means or variances can lead to large differences in relative odds at the tails
- Accuracy of decision/prediction is higher for extreme values.
- Do we infer trait mean differences from observing differences of extreme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

- (code for these graphs at personality-project.org/r/extreme.r)

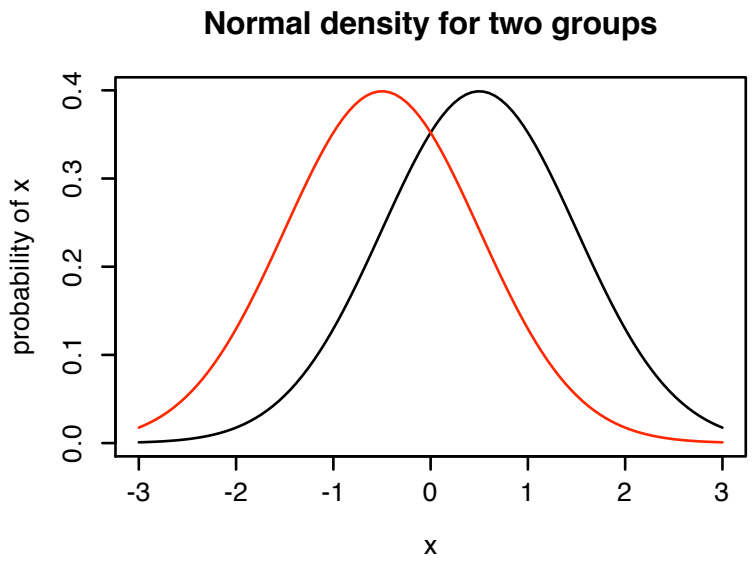
Odds ratios as f(mean difference, extremity)

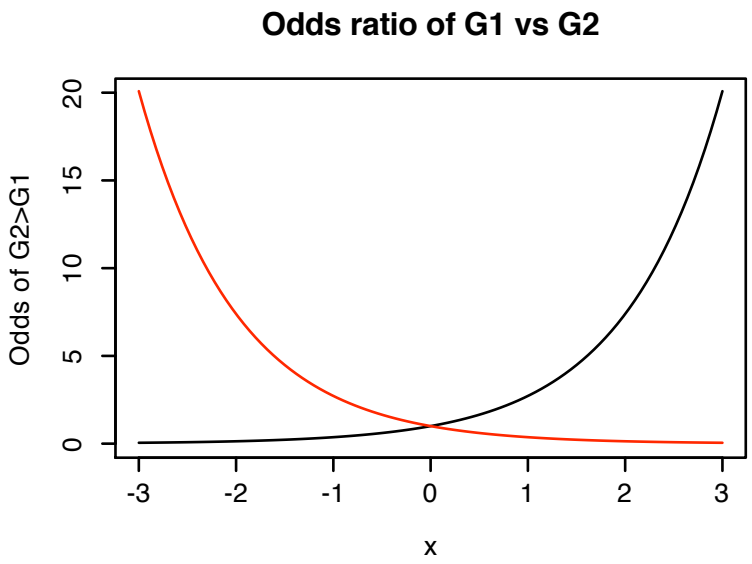
Difference =.5 sigma





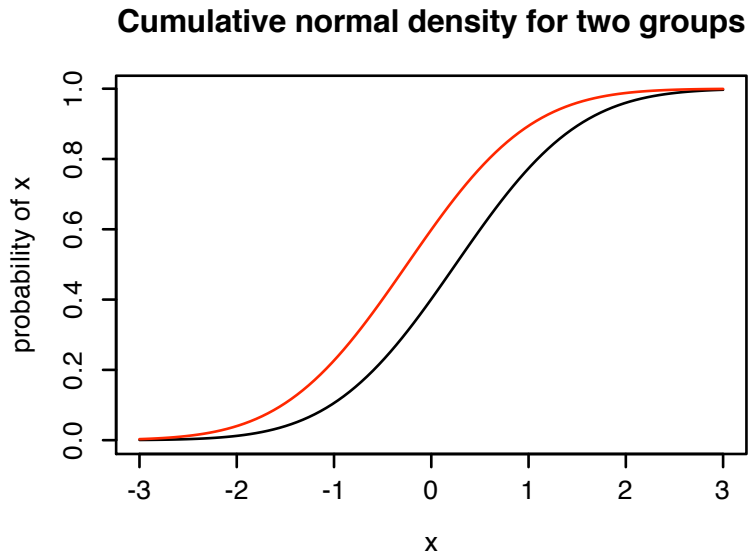
Difference =1.0 sigma

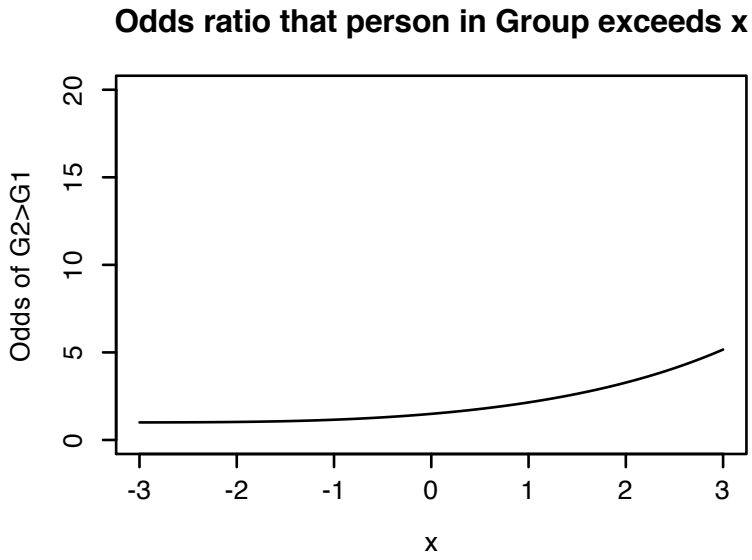




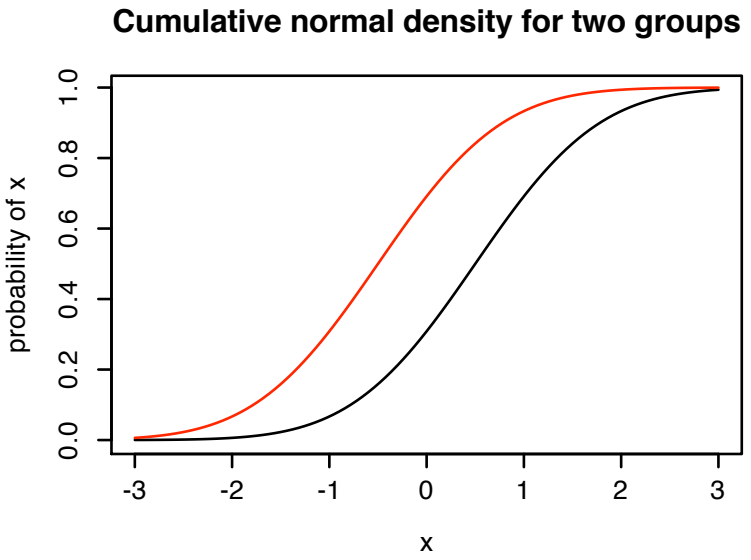
The effect of group differences on likelihod of extreme scores

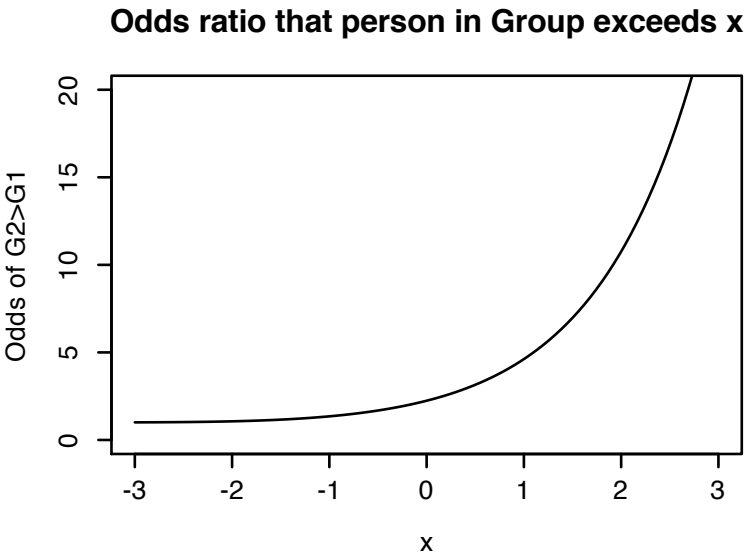
Difference =.5 sigma



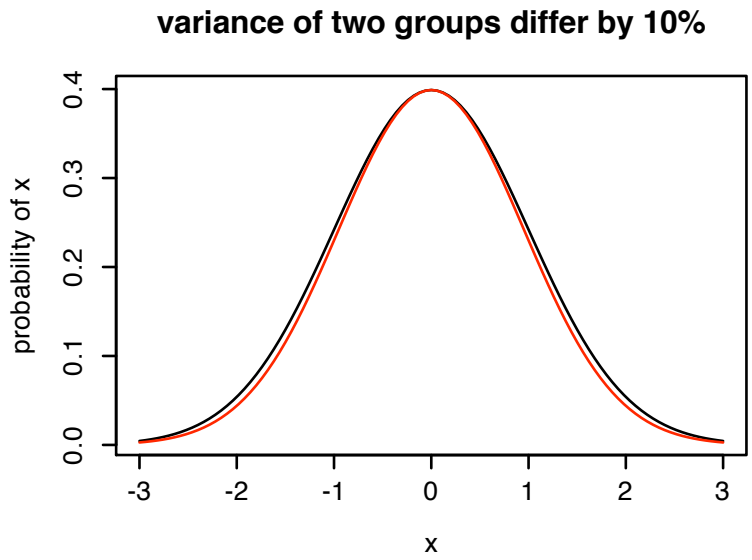


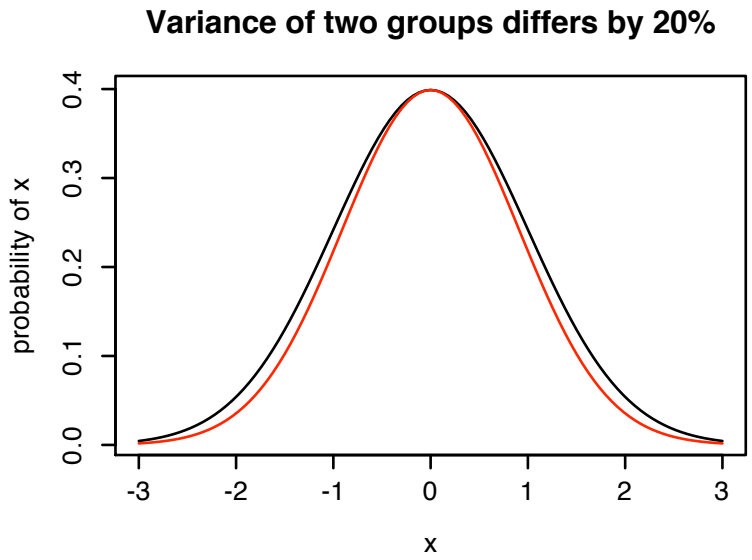
Difference =1.0 sigma

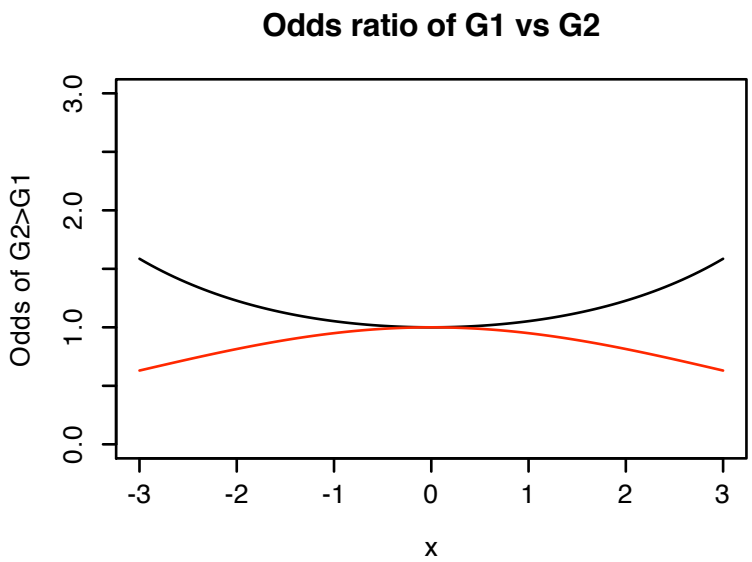


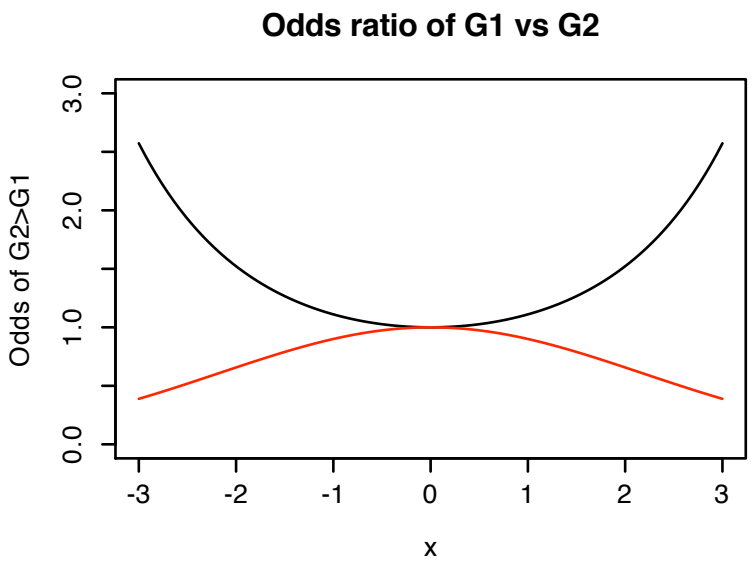


The effect of differences of variance on odds ratios at the tails





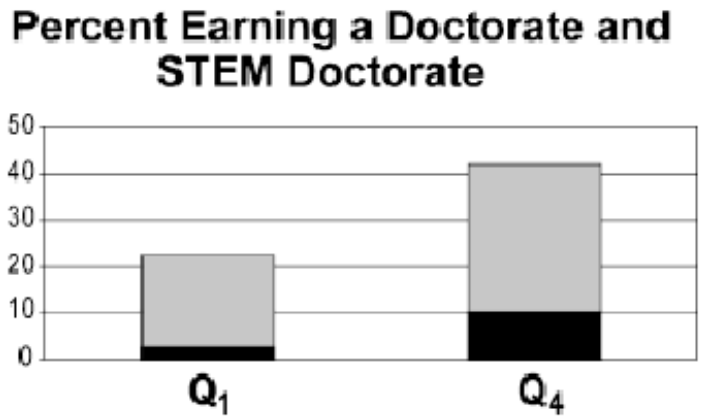


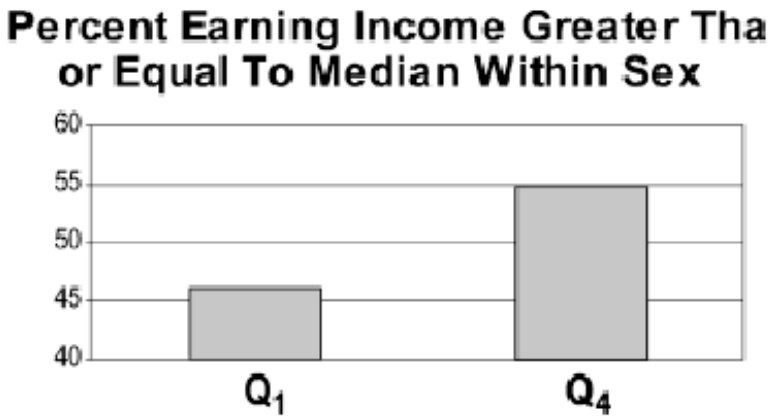


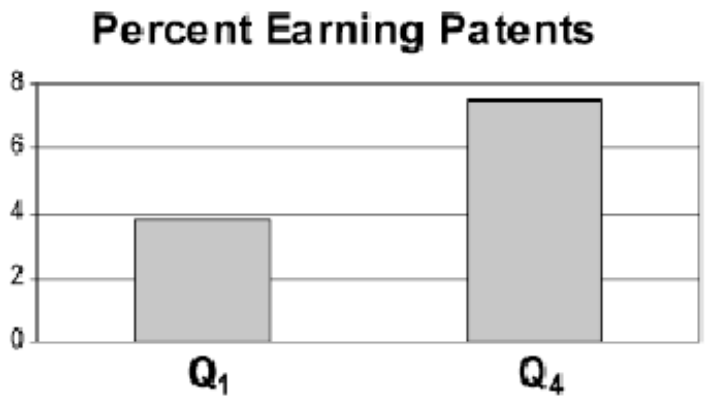
Restriction of range

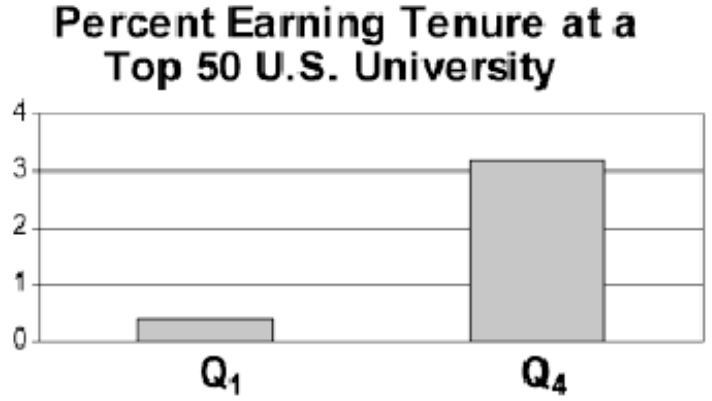
- Validity of SAT is partially limited by range restriction. (see Lubinski and Benbow)
- Consider giving SATs to 12-13 year olds
  - –SAT M ≥ 390 or SAV V ≥ 370 (top 1 in 100)
  - –SAT M ≥ 500 or SAV V ≥ 430 (top 1 in 200)
  - –SAT M ≥ 700 or SAV M ≥ 430 (top 1 in 10,000)

Predictions within top student group



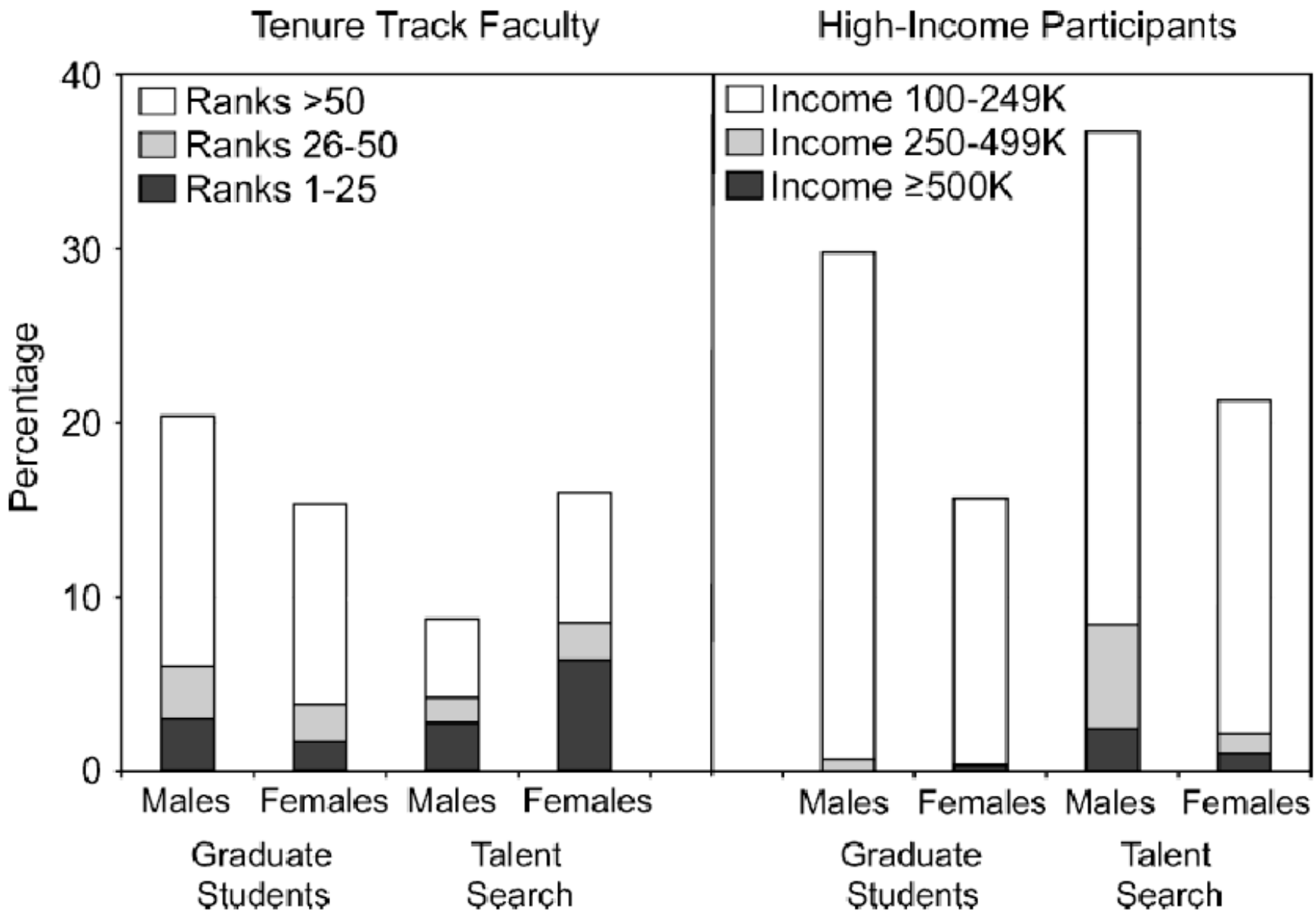




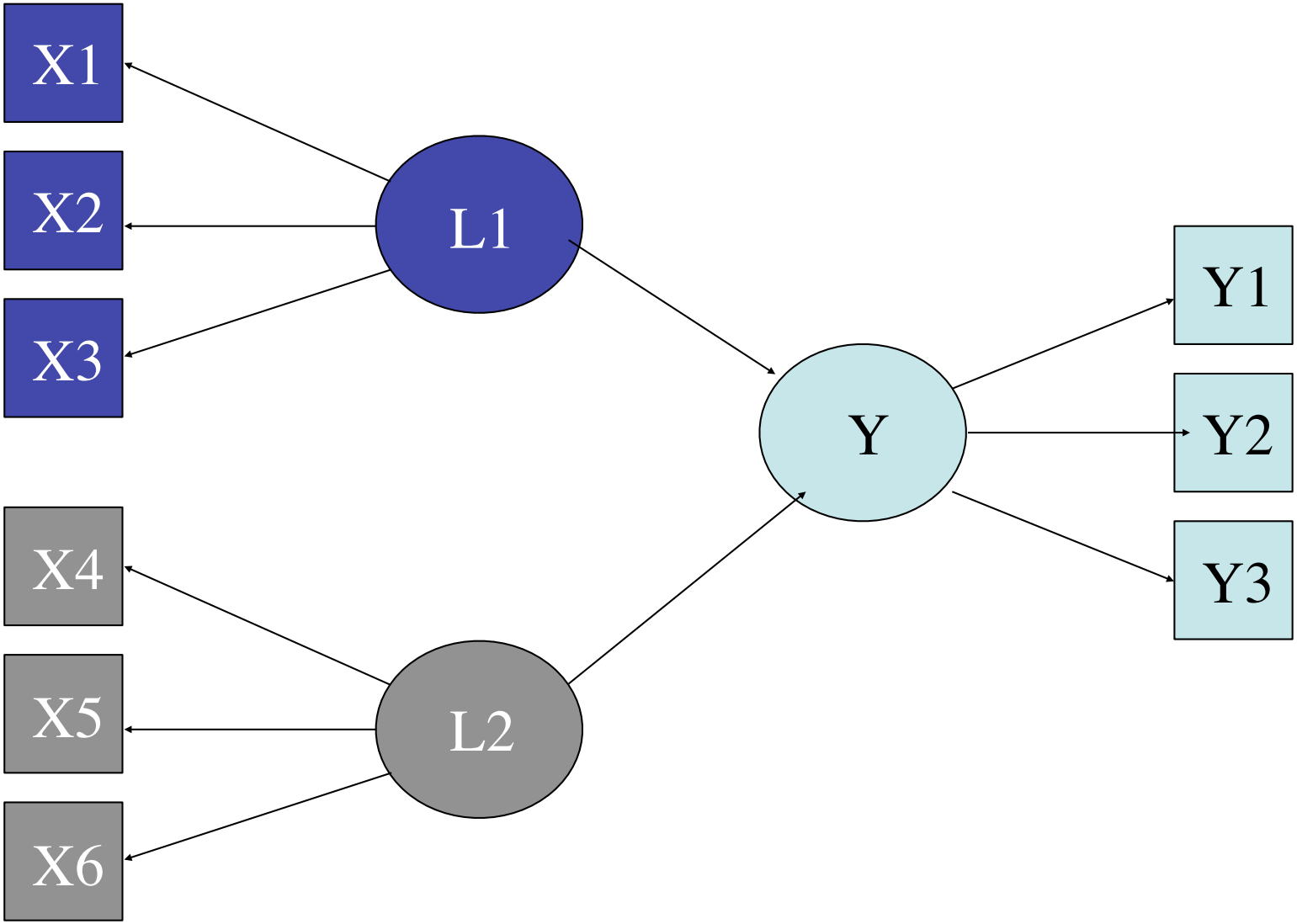


Validity continues even among top 1%

Validity over 25 years



Construct Validity: Convergent, Discriminant, Incremental



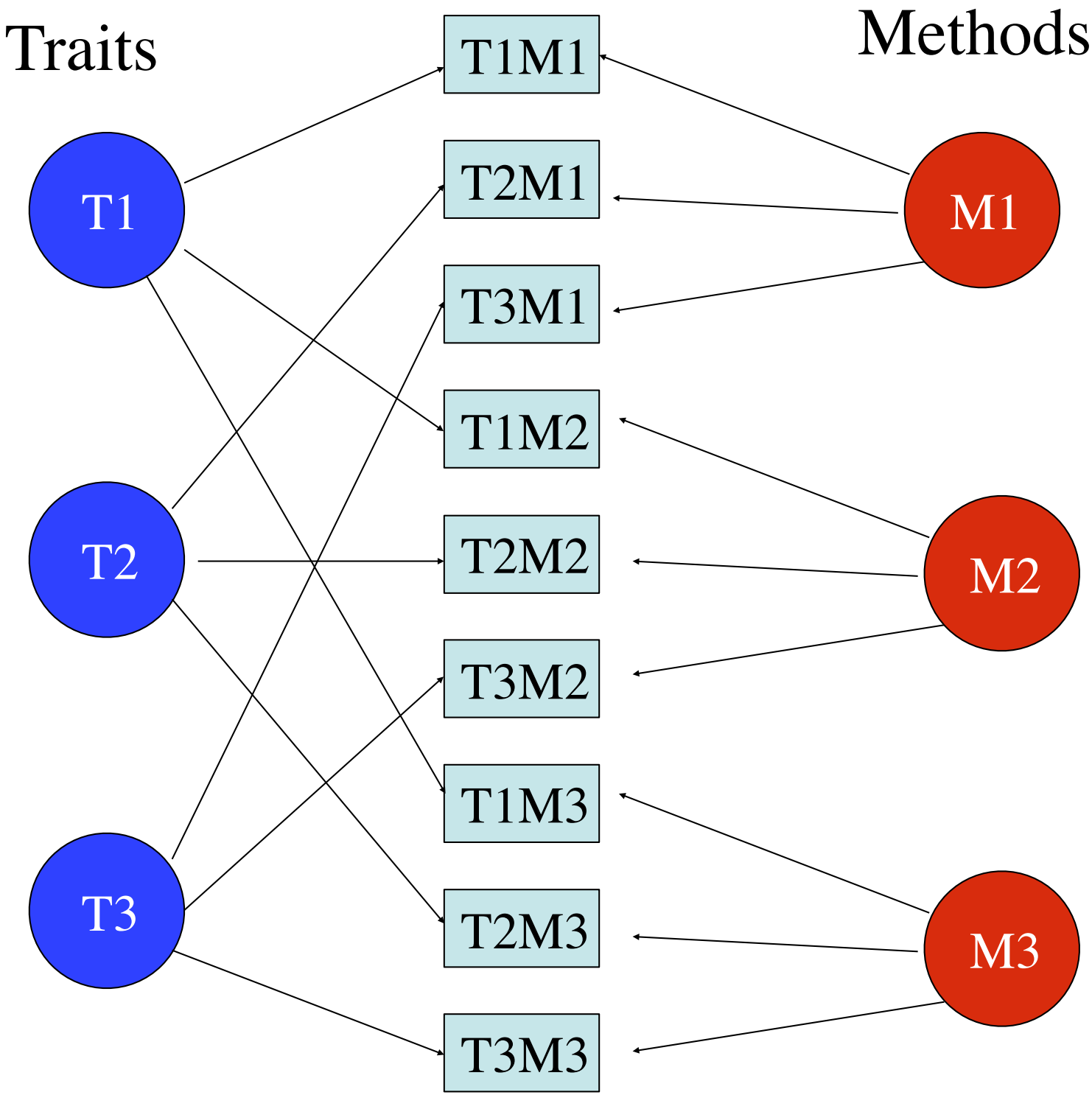
Multi-Trait, Multi-Method Matrix

	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


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