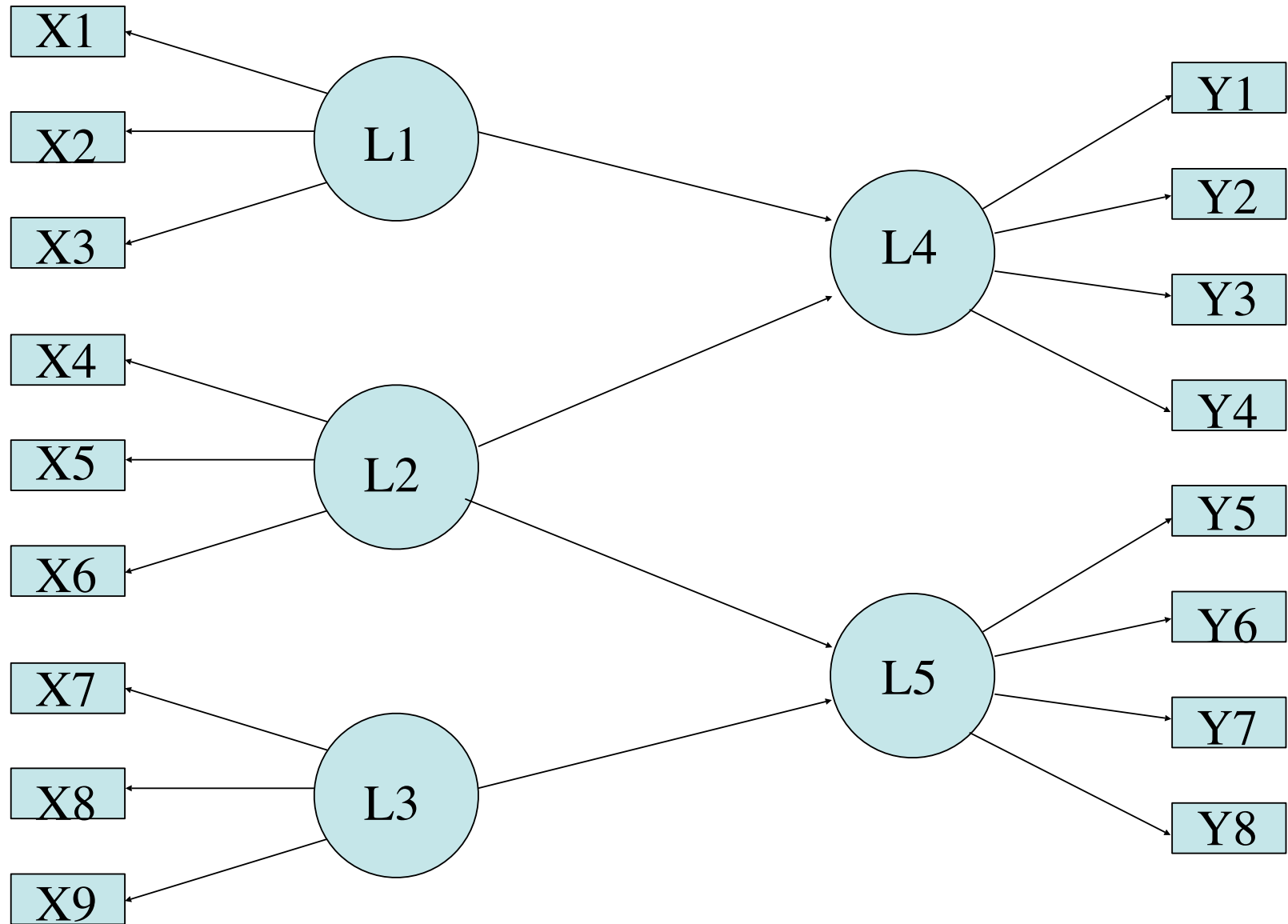


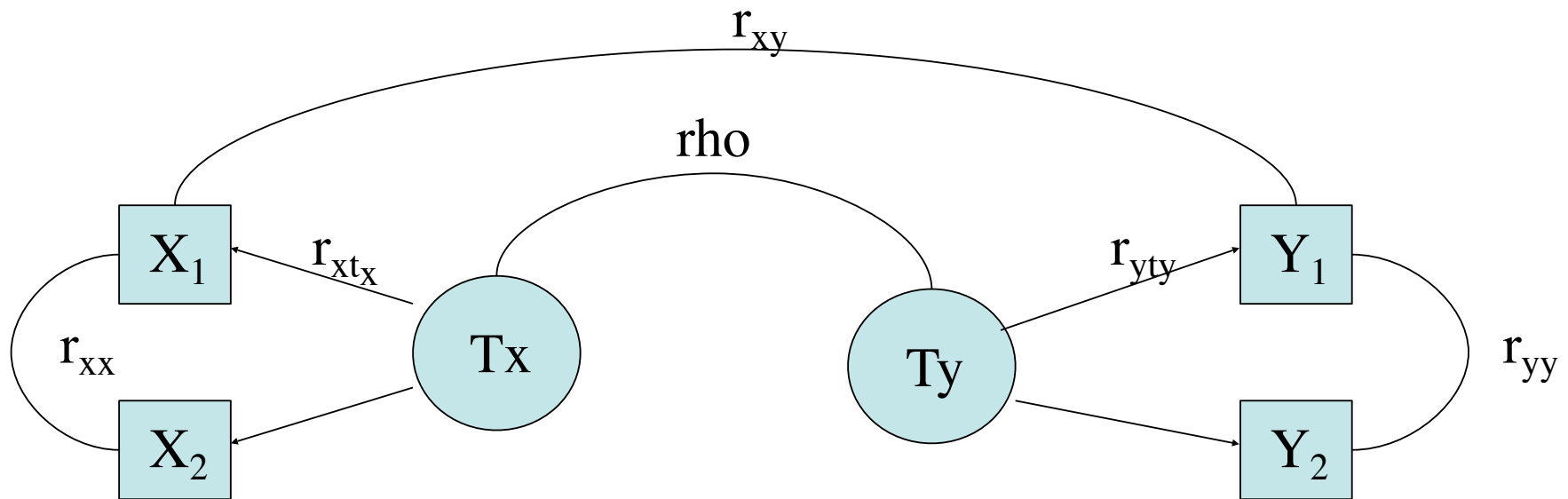
Validity

Face, Concurrent, Predictive, Construct

Psychometric Theory: A conceptual Syllabus



Reliability- Correction for attenuation

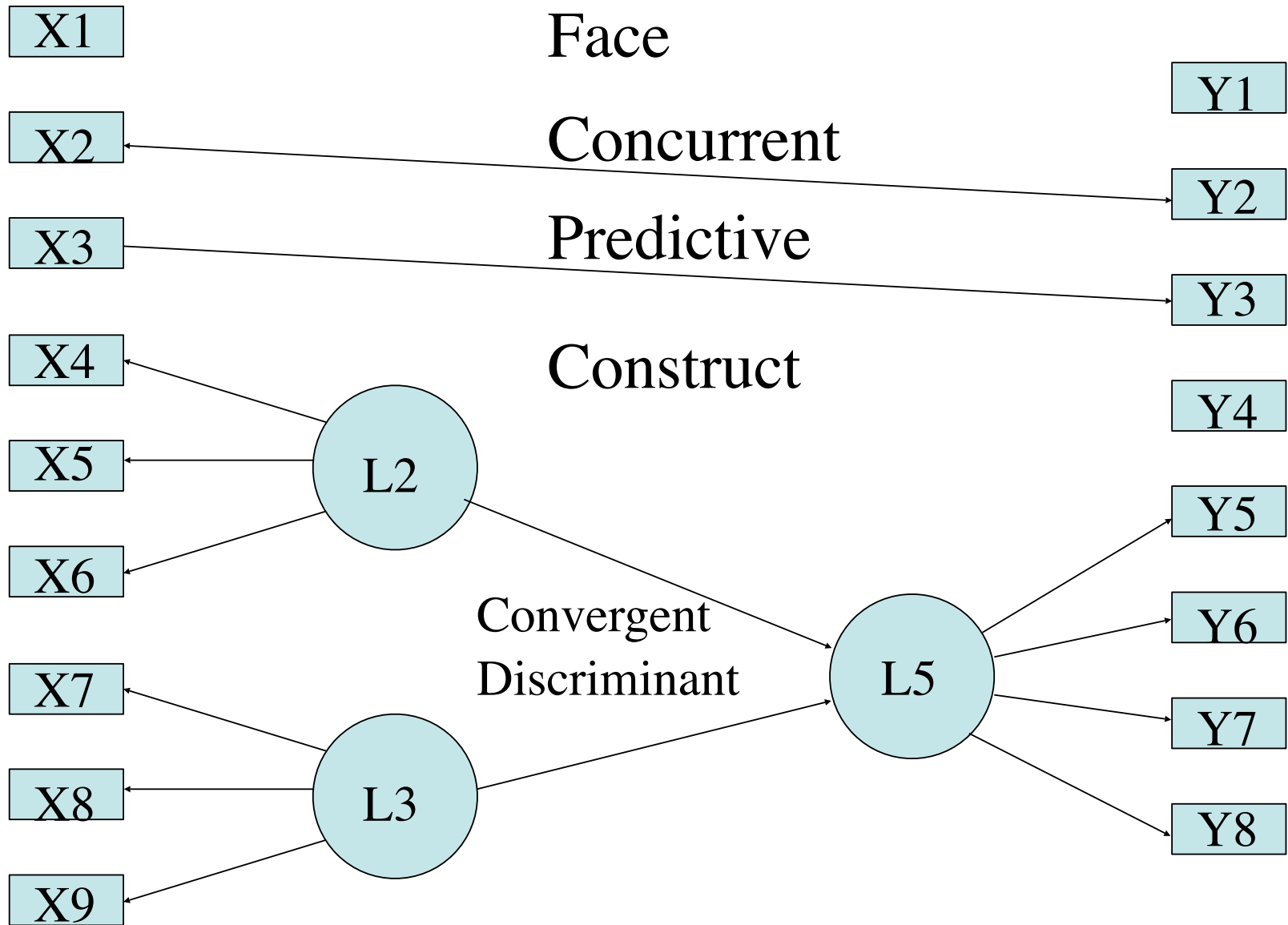


$$r_{xt_x} = \sqrt{r_{xx}}$$

$$r_{yty} = \sqrt{r_{yy}}$$

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

Types of Validity: What are we measuring

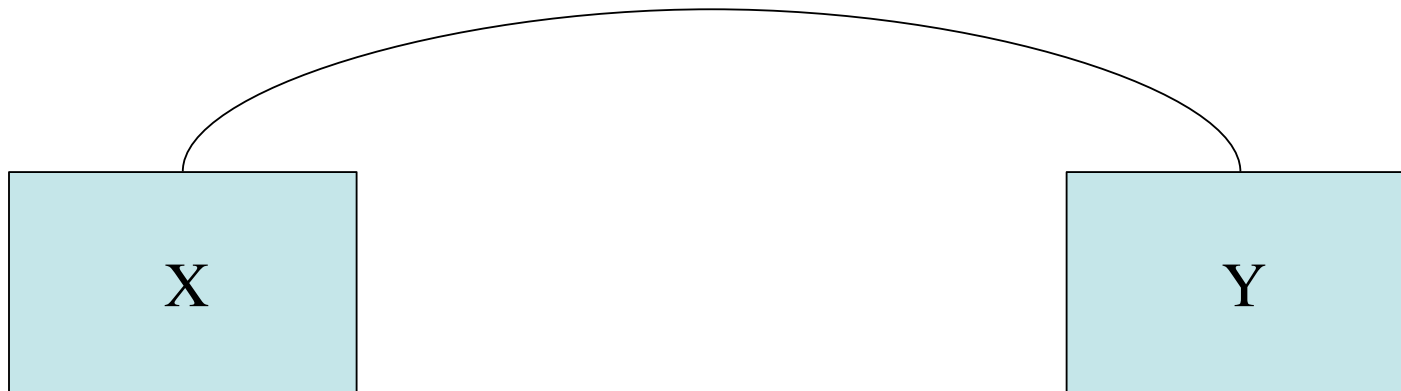


Face (Faith Validity)



- 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.
- Requires waiting for time to pass.

Type of correlation

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

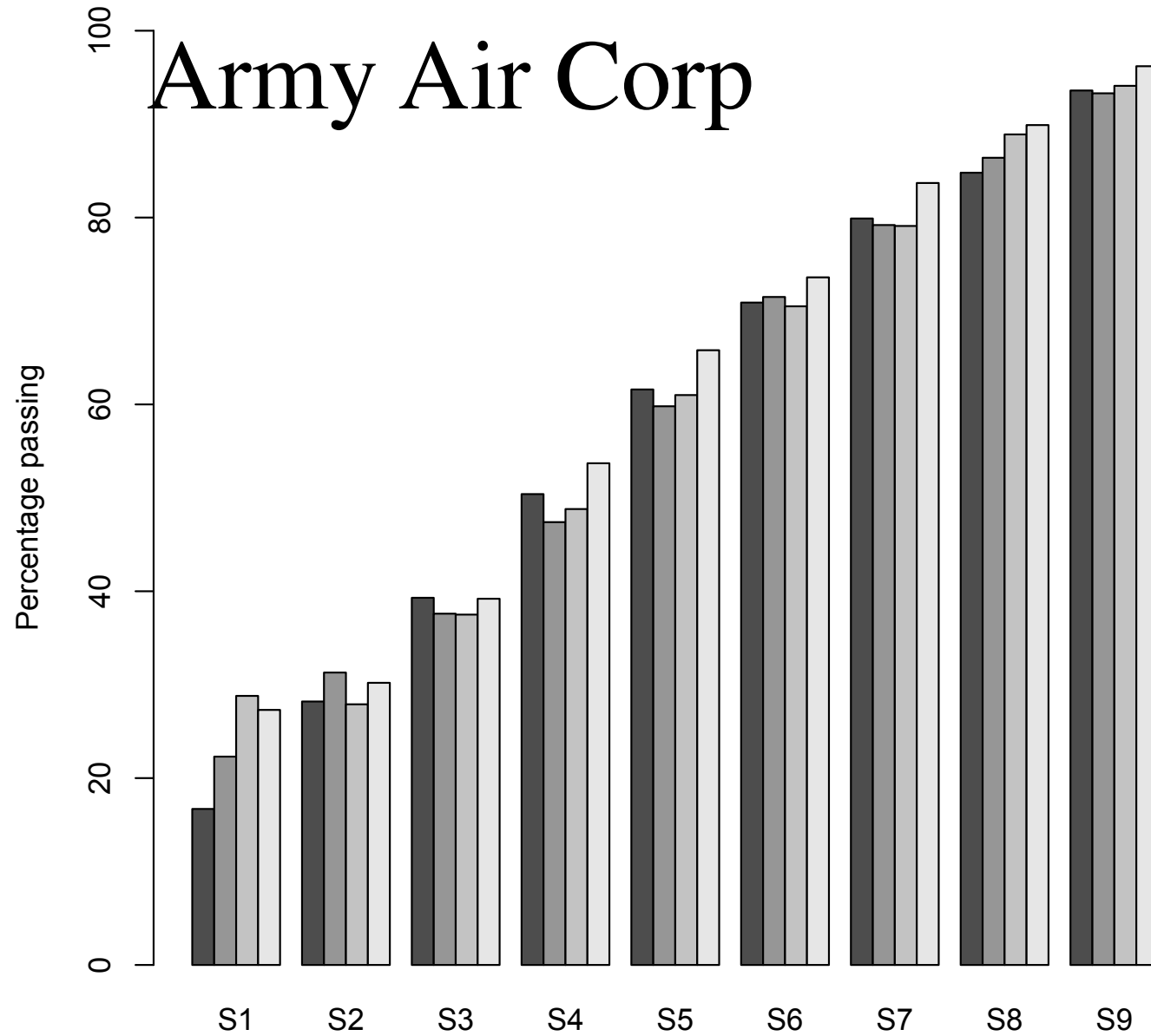
Classics in Selection/Assessment

- Gideon's selection of soldiers
- OSS and Army Air Corps selection studies
- Kelly and Fiske (1950) selection of psychology students
- Astronaut selection
- Peace Corps selection

Gideon's assessment technique



Army Air Corp



$r = .45$

The power of a good graphic

Ability by Stanine

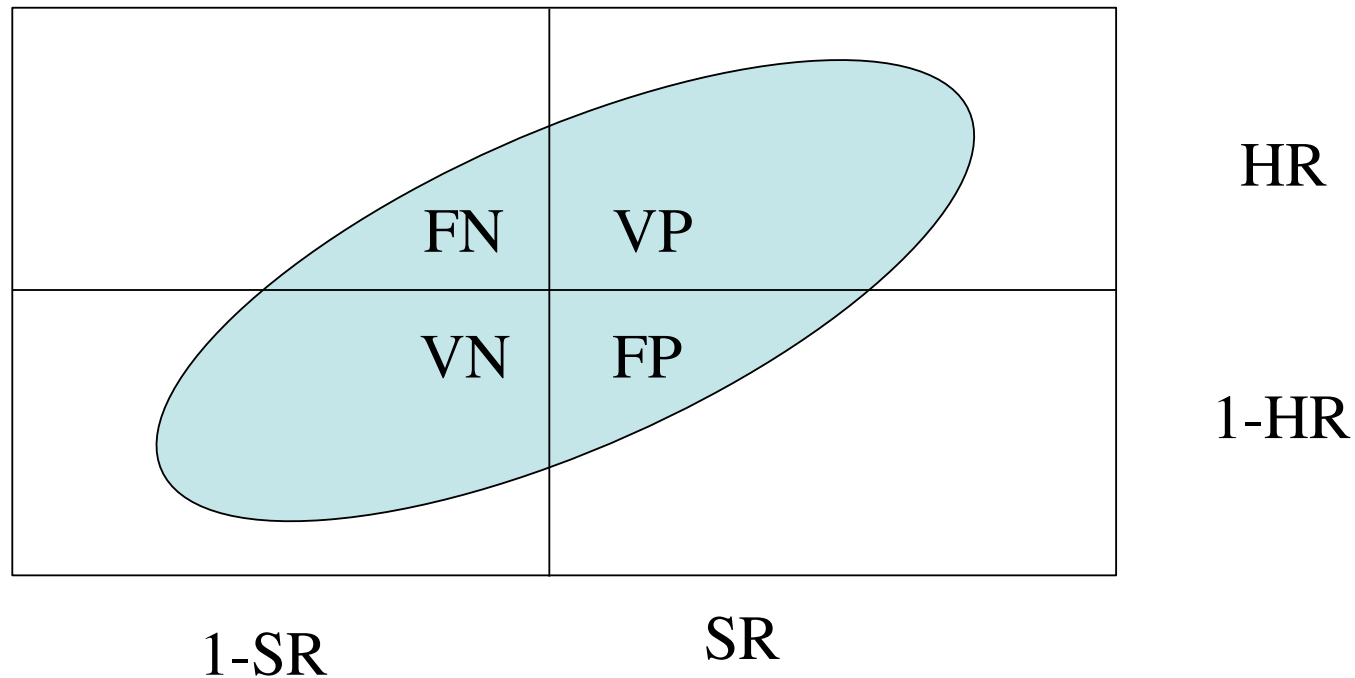
Kelly and Fiske (1950)

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

Predictive and Concurrent Validity and Decision Making

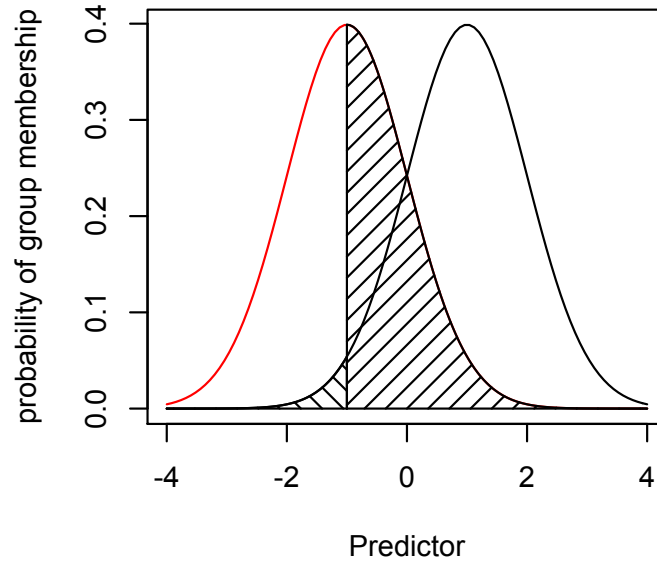
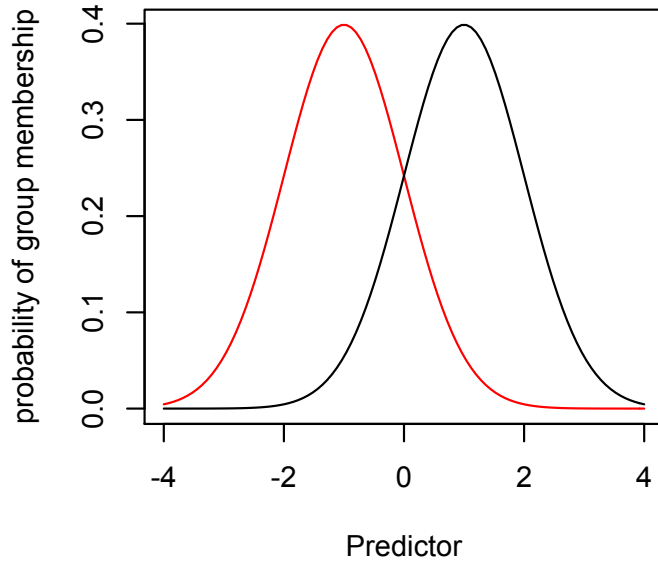
Hit Rate = Valid Positive + False Negative

Selection Ratio = Valid Positive + False Positive

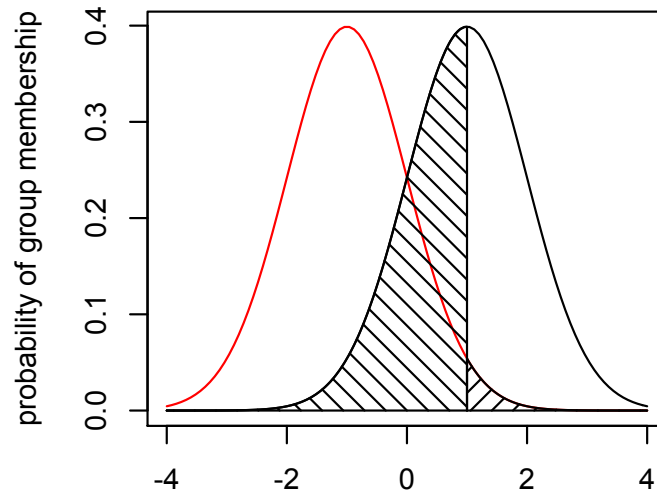
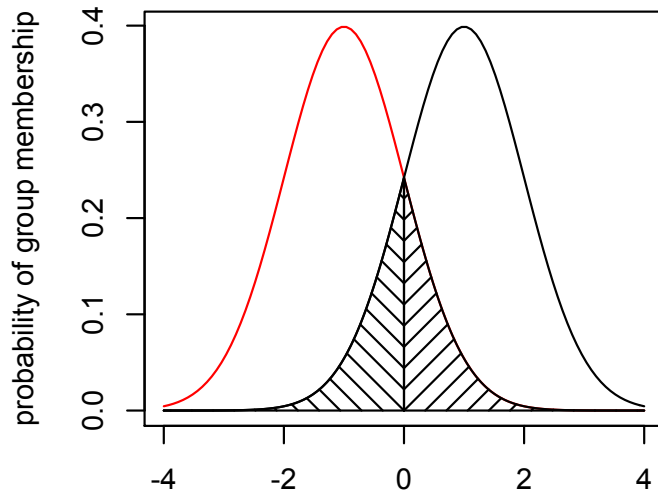


$$\text{Phi} = \frac{\text{VP} - \text{HR} * \text{SR}}{\sqrt{\text{HR} * (1 - \text{HR}) * (\text{SR}) * (1 - \text{SR})}}$$

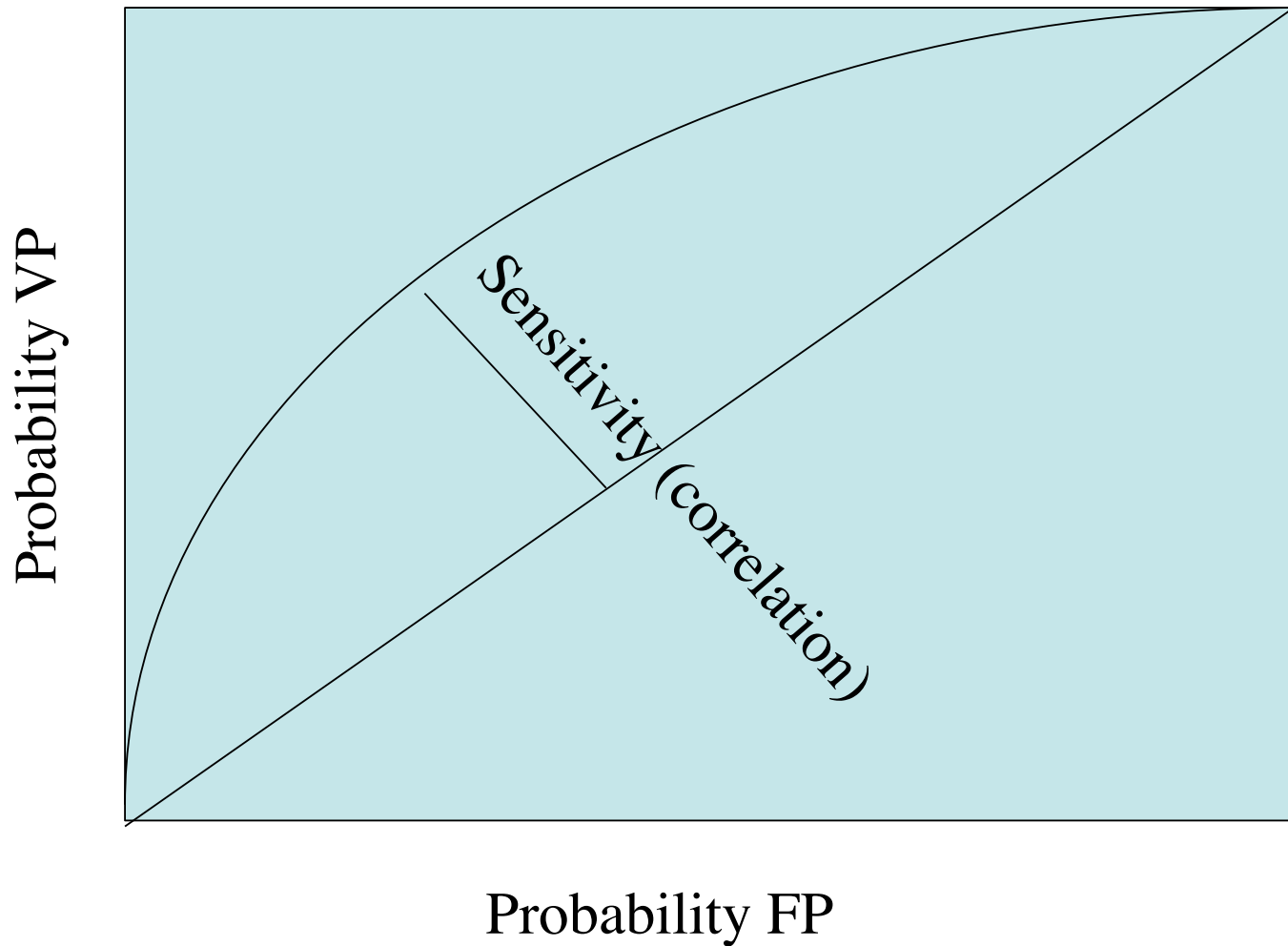
Validity as decision making



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	I-HR
Decision	I-SR	Selection Ratio	

Predictive Validity, Utility and Decision Theory

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

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

Decisions for institutions, advice for individuals

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

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

Decision making and the benefit of extreme selection ratios

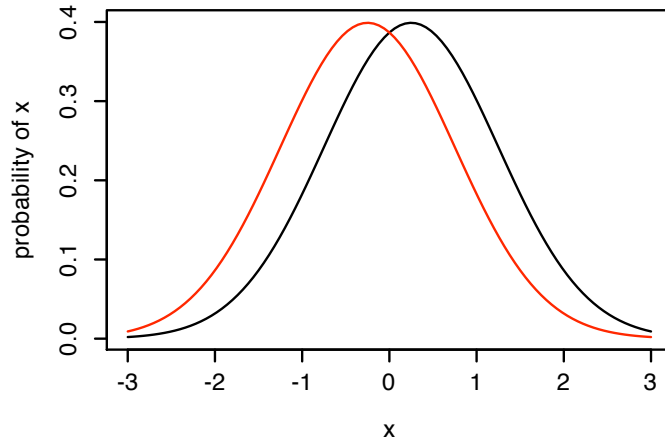
- Typical traits are approximated by a normal distribution.
- Small differences in means or variances can lead to large differences in relative odds at the tails
- Accuracy of decision/prediction is higher for extreme values.
- Do we infer trait mean differences from observing differences of extreme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

Odds ratios as f(mean difference, extremity)

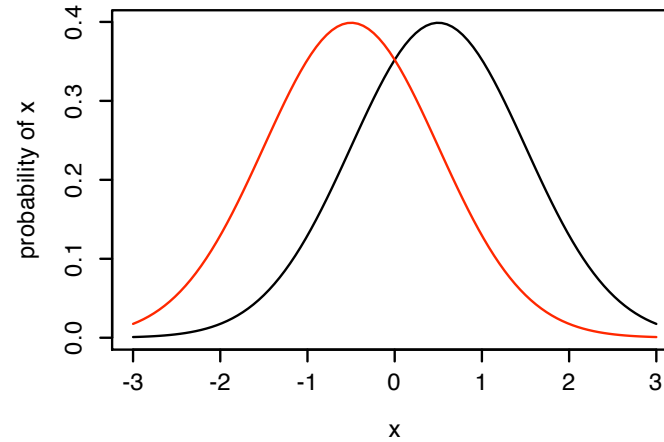
Difference = .5 sigma

Difference = 1.0 sigma

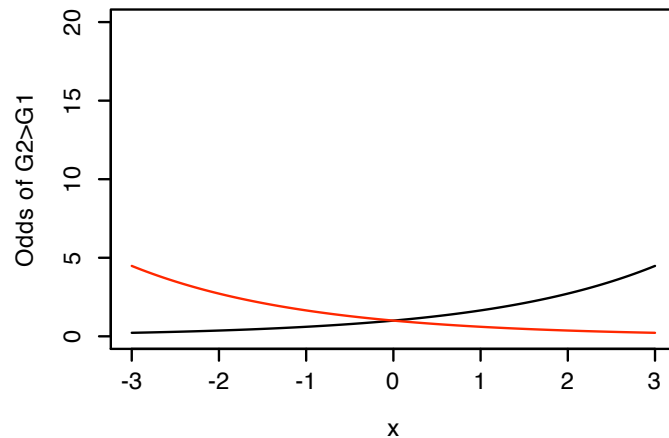
Normal density for two groups



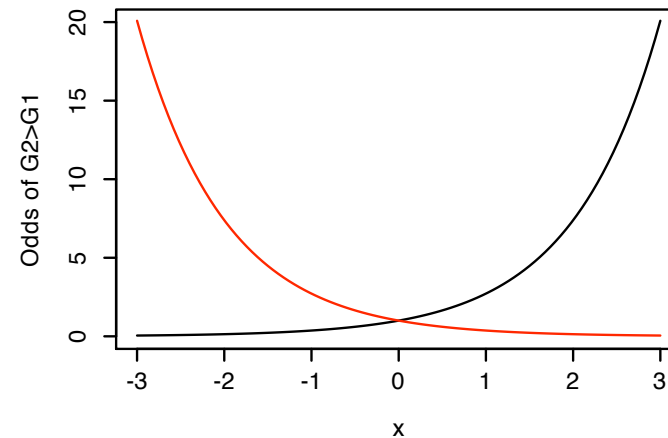
Normal density for two groups



Odds ratio of G1 vs G2



Odds ratio of G1 vs G2

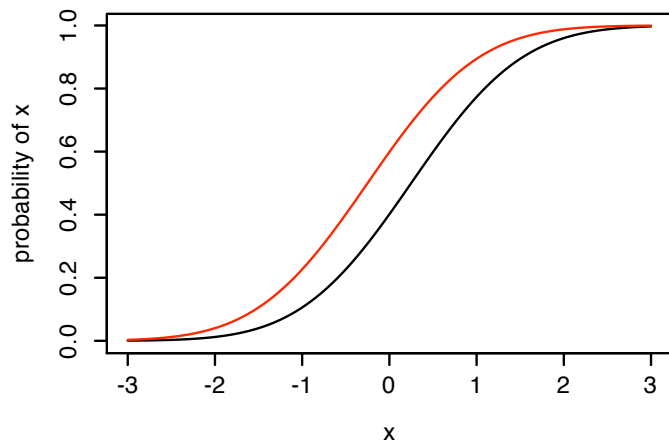


The effect of group differences on likelihood of extreme scores

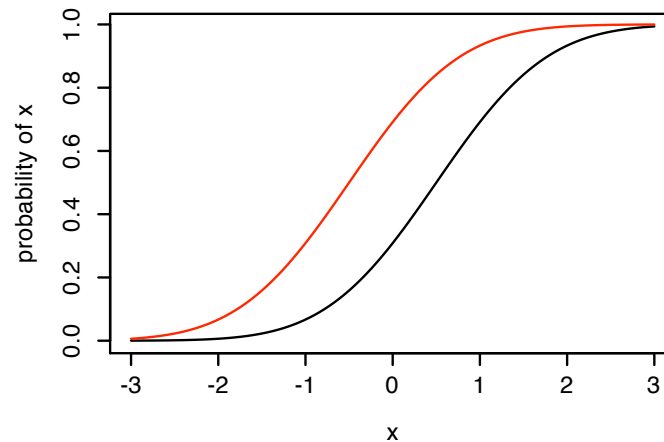
Difference = .5 sigma

Difference = 1.0 sigma

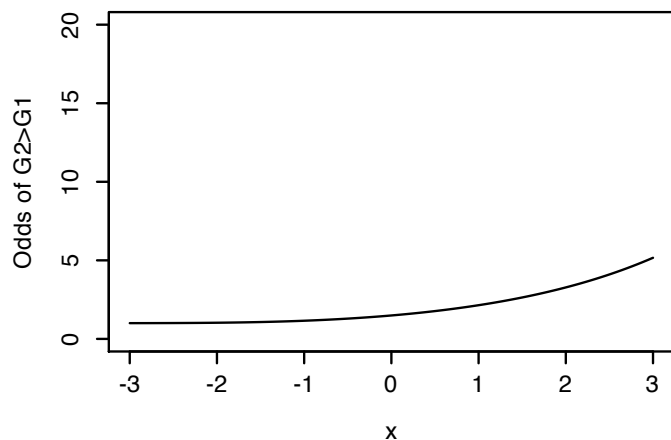
Cumulative normal density for two groups



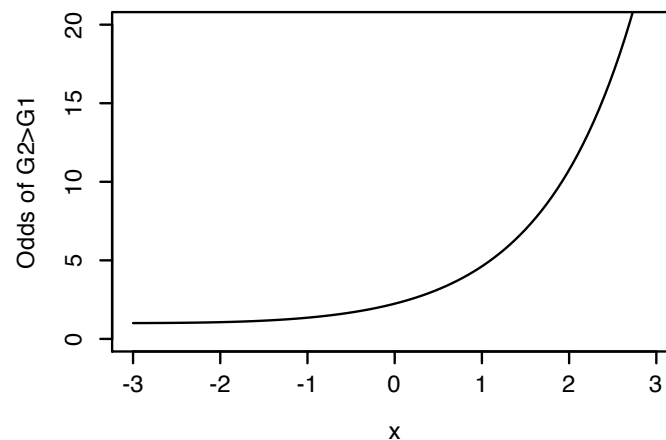
Cumulative normal density for two groups



Odds ratio that person in Group exceeds x

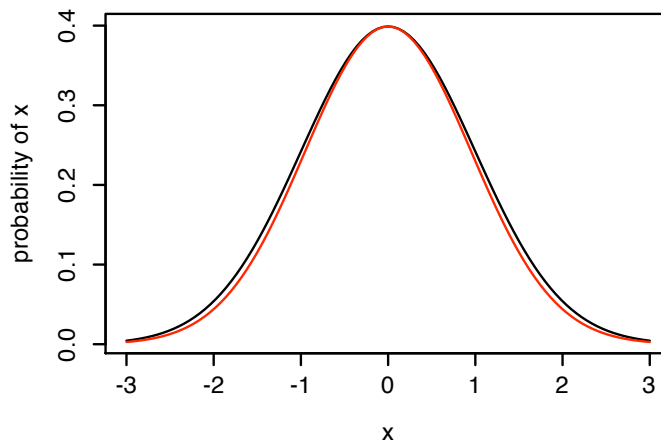


Odds ratio that person in Group exceeds x

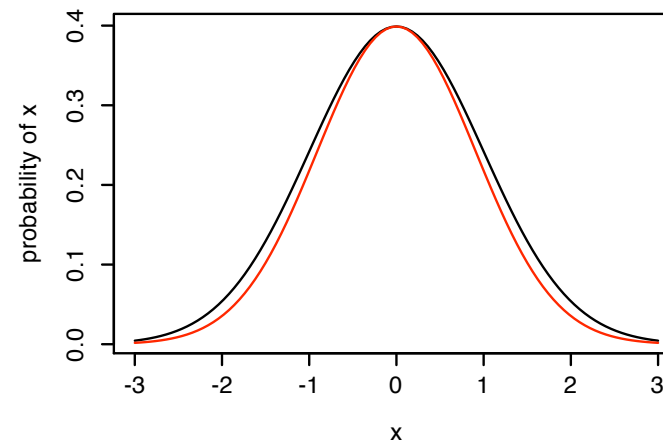


The effect of differences of variance on odds ratios at the tails

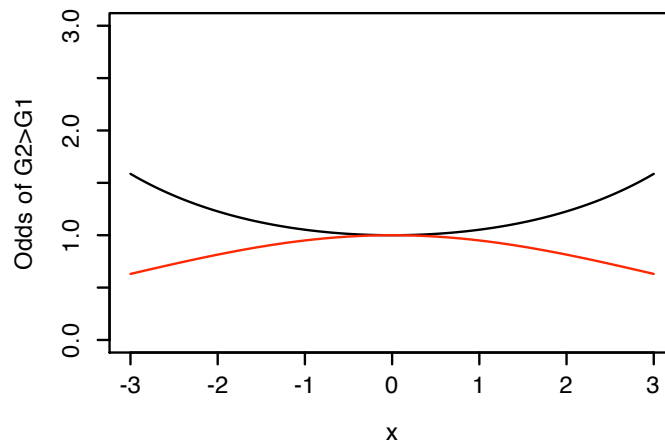
variance of two groups differ by 10%



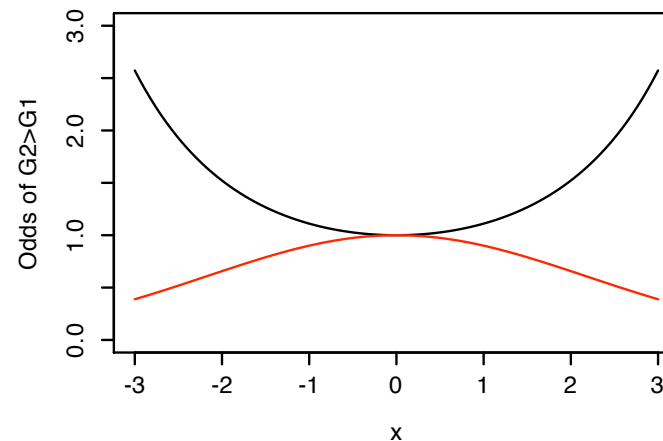
Variance of two groups differs by 20%



Odds ratio of G1 vs G2



Odds ratio of G1 vs G2

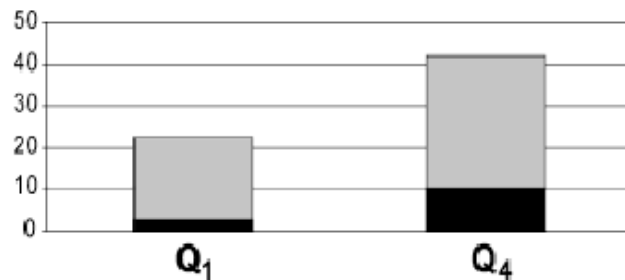


Restriction of range

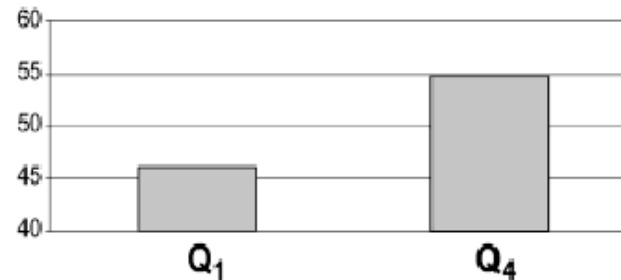
- Validity of SAT is partially limited by range restriction. (see Lubinski and Benbow)
- Consider giving SATs to 12-13 year olds
 - SAT M \geq 390 or SAV V \geq 370 (top 1 in 100)
 - SAT M \geq 500 or SAV V \geq 430 (top 1 in 200)
 - SAT M \geq 700 or SAV M \geq 430 (top 1 in 10,000)
 -

Predictions within top student group

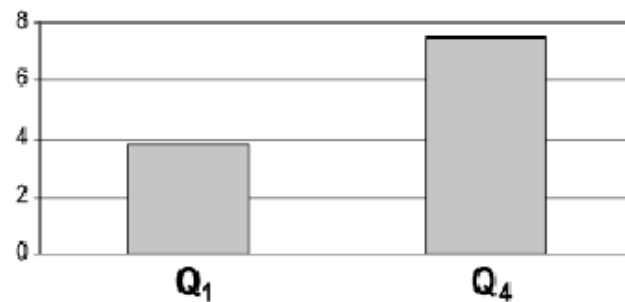
Percent Earning a Doctorate and STEM Doctorate



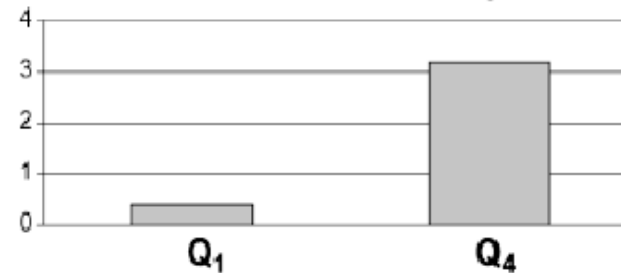
Percent Earning Income Greater Than or Equal To Median Within Sex



Percent Earning Patents

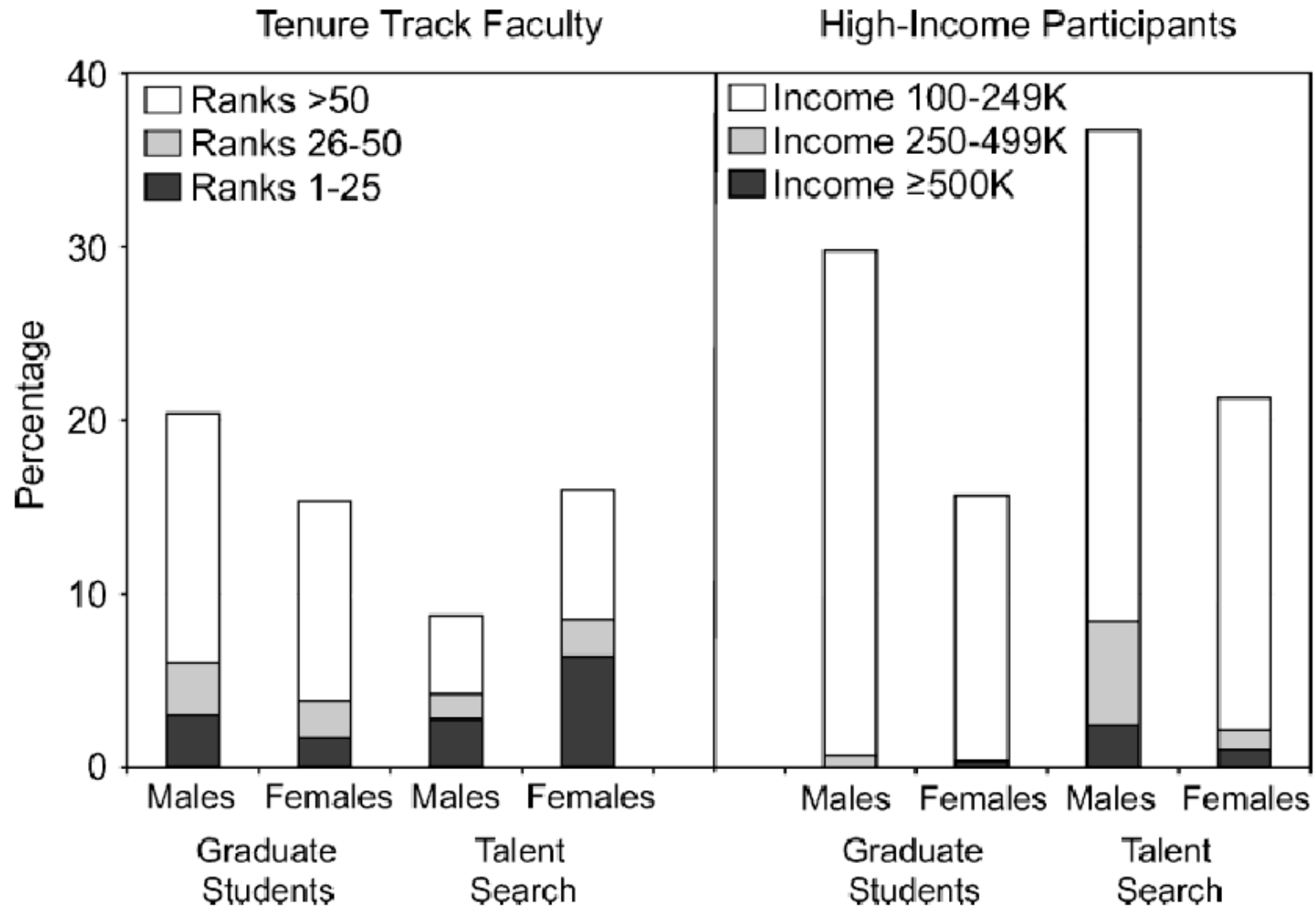


Percent Earning Tenure at a Top 50 U.S. University

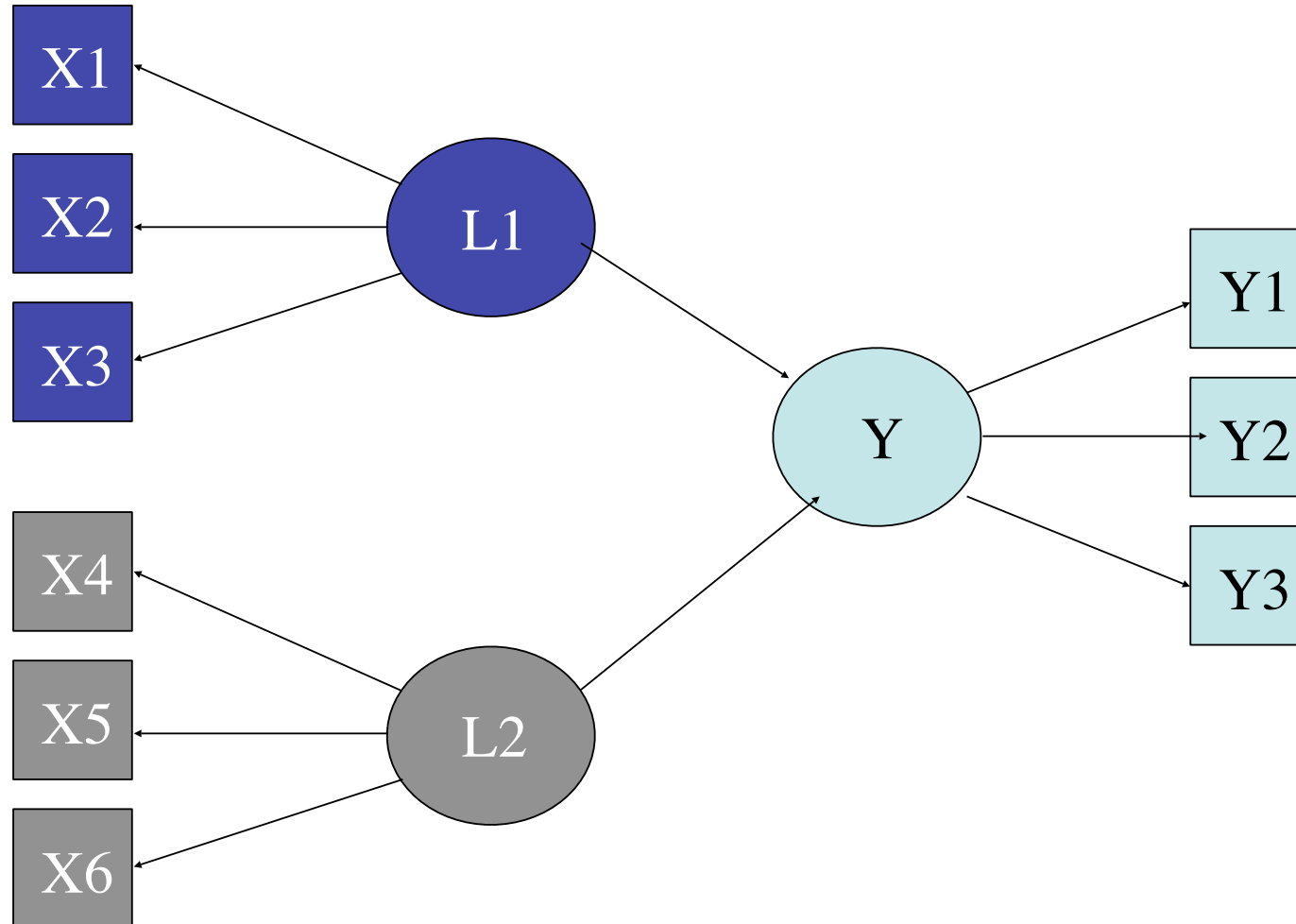


Validity continues even among top 1%

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

