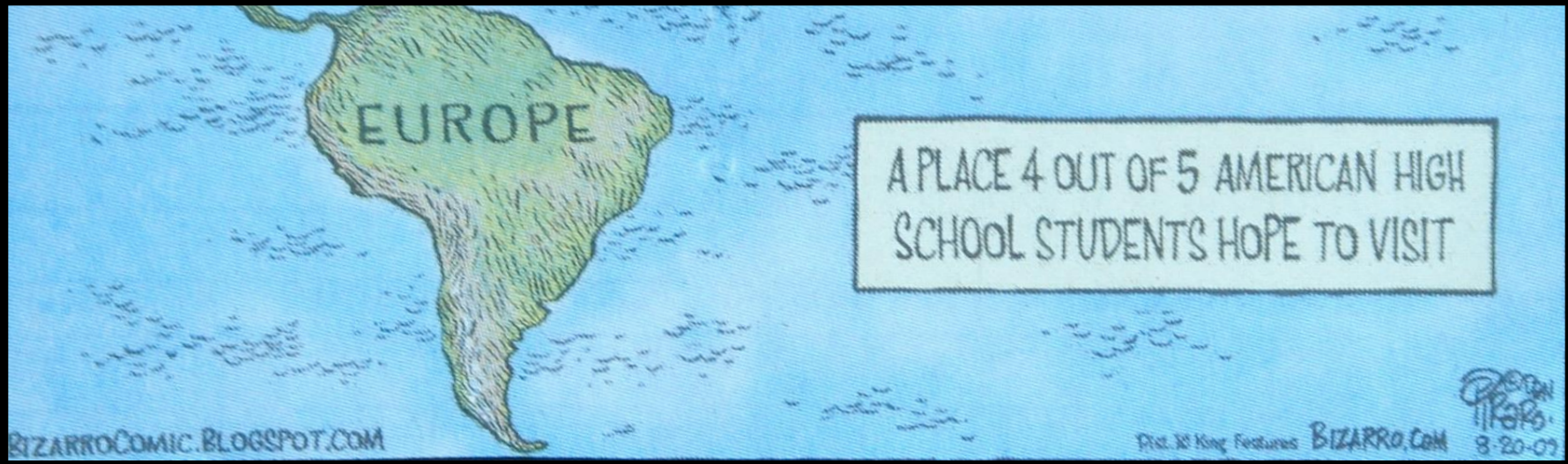
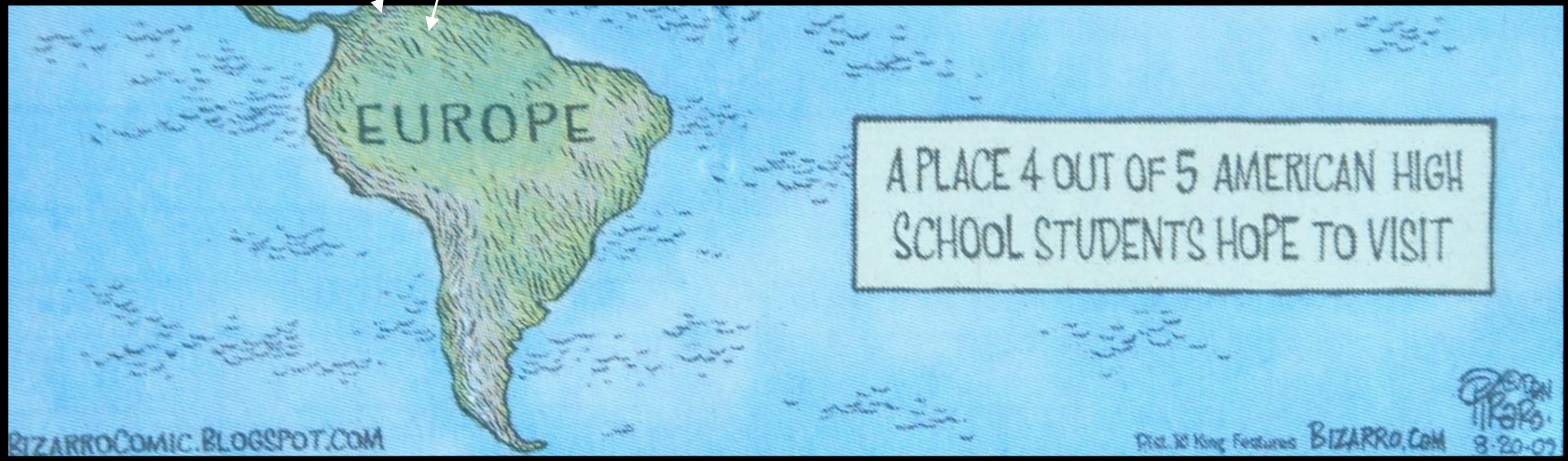
Fairness of Selection: Paradoxes and Dilemmas

Wim J. van der Linden
CTB/McGraw-Hill


EUROPE
A PLACE 4 OUT OF 5 AMERICAN HIGH SCHOOL STUDENTS HOPE TO VISIT
BIZARROCOMIC.BLOGSPOT.COM
Dist. by King Features BIZARRO.COM
8-20-09

Norbotten

Västerbotten



Outline

- Analysis of fairness
- Any solutions?

How to Analyze Fairness?

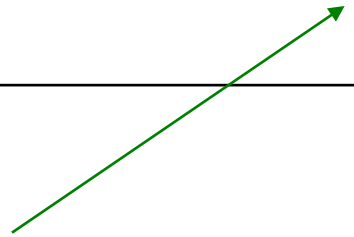
- Alternative approaches to the analysis of fairness of selection
  - analysis of probability of admission
  - regression analysis
- First dilemma: should criterion be based purely on the event of admission or predicted success during the study?

Berkeley Study

- Bickel, P.J, Hammel, E.A., O'Connel, J.W., (1975). Sex bias in graduate admission: Data from Berkeley. *Science*, *7*, 398-404.
- Data
  - 12,763 applicants (8,442 males; 4,321 females)
  - 101 different PhD programs
- 2x2-table with counts of admissions showed substantial sex bias (χ^2=110.8, *df*=1, *p*=0)

Berkeley Study *Cont'd*

	Admitted	Rejected
Males	3.738	4.704
	(3.460,7)	(4.981,3)
Females	1.494	2.827
	(1.771,3)	(2.549,7)



Observed-Expected
equal to -277.3

Berkeley Study *Cont'd*

- Repetition of same analysis for 101 separate PhD programs showed 10 programs with significant differences
  - 4 programs discriminated against females (26 fewer females admitted than expected)
  - 6 programs discriminated against males (64 more females admitted than expected)
- What has happened to the 277,3 + 26 - 64 = 239.3 missing females?

High School Exit Exam Study

- Study of developments in scores for German and French as a foreign language on Dutch high school exit exam between 1985-1990 (de Jong)
  - All scores equated (adjusted for differences in difficulty, etc.)
  - Scores in each track (Mavo/Havo/VWO) decreased systematically

High School Exit Exam Study

Cont'd

- Study of developments in scores
  - Score distribution for whole population remained constant
- What has happened to the students with the decreasing scores?

Basketball Study

- Belief in *streak shooting: hits* have a higher probability of being followed by another *hit* than by a *miss*
- Several explanations, including the idea of a "hot hand"

Basketball Study *Cont'd*

- Kahneman-Tversky study of the perception of runs: tendency to perceive longer runs no longer as random
- Controlled experiments
  - Runs in coin tosses
  - Basketball throwing (Gilovich & Tversky)

Basketball Study *Cont'd*

- Wardrop, R. L. (1995). Simpson's paradox and the hot hand in basketball. *The American Statistician*, *49*, 24-28.
  - Re-analysis of the Gilovich-Tversky data
  - Data for the individual players do not show any streak shooting
  - Combined data for all players do show streak shooting

Medical Studies

- Success rates for kidney-stone procedures
  - All open procedures (A)
  - Percutaneous nephrolithotomy (B)

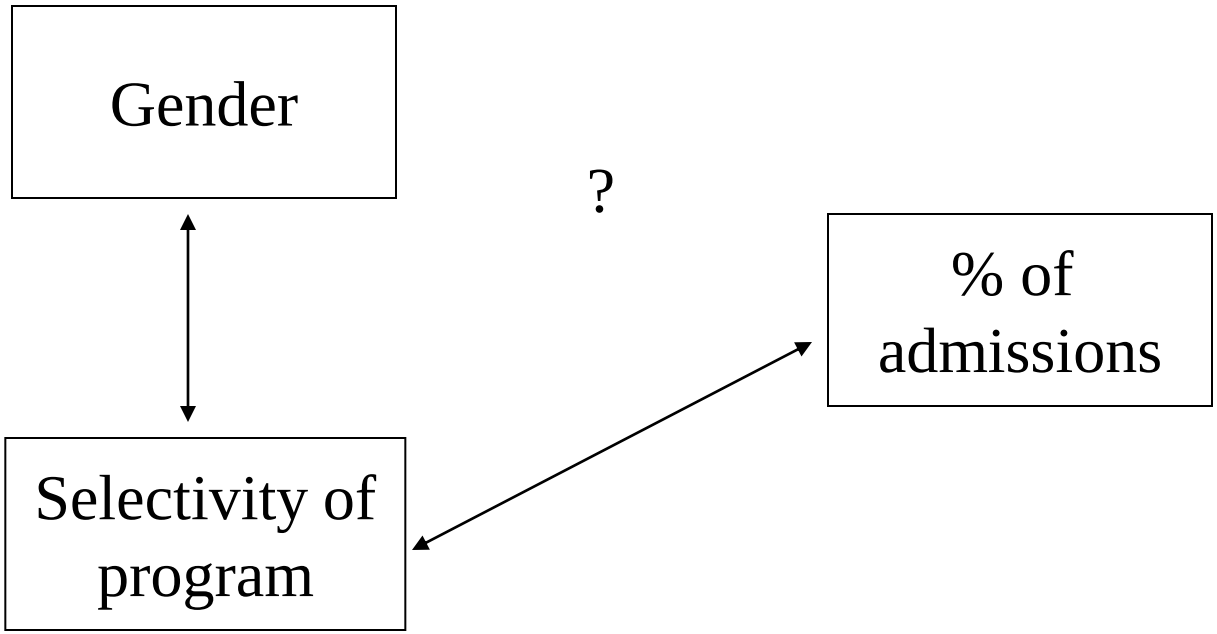
	A	B
Small Stones	**93**%	87%
Large Stones	**73**%	69%
Combined	78%	**83**%

Simpson's Paradox

- The previous phenomena arise when covariates are present that correlate both with the independent and dependent variables
- Unawareness of such covariates leads to misleading results; acknowledgment leads to "paradoxes''
- Simpson's paradox occurs in correlational studies it is not a causal paradox

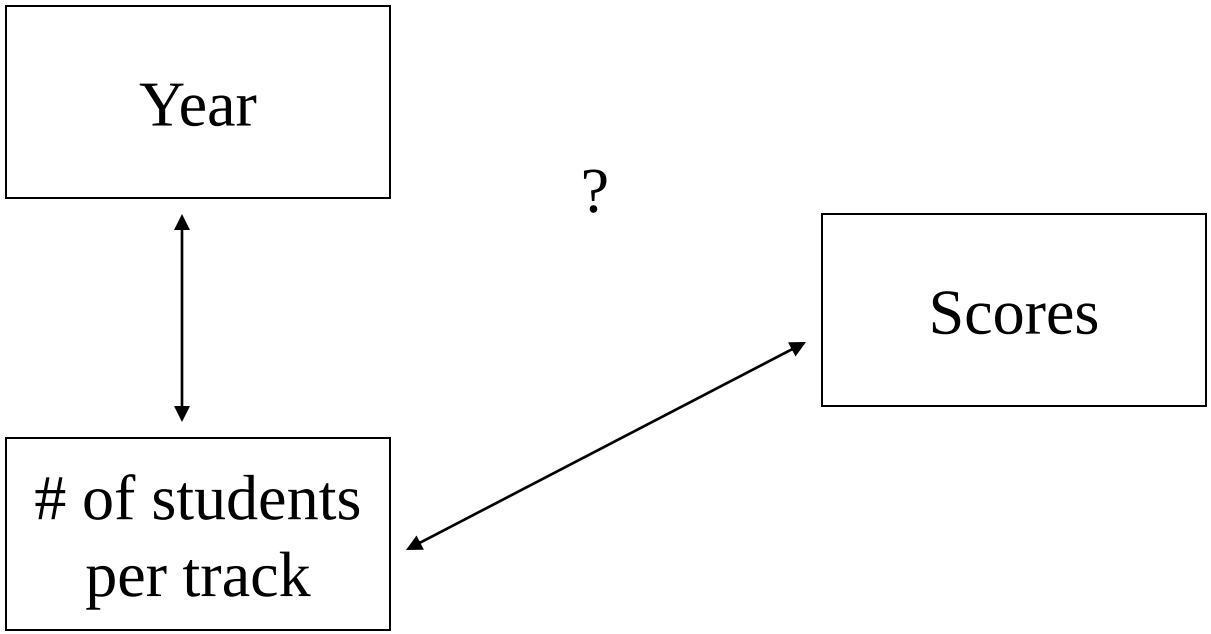
Simpson's Paradox *Cont'd*

- Berkeley study



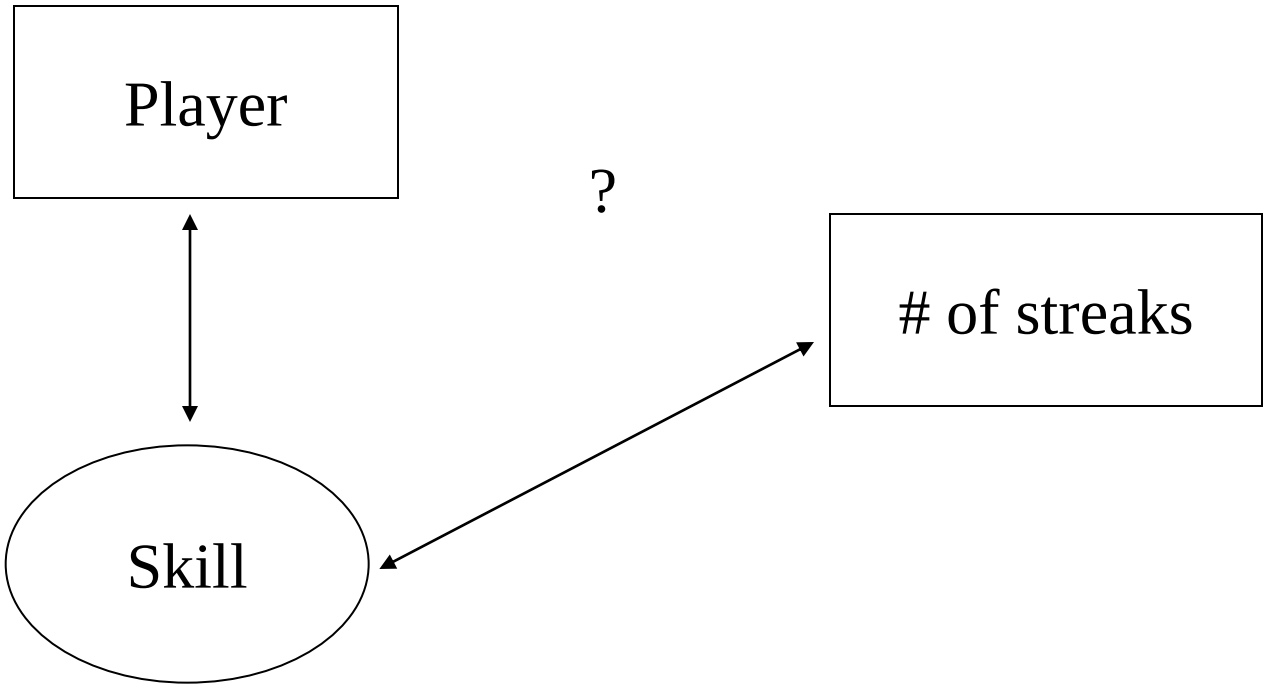
Simpson's Paradox *Cont'd*

- High school exit exam



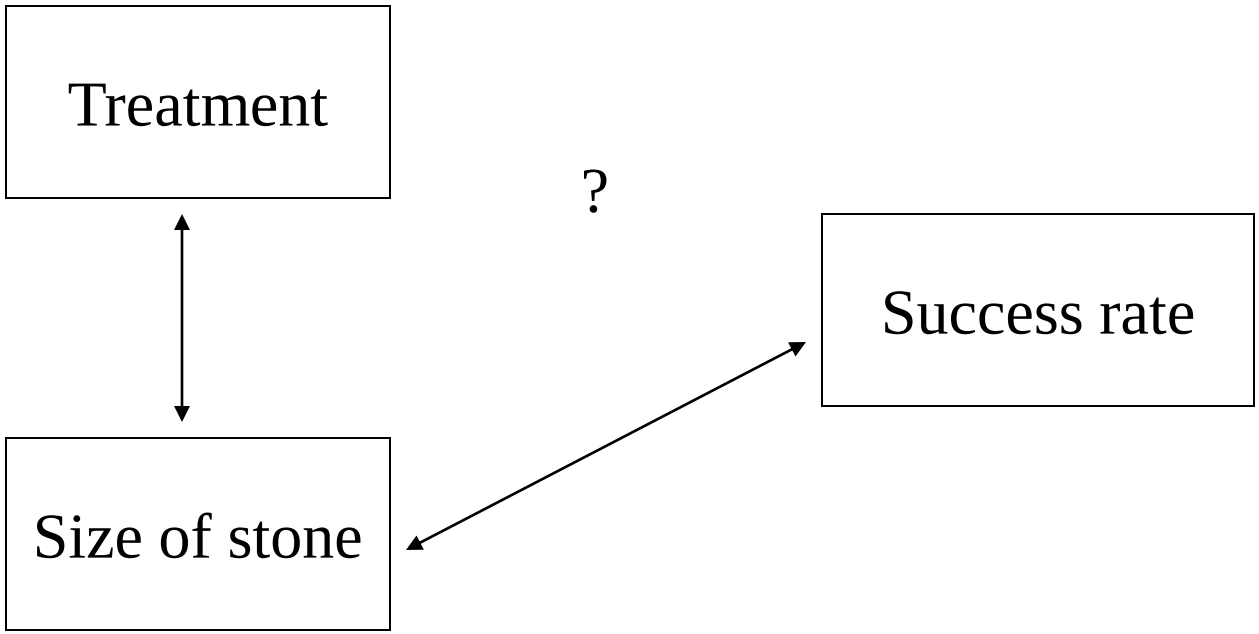
Simpson's Paradox *Cont'd*

- Hot-hand in basketball



Simpson's Paradox *Cont'd*

- Kidney stones



Simpson's Paradox *Cont'd*

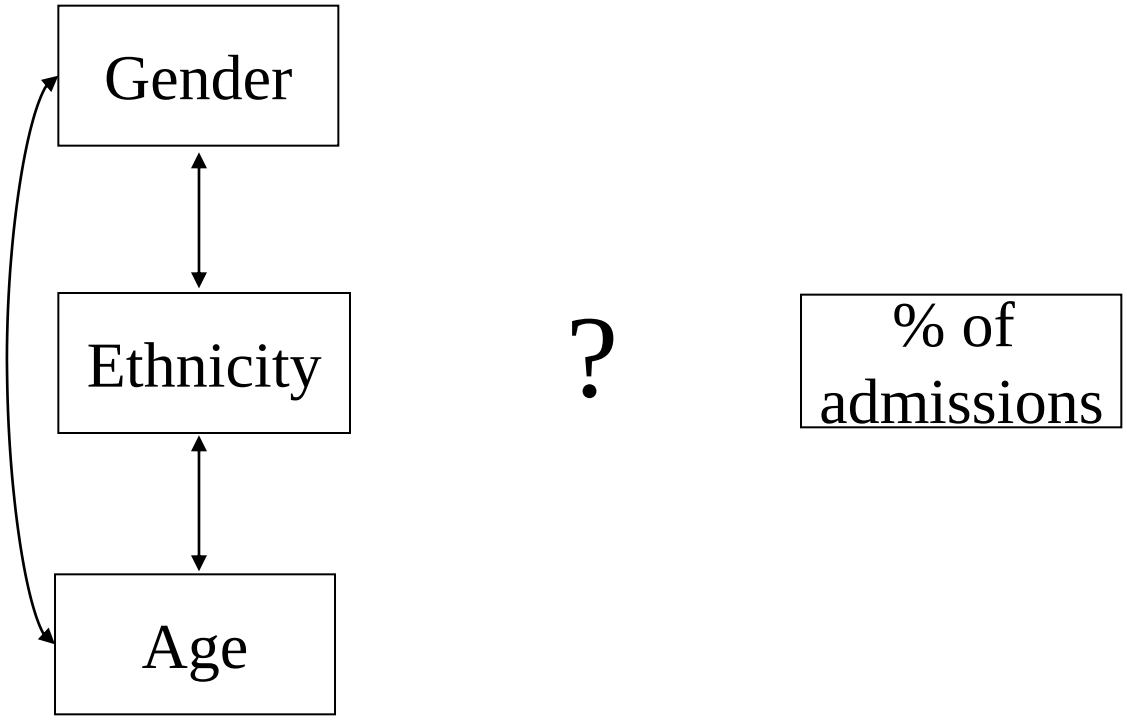
- Simpson's paradox arise easily in studies without experimental control
  - Observation studies
  - Surveys
  - Case studies
- Studies on fairness of selection typically belong to this category.

Simpson's Paradox *Cont'd*

- A even more embarrassing version of the paradox can arise in studies of fairness with multiple groups, e.g., gender, ethnicity, and age
- Depending on the correlational structure between these groups, we may then find bias against a group in the total data set, but none, or even the opposite, in some (combination) of its subgroups

Simpson's Paradox *Cont'd*

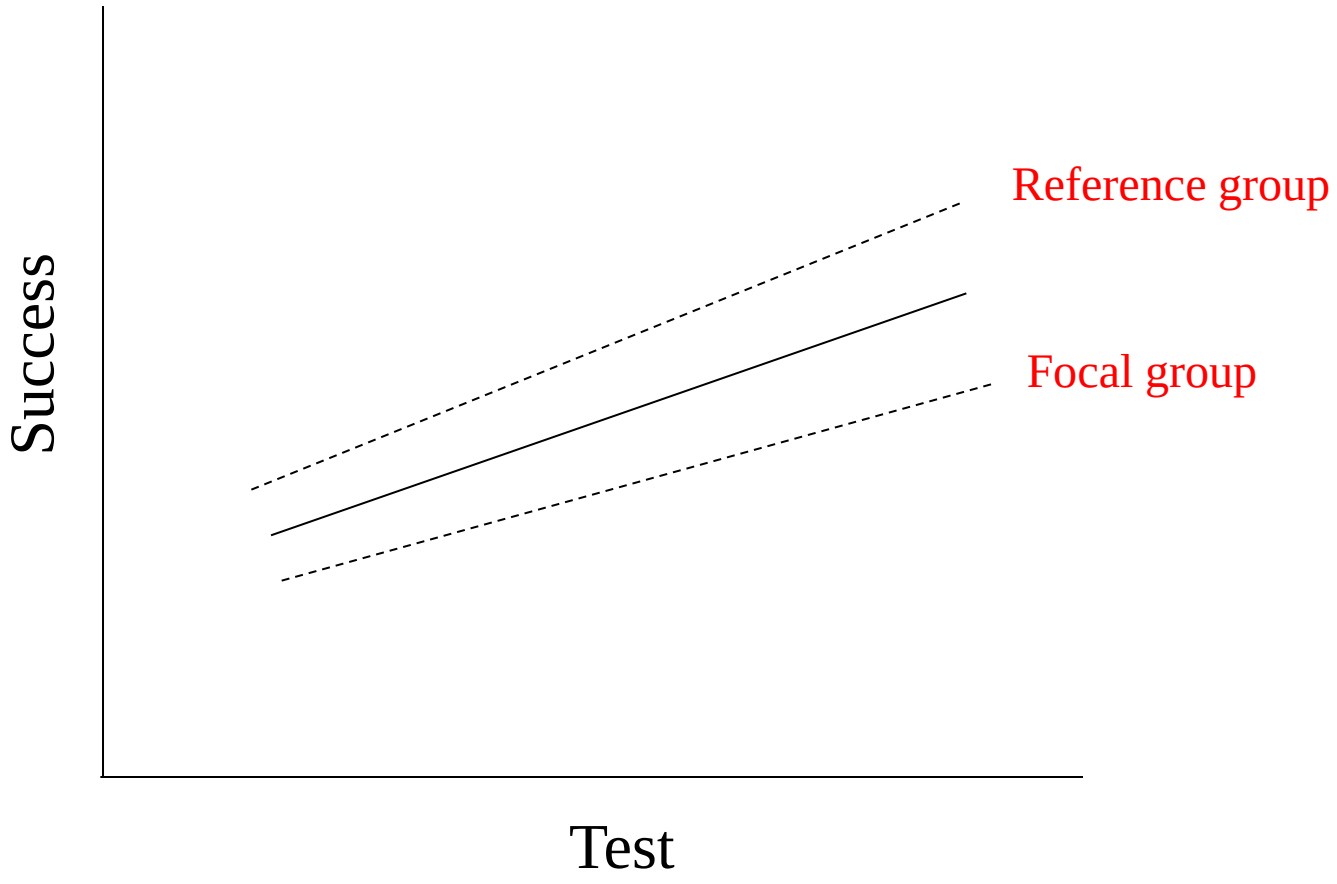
- Multiple groups



Regression Analysis

- Evaluation of admission test as a predictor of success in the educational program
- An admission test is considered to be fair when it shows identical regression lines for the success on the test scores in the subgroups
- No distinction between bias due to
  - tests scoring (measurement bias)
  - use of inadequate predictors (prediction bias)

Regression Analysis *Cont'd*



Regression Analysis *Cont'd*

- Regression example from Neter & Wasserman, 1974, sect. 4.2
  - Residual plots for regression of productivity on the age of workers using machines produced by two different companies (A and B)
  - Are the age measures biased? Or should we adopt another explanation?

Regression Analysis *Cont'd*

- The typical reaction of the statistician is not to accuse the age measure of bias but look for important *missing variables* in the regression equation
- If the missing variable (here: level of experience or training) is entered in the equation, the residual plots for machine A and B become identical

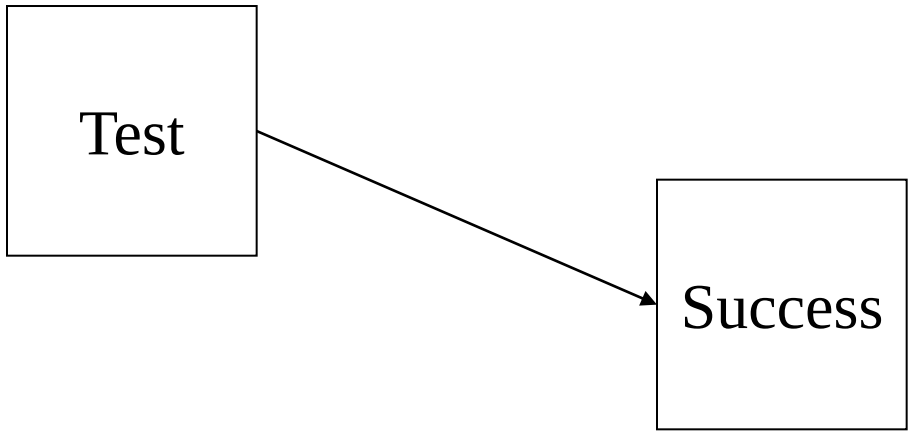
Regression Analysis *Cont'd*

- Principle of complete modeling: If for a population of candidates all relevant predictors are in the regression equation, then the lines for any two subpopulations are identical.
  Hence,
  - if the model is complete, no "prediction bias"
  - if there is "prediction bias", the model is not complete

Regression Analysis *Cont'd*

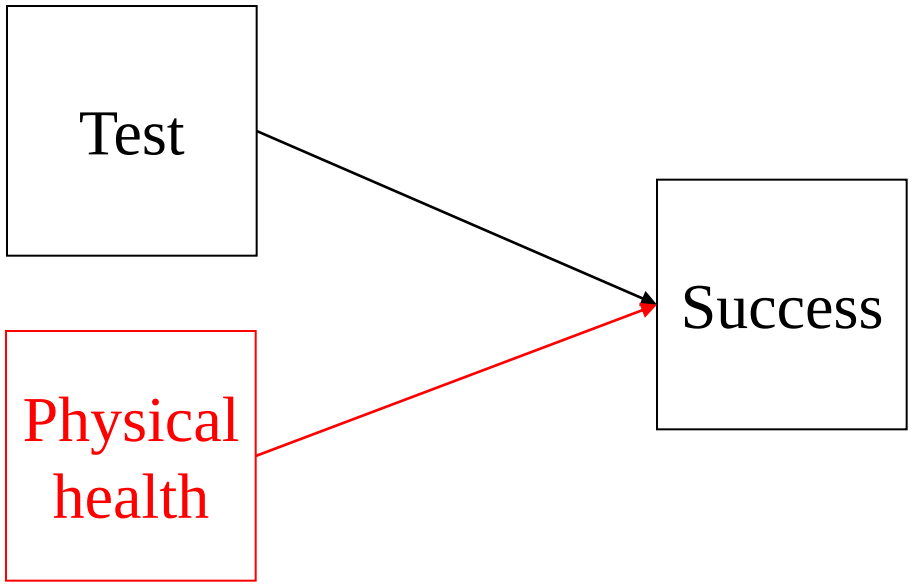
- Examples of likely missing predictors in selection to higher education: other abilities, physical health, personal income, age, gender, psychological stability, religious background, etc.
- Ethical dilemma: fairness of selection vs. acceptance of undesirable predictors

Complete Prediction Model



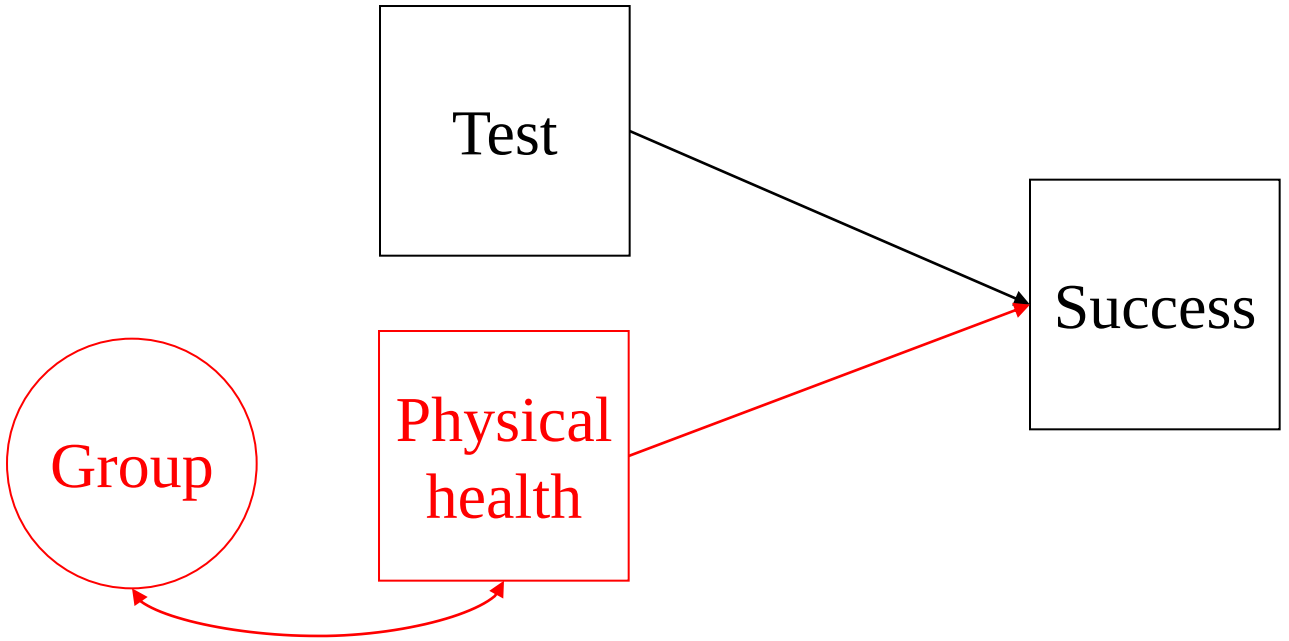
Except for random error, test scores predict all differences in success between candidates

Incomplete Prediction Model



Individual prediction bias: success of candidates with better health is underestimated

Incomplete Prediction Model with Groups Variable



If a group variable correlates with health, success of candidates in underestimated in one group and overestimated in the other

Solutions?

- Controlled experiments
  - Role players as applicants
  - Manipulation of personal files
- Transition from selection testing to mastery testing
  - Two different educational systems
  - Stimulation of critical program choices, etc.

Two Different Educational Systems

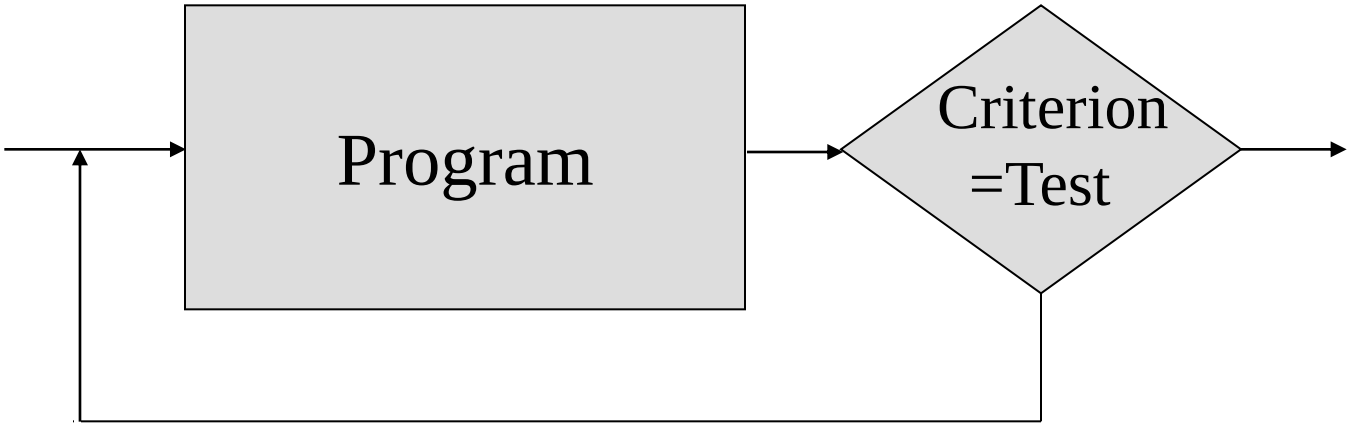
- Output control
- National standards
- National exit exams
- High-school diplomas give admission rights to higher education
- Diploma includes certification

- Input control
- Local standards
- National admission tests
- High-school diplomas without admission rights to higher education
- Separate certification

Two Different Educational Systems *Cont'd*

- Mastery testing vs. selection testing
- Achievement testing vs. aptitude testing
- Content validity vs. predictive validity
- Weak vs. fierce debates on item and test bias

Mastery Decision



Selection Decision