HUDK 4050:CORE METHODS IN EDM

11/15/16 10:53 AM

In the news

AMERICAN SENTINEL UNIVERSITY

Lexia®
A Rosetta Stone® Company

Readers' Choice Awards 2016

Our tech-savvy readers name their favorite technologies in dozens of categories, from flipped learning software to tablets and convertible laptops.

DATA
Why It's So Hard for Us to Visualize Uncertainty
Harvard Business Review

WNYC

City's Special Ed Compliance Still Hobbled by Bad Data

Early Learning and Educational Technology Policy Brief

EdSurge

Which Edtech Companies Are Producing the Best Research-Based Products?

U.S. Department of Education

How Can We "Leapfrog" Educational Outcomes?
Stanford SOCIAL INNOVATION REVIEW

LinkedIn Data Shows More Cash-Strapped Millennials Turning To Part-Time Freelancing
FC

Top Hat Hires First Chief Marketing Officer, Nick Stein

BusinessWire
A Berkshire Hathaway Company

http://bit.ly/2fS6tLA

HealthGrove
by GRAPHIQ

CDC

Researchers Use CDC Data to Rank States' Sexual Education Programs

THE JOURNAL

What Will a Trump Presidency Mean for K–12 and Ed Tech?

Meetup

Big Data, NYC
New York Machine learning and Data Meetup

CHAR(16)
CLUSTERING • HIGH AVAILABILITY • REPLICATION

bigml®

INSIGHT

Who Will Be Donald Trump's Secretary Of Education? A Familiar Name Is A Likely Pick For The Cabinet Position

Insight Artificial Intelligence Fellowship

Twitter

Podcast

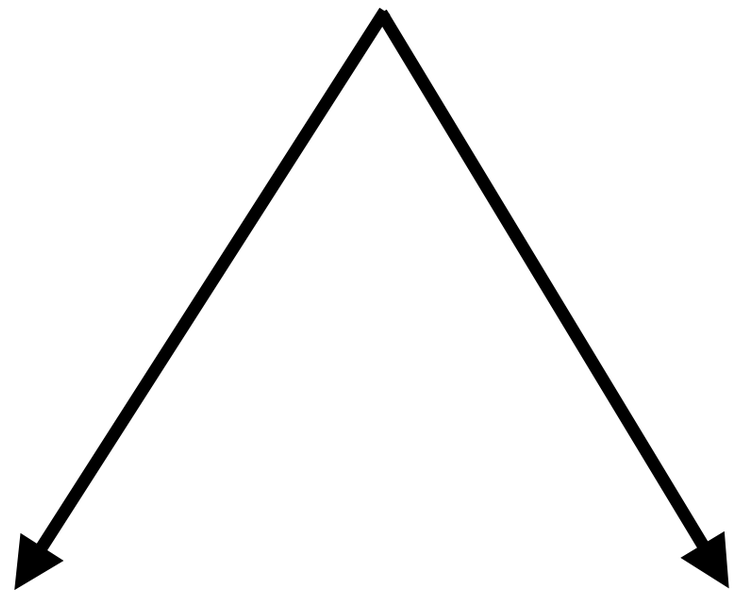
Inference

11/15/16 10:53 AM

“Statistics is the study of uncertainty”

– LJ Savage, 1977

Variation



Patterns

Uncertainty

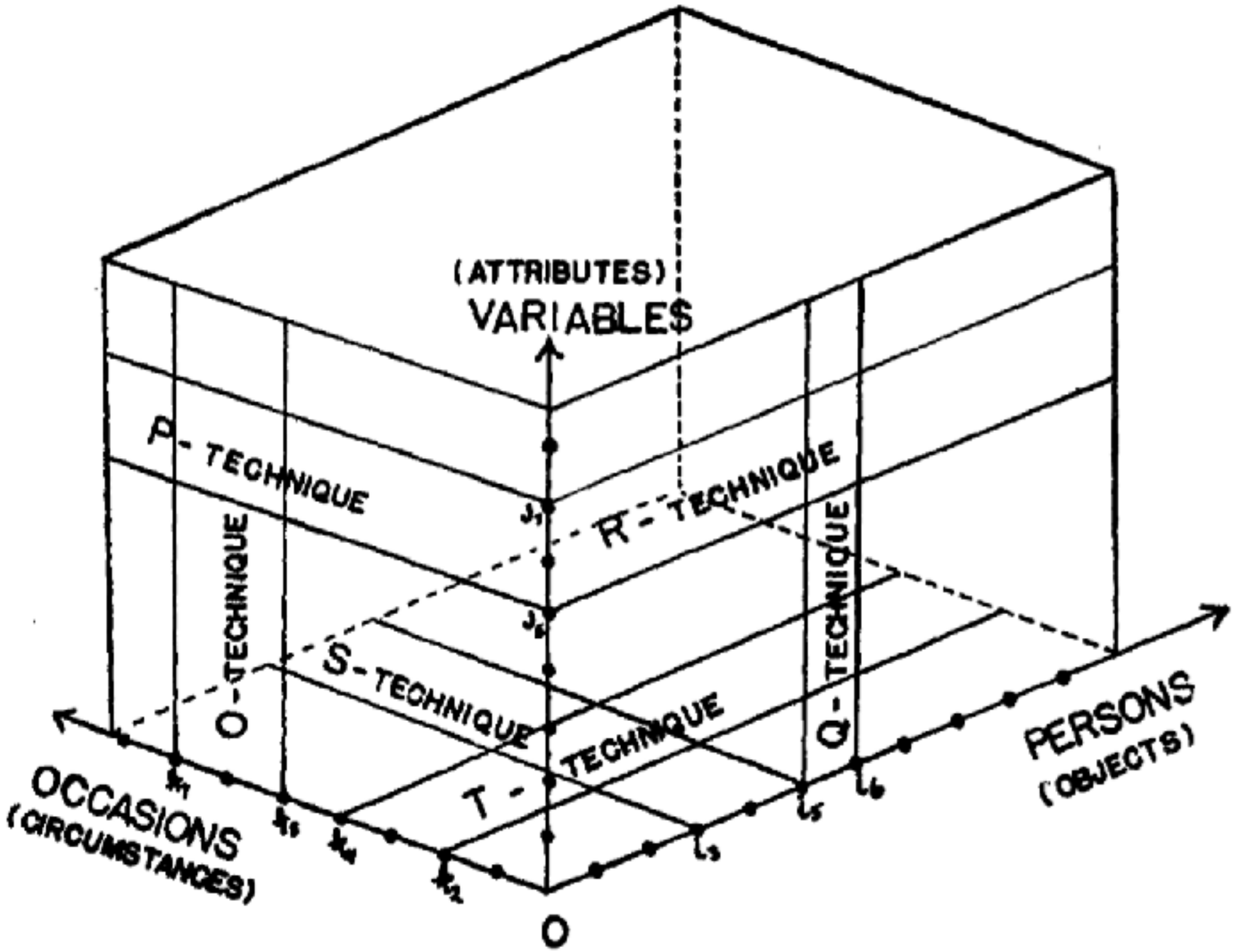


FIG. 1. THE COVARIATION CHART

Cattell, 1952

Swimming

Experimental Psych

Context

Classroom

Ed Stat, Economics

Bad

Spell “cat”

Social Psych

Good

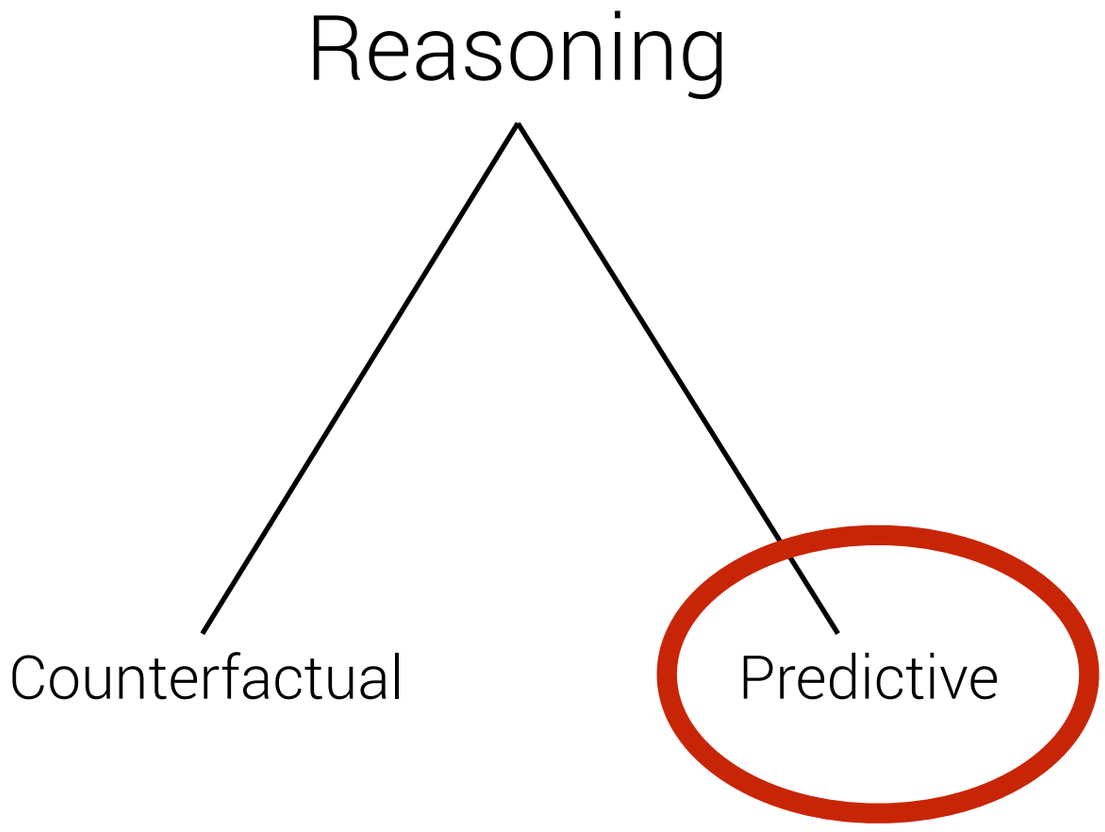
Behavior

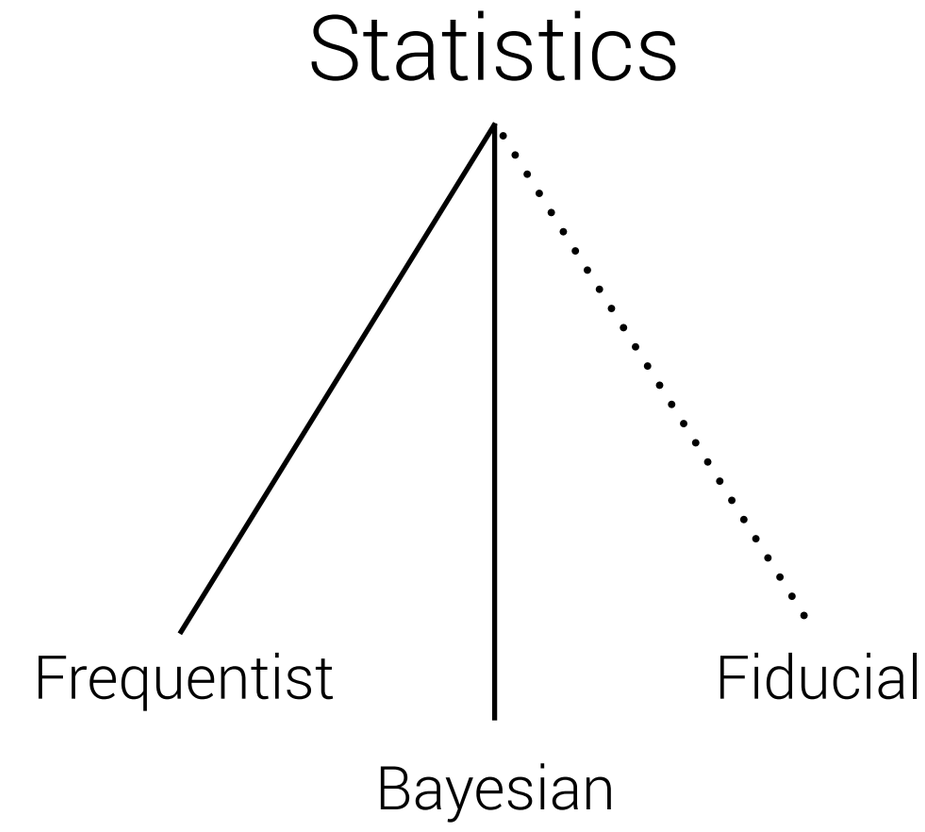
1 2 3 4 5 6 7 8 …

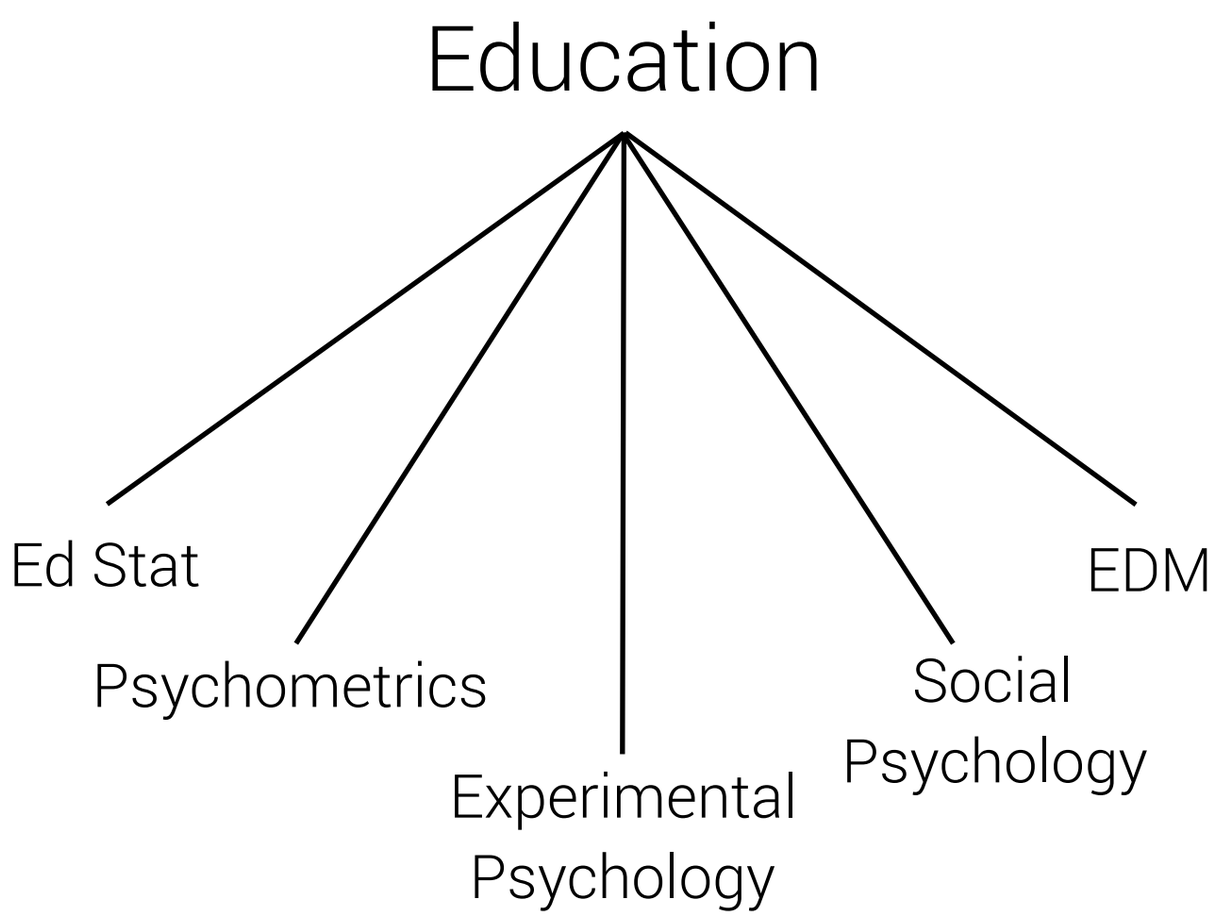
1 2 3 4 5 6 7 8 9 …

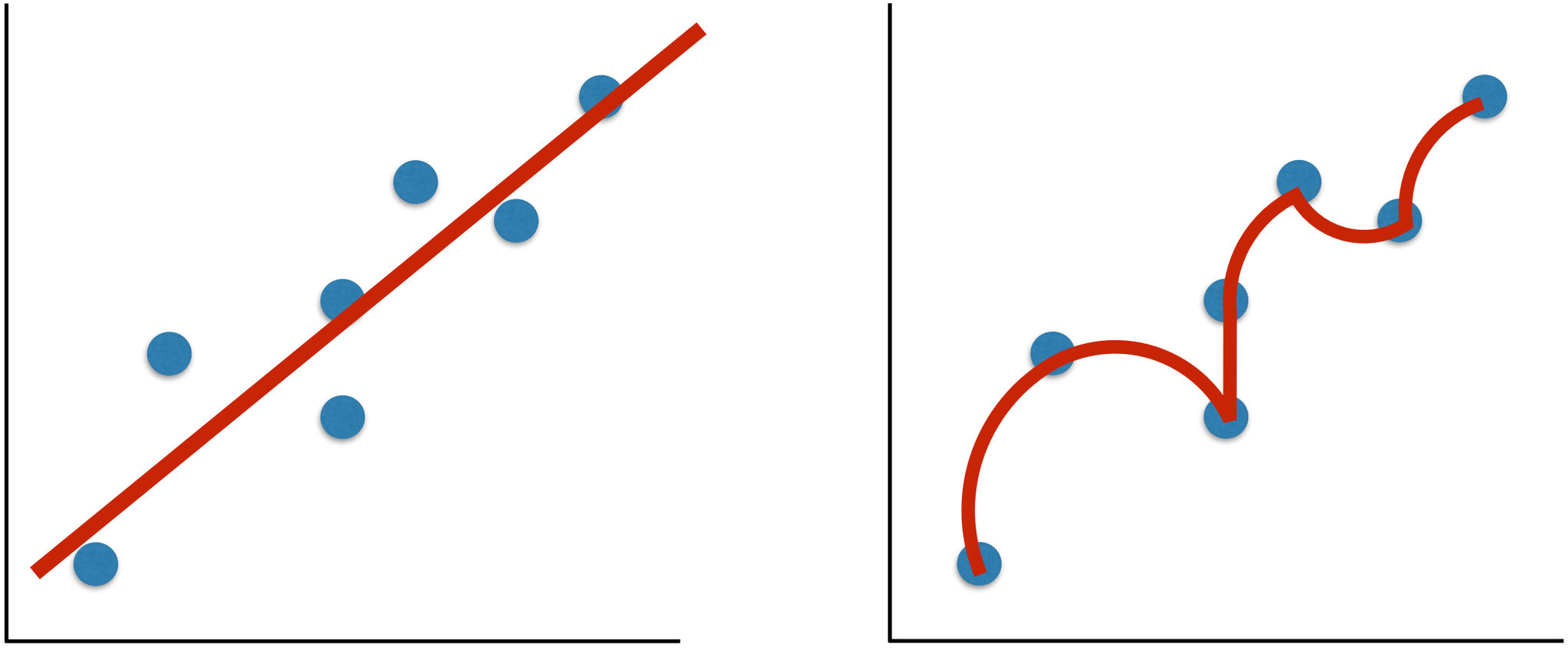
Psychometrics/ BKT

Adaptive


Reasoning
Counterfactual
Predictive


Statistics
Frequentist
Bayesian
Fiducial


Education
Ed Stat
Psychometrics
Experimental Psychology
Social Psychology
EDM



Which is more “accurate”?

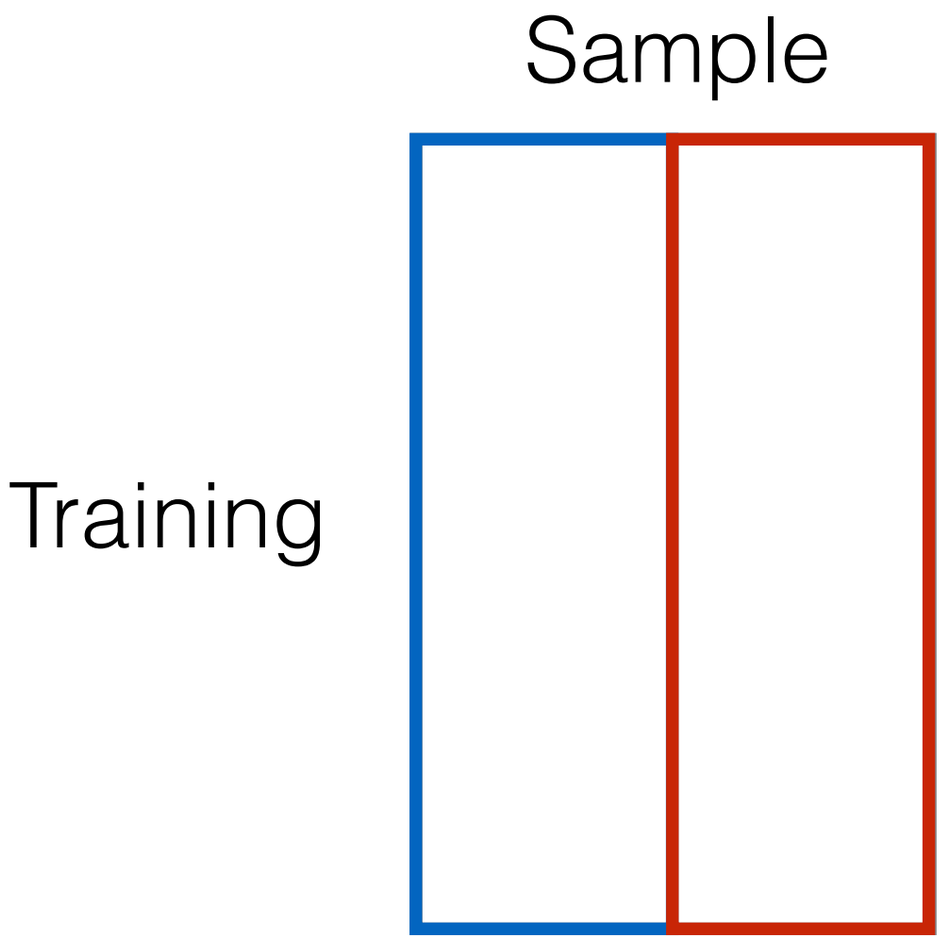
Which is more “useful”?

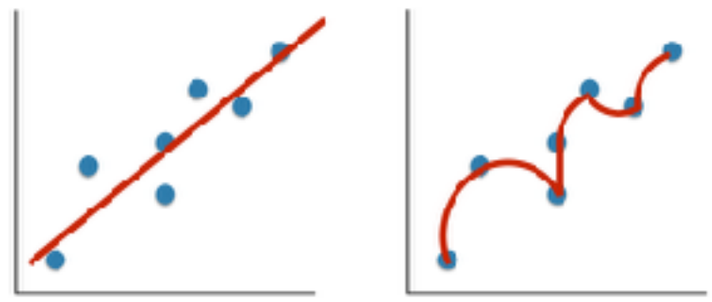
How can we tell?

Cross Validation

- Estimate how accurately a predictive model will perform in practice
- Give an insight on how the model will generalize to an independent dataset

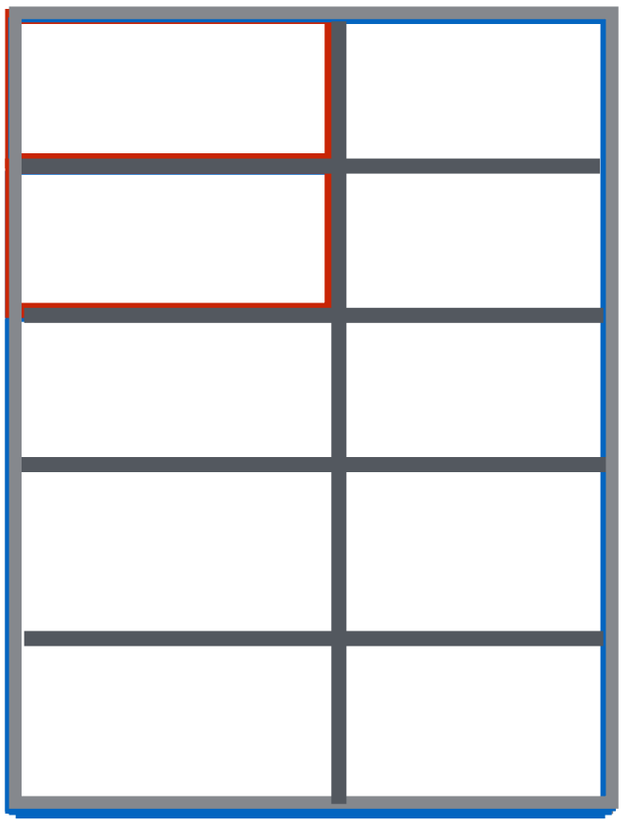
Hold-out Validation

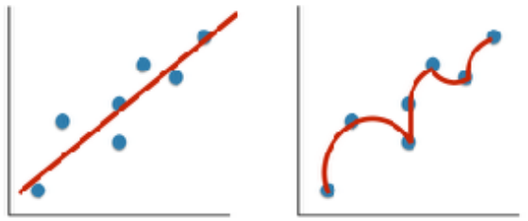




Problem: very dependent on which data are in each group

K-Fold Cross Validation





Training 1	5	2
Training 2	4	2
Training 3	3	1
Training 4	5	4
Training 5	4	2
	4.2	2.2

Calculate how accurate we are in each “fold” and average the answer

Activity

1. Everyone choose a preference: cats or dogs
2. Count how many in each category at your table
3. Write on the board the answer for each table
4. Table 1 is the prediction of Table 2, Table 3 is the prediction of Table 4, etc.
5. What is the error rate of all the predictions?