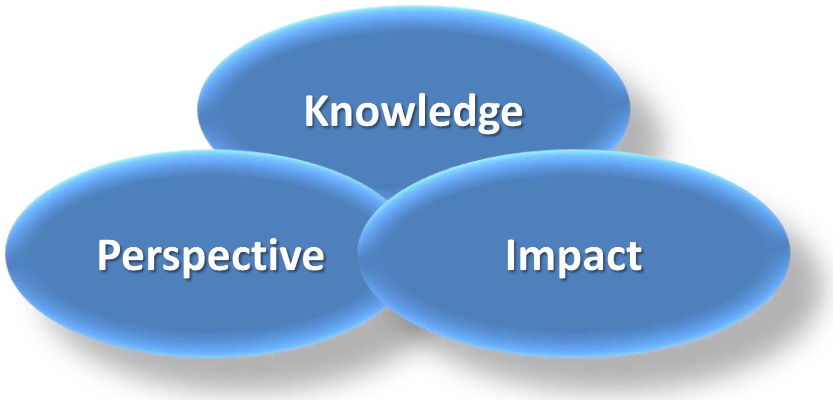
Statistics: A Life Cycle View

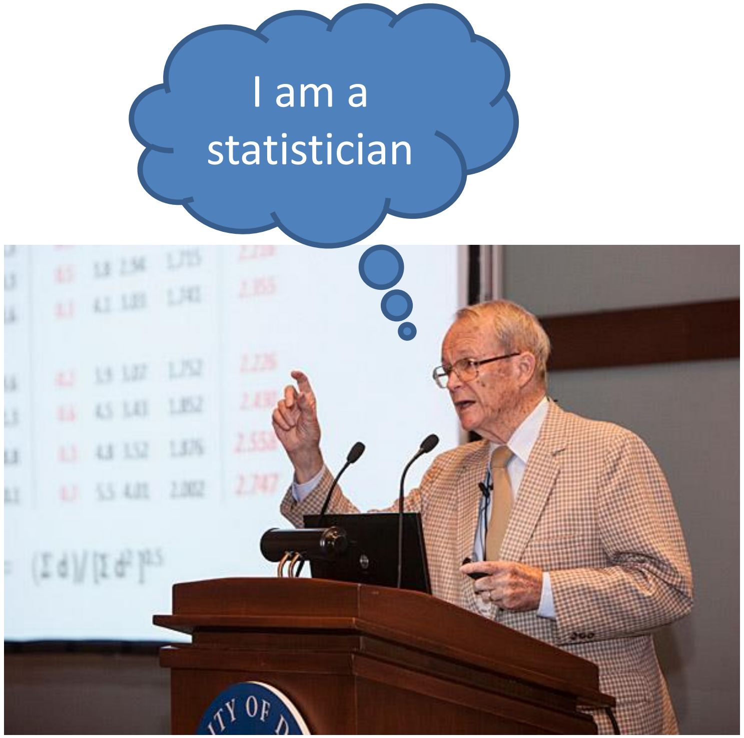
Ron S. Kenett

KPA Ltd., Raanana, Israel
Univ. of Torino, Torino, Italy
NYU Poly, New York, USA
ron@kpa-group.com



Agenda

- Background
- The Trilogy
- Who, how and what
- Discussions (can be ongoing….)



Background

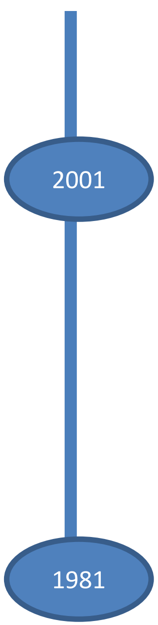
"Much fine work in statistics involves minimal mathematics; some bad work in statistics gets by because of its apparent mathematical content."

1981

David Cox (1981),

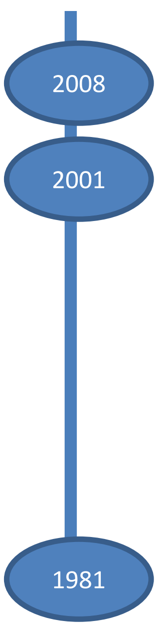
Theory and general principle in statistics, *JRSS(A)*, 144, pp. 289-297.

"…most statisticians seem to agree that statistics is becoming relatively less influential among the information sciences."



Jerry Friedman (2001)
The role of statistics in the data revolution,
International Statistical Review, 69, 1, pp. 5–10.

“The current status of statistics in industry is strong; however, the status of statisticians in industry is possibly at an all-time low.”



Sallie Keller-McNulty (2008)
ASA Presidential Address

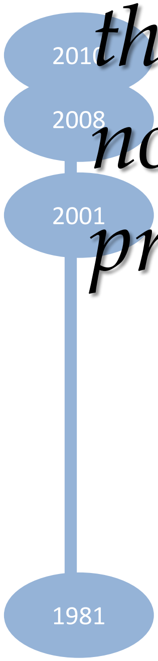
"The state of research in engineering and industrial statistics is not as healthy as it was two to three decades ago. The short-term focus in business and industry has led to drastic cutbacks in industry-based research. Within most statistics departments in academia, engineering and industrial statistics are viewed as mature areas, and do not attract much interest."

V.J. Nair (2008)

Industrial statistics: The gap between research and practice,

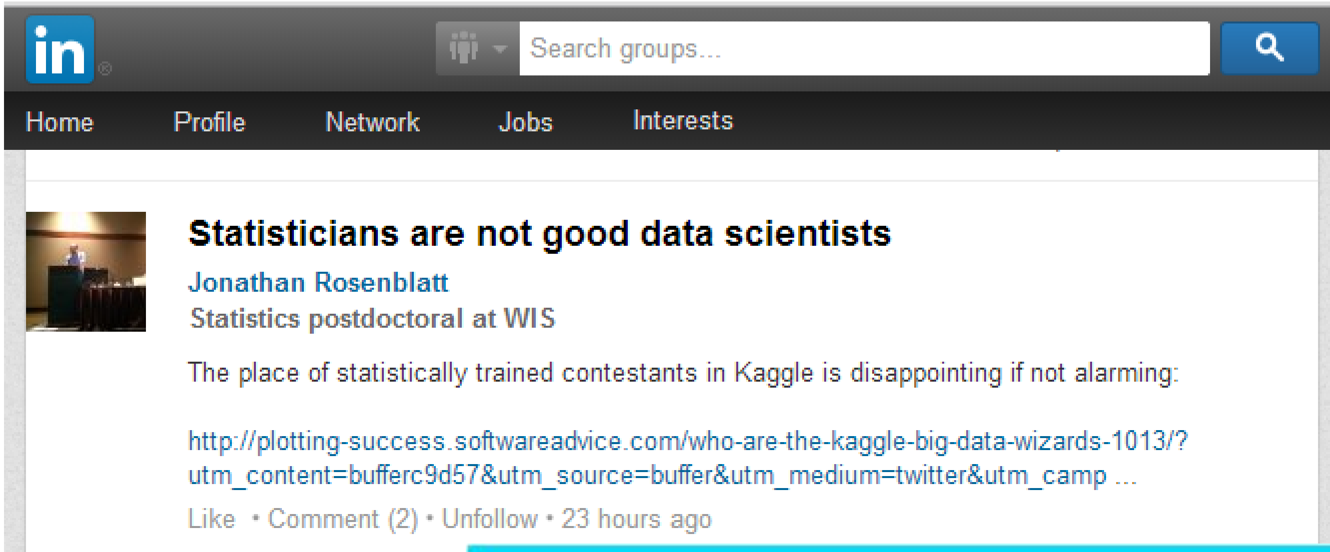
Youden Memorial Address. *ASQ Statistics Division Newsletter*, 27, 1, pp. 5–7.

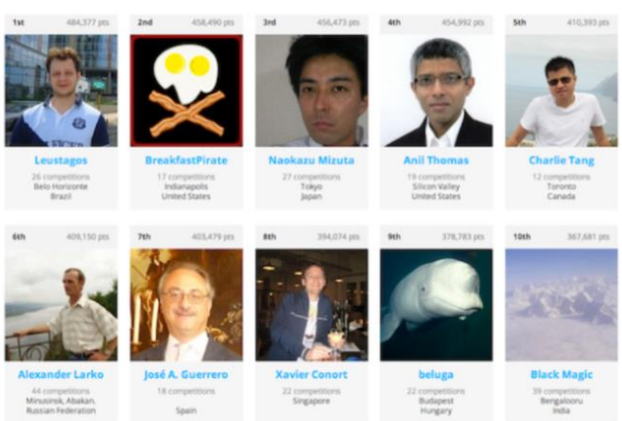
"There appears to be a serious disconnect between academic research in statistics and quality control improvement and actual practice. That is, quality practitioners are not utilizing the latest published research, and researchers are not addressing the research needs perceived by practitioners."



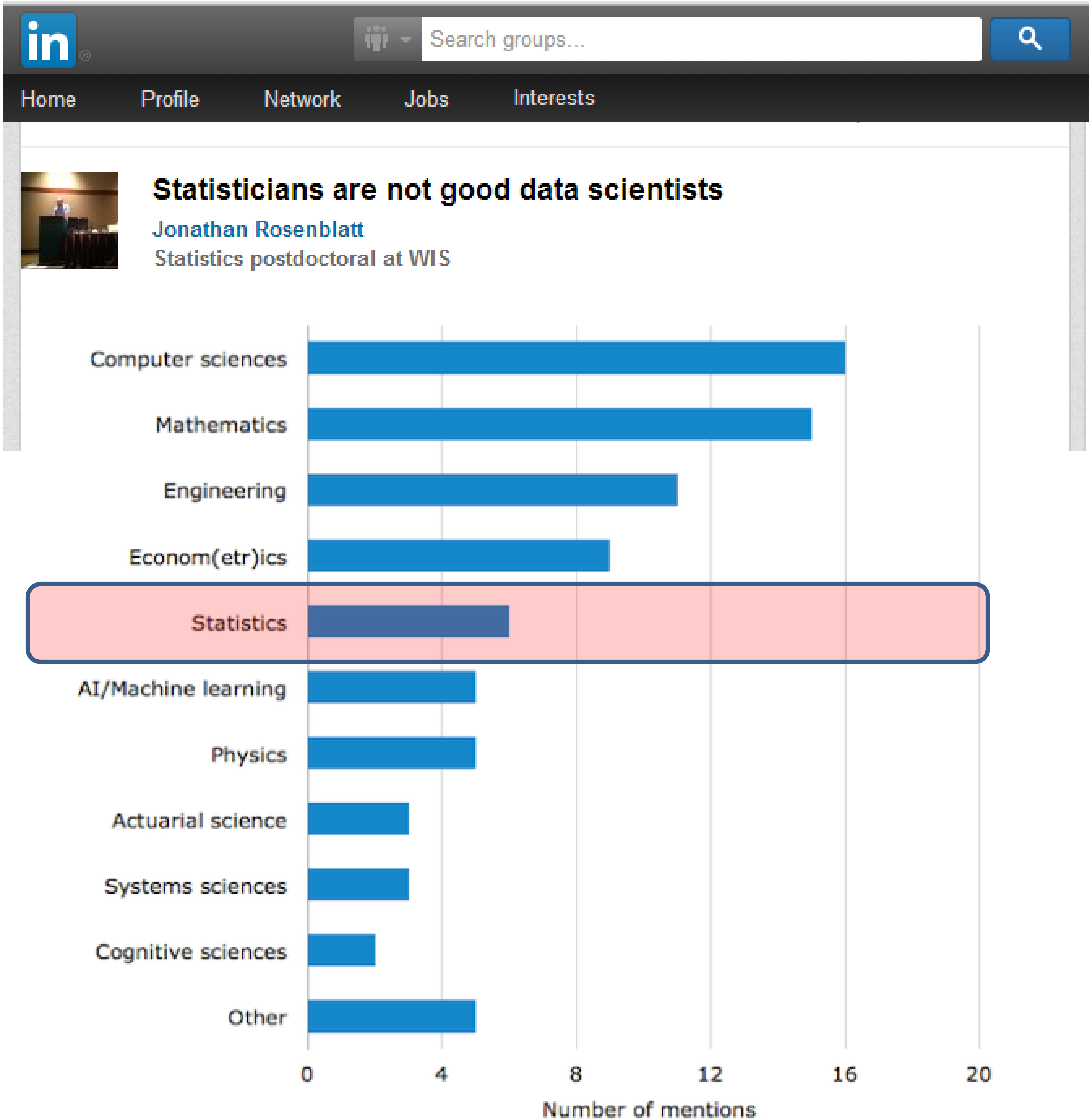
Roger Hoerl and Ron Snee (2010)

Statistical Thinking and Methods in Quality Improvement: A Look to the Future, *Quality Engineering*, 22, 3, pp. 119 -129





The top 10 ranked Kaggle users (as of 10/25/13)

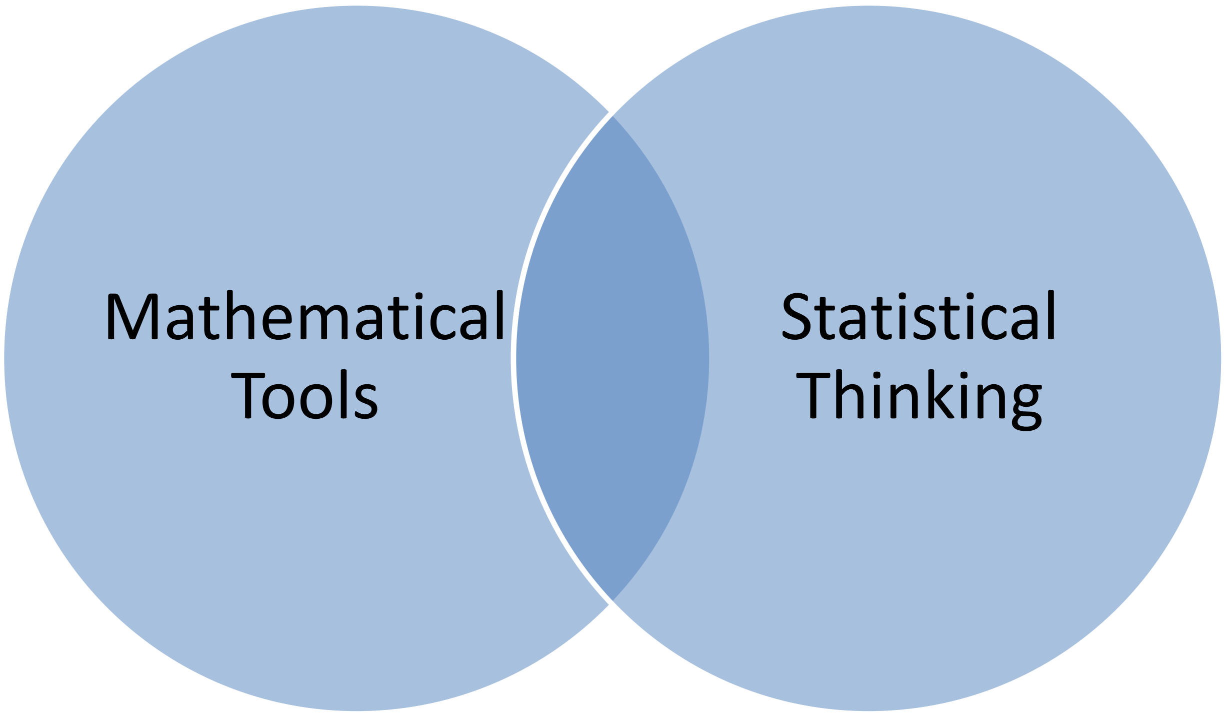

in
Search groups...
Home
Profile
Network
Jobs
Interests
Statisticians are not good data scientists
Jonathan Rosenblatt
Statistics postdoctoral at WIS
Computer sciences
Mathematics
Engineering
Econom(etr)ics
Statistics
AI/Machine learning
Physics
Actuarial science
Systems sciences
Cognitive sciences
Other
0
4
8
12
16
20
Number of mentions

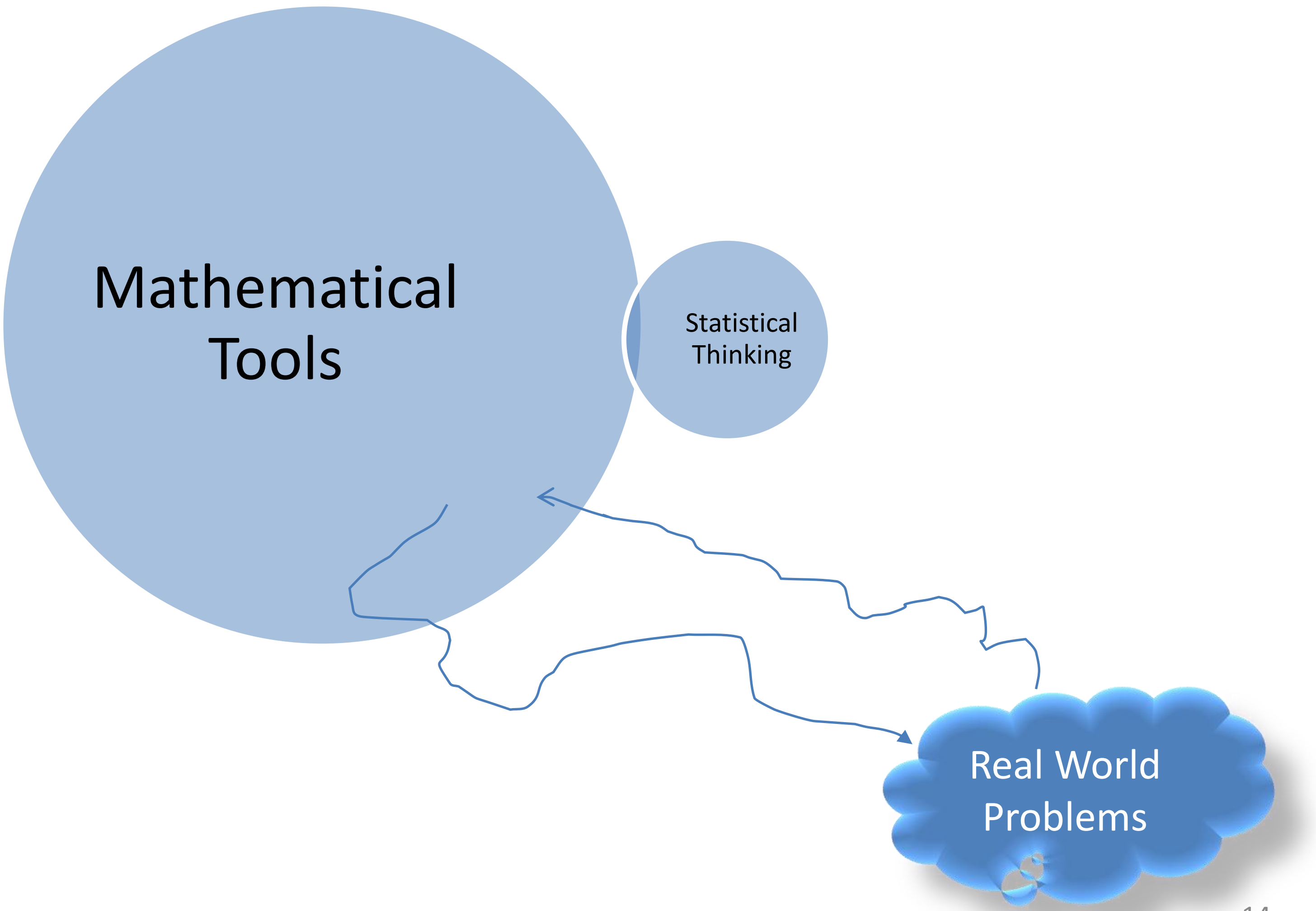
The Market View of Statistics

Take a step backwards, and think of statistics as a product or service (let's call it "S") that you are trying to promote. There are apparently many problems with "S". Basic market research shows that "S" is misused, misunderstood, not appreciated and worst yet, others are offering much more attractive and successful versions of it. For statisticians this is an issue that requires serious considerations.

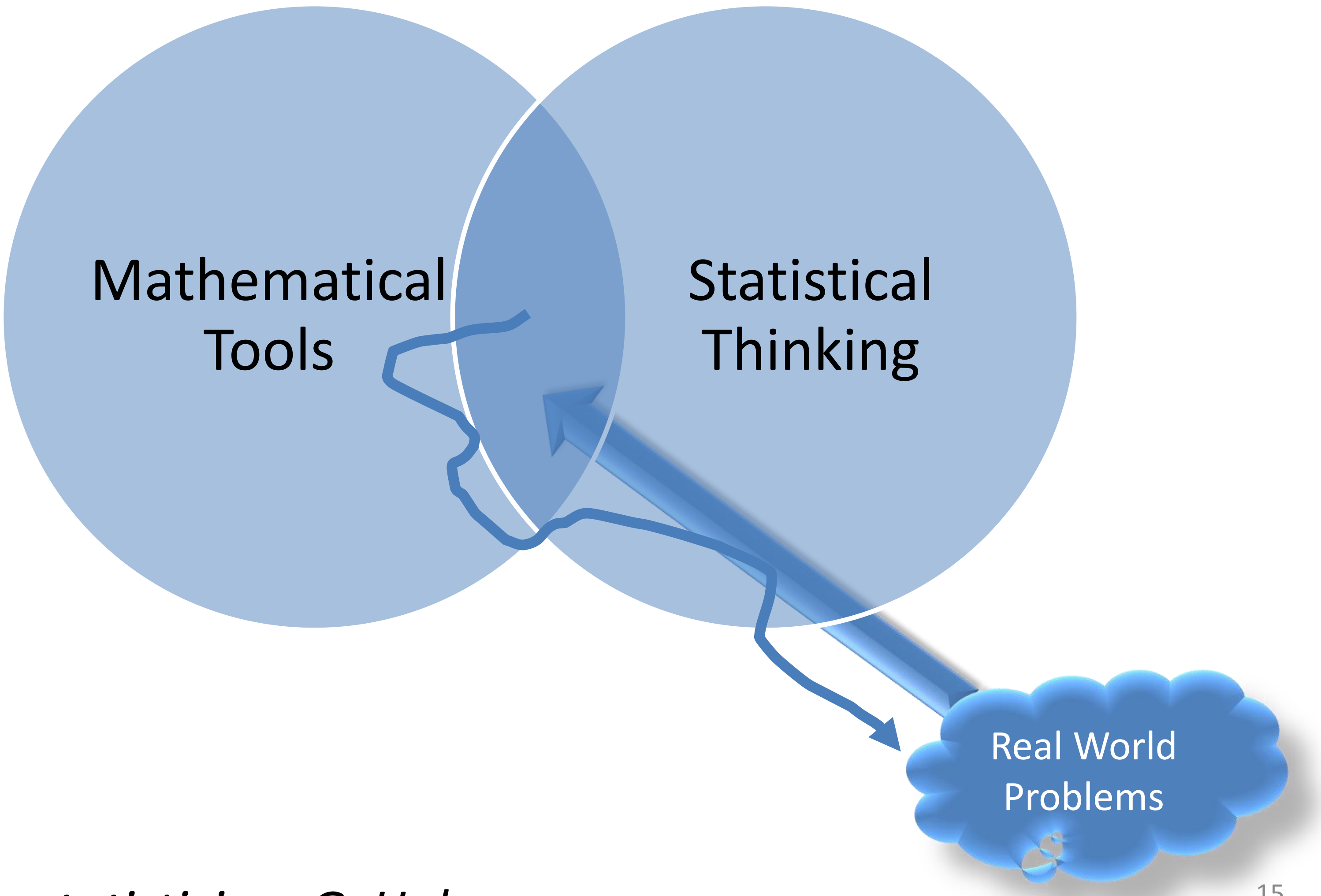
Kenett, R.S. (2012) A Note on the Theory of Applied Statistics . Available at SSRN: http://ssrn.com/abstract=2171179

Applied statistics is about meeting the challenge of **solving real world problems** with **mathematical tools** and **statistical thinking**


Mathematical
Tools
Statistical
Thinking



The mathematical statistician



The proactive statistician, G. Hahn

The Gap

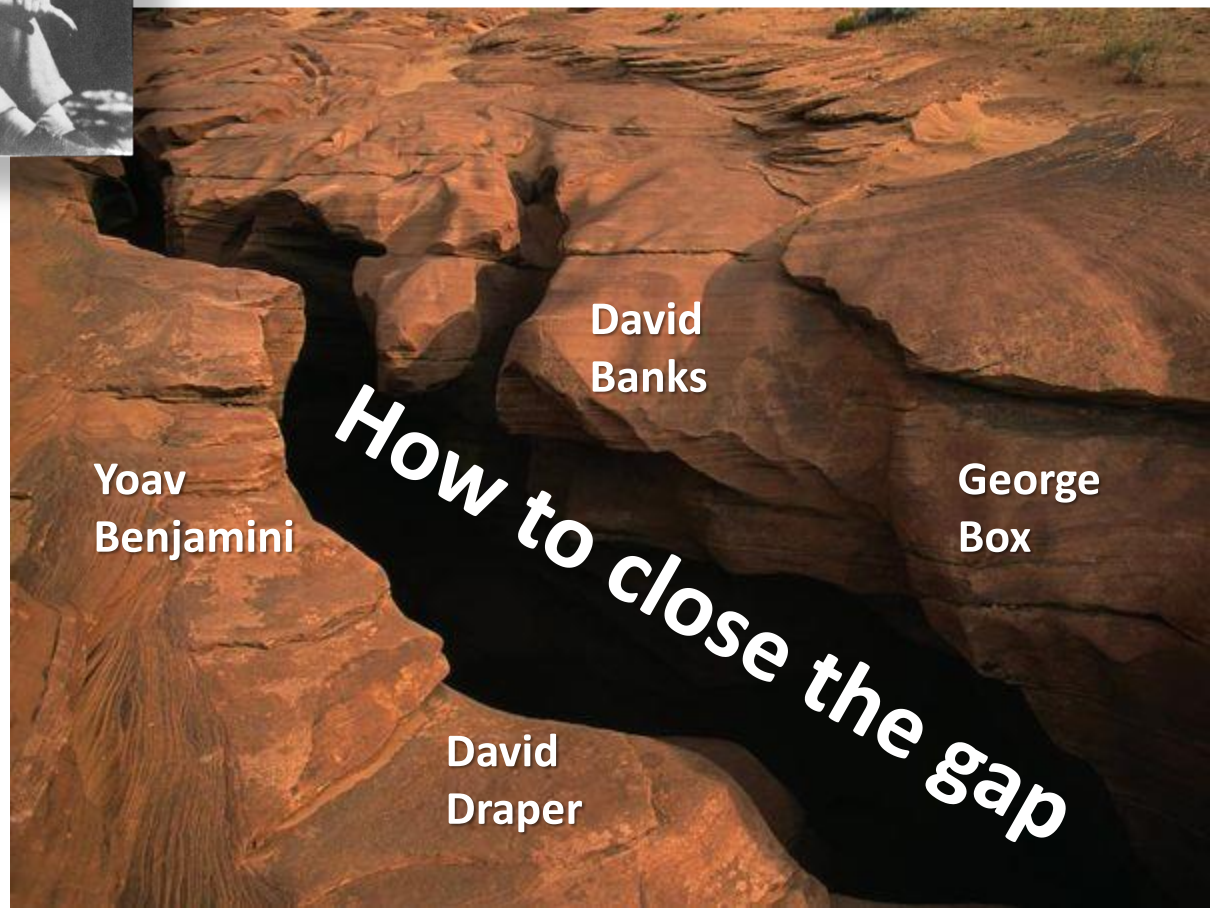
Science

Businesses

Industry

Services

1890-1962



The Theory of Applied Statistics
(á la Draper)

Bayesian Model Specification: Toward a Theory of Applied Statistics

David Draper

Department of Applied Mathematics and Statistics
University of California, Santa Cruz

draper@ams.ucsc.edu
www.ams.ucsc.edu/∼draper

STANFORD UNIVERSITY: WORKSHOP IN BIOSTATISTICS

26 May 2011

The Theory of Applied Statistics *(á la Draper)*

(1) An **axiomatization** of **statistics**.

(2) **Foundations** of **probability** seem (to me) to be **secure**:
(RT Cox, 1946) **Principles** → **Axioms** → **Theorem**:

Logical consistency in **uncertainty quantification** →
justification of **Bayesian reasoning**.

(3) **Foundations** of **inference**, **prediction** and **decision-making** not yet **secure**: fixing this would yield a **Theory of Applied Statistics**, which we **do not yet have**; two remaining **challenges:**

(a) **Cox's Theorem** doesn't **require** You to **pay attention** to a **basic scientific issue**: how **often** do You get the **right answer**?

(b) Too much **ad hockery** in **model specification**: still lacking
Principles → **Axioms** → **Theorems**.

(4) A **Calibration Principle** fixes **3 (a)** via **decision theory**.

(5) **Log scores** help with **3 (b)** via a **Modeling-As-Decision Principle** and a **Prediction Principle**.

The Theory of Applied Statistics *(á la Draper)*

— RT Cox (1946): following **Laplace**, **probability** is a **quantification** of **information** about the **truth** of a **proposition**, constrained to obey **axioms** guaranteeing **internal logical consistency**; this is both **fundamental to science** and **as general as You can get**.

Cox's goal was to identify what **basic rules** $pl(A|B)$ — the **plausibility** (**weight of evidence** in favor) of (the **truth** of) A given B — should follow so that $pl(A|B)$ behaves **sensibly**, where A and B are **propositions** with B **assumed** by You to be **true** and the truth status of A **unknown** to You.

He did this by **identifying** a set of **principles** making **operational** the word **"sensible"** (Jaynes, 2003):

• Suppose You're **willing** to represent **degrees of plausibility** by **real numbers** (i.e., $pl(A|B)$ is a function from propositions A and B to $\Re$);

• You insist that **Your reasoning** be **logically consistent**:

— If a **plausibility assessment** can be arrived at in **more than one way**, then **every possible way** must lead to the **same value**.

The Theory of Applied Statistics *(á la Draper)*

— You always take into account **all of the evidence** You judge to be **relevant** to the **plausibility assessment** under consideration (this is the **Bayesian** version of **objectivity**).

— You always represent **equivalent states of information** by **equivalent plausibility assignments**.

From these **principles** Cox derived a set of **axioms**:

• The **plausibility** of a **proposition** determines the **plausibility** of the proposition's **negation**; each **decreases** as the other **increases**.

• The **plausibility** of the **conjunction** $AB = (A \text{ and } B)$ of **two propositions** A, B **depends** only on the **plausibility** of B and that of $\{A$ **given** that B is true$\}$ (or **equivalently** the **plausibility** of A and that of $\{B$ **given** that A is true$\}$).

• Suppose AB is **equivalent** to CD; then if You acquire **new information** A and later acquire **further new information** B, and **update** all **plausibilities** each time, the **updated plausibilities** will be the **same** as if You had **first acquired new information** C and **then acquired further new information** D.

The Theory of Applied Statistics *(á la Draper)*

From these **axioms** Cox proved a **theorem** showing that **uncertainty quantification** about **propositions** behaves in **one and only one way**:

Theorem: If You accept **Cox's axioms**, then to be **logically consistent** You **must** quantify uncertainty as follows:

- Your **plausibility operator** $pl(A|B)$ — for **propositions** A and B — can be referred to as Your **probability** $P(A|B)$ that A is true, **given** that You regard B as true, and $0 \leq P(A|B) \leq 1$, with **certain truth** of A (given B) represented by **1** and **certain falsehood** by **0**.
- **(normalization)** $P(A|B) + P(\overline{A}|B) = 1$, where $\overline{A} = (\text{not } A)$.
- **(the product rule)**:

$$P(A\,B|C) = P(A|C) \cdot P(B|A\,C) = P(B|C) \cdot P(A|B\,C).$$

The **proof** (see, e.g., Jaynes (2003)) involves deriving two **functional equations** $F[F(x,y),z] = F[x, F(y,z)]$ and $x\,S\left[\frac{S(y)}{x}\right] = y\,S\left[\frac{S(x)}{y}\right]$ that $pl(A|B)$ must satisfy and then **solving** those equations.

The Theory of Applied Statistics *(á la Draper)*

Cox's Theorem and its **corollaries** provide **no constraints on the specification process**, apart from the requirement that **all probability distributions** be **proper** (integrate or sum to **1**).

In my view, in seeking **answers** to these **specification questions**, as a **profession** we're approximately where the **discipline of statistics** was in arriving at an **optimal theory of probability before Cox's work**: many people have made **ad-hoc suggestions** (**some** of them **good**), but **little formal progress** has been made.

Developing (1) **principles**, (2) **axioms** and (3) **theorems** about **optimal specification** could be regarded as creating a Theory of Applied Statistics, which we **do not yet have**.

J. R. Statist. Soc. A, (1980),
143, *Part* 4, *pp.* 383–430

The Theory of Applied Statistics
(á la Box)

Sampling and Bayes' Inference in Scientific Modelling and ... ess

[... up on

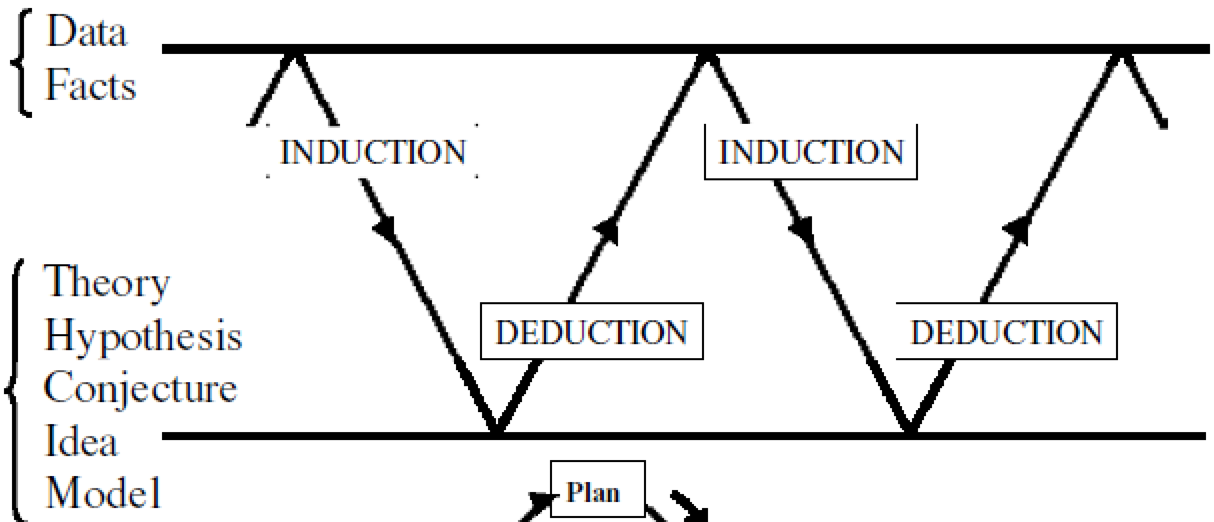
Warning

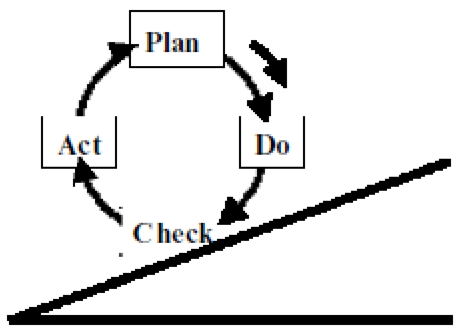
We do not teach tools and methods for doing that

To clear up some misunderstandings and to set my reply in context, let me first make clear what I regard as the proper role of a statistican. This is not as the analyst of a single set of data, nor even as the

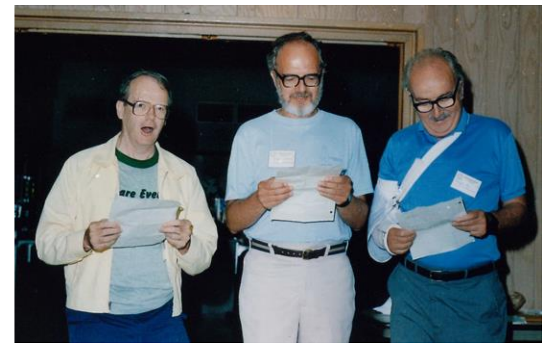
"This is not as the analyst of a single set of data, nor even as the designer and analyzer of a single experiment, but rather as a colleague working with an investigator throughout the whole course of iterative deductive-inductive investigation."

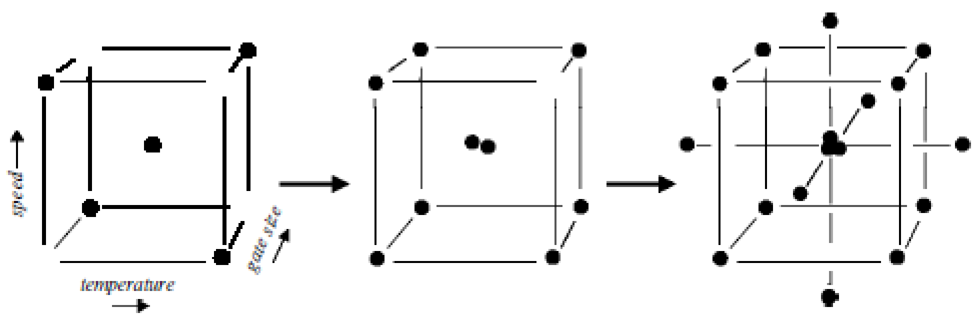
The Theory of Applied Statistics *(á la Box)*

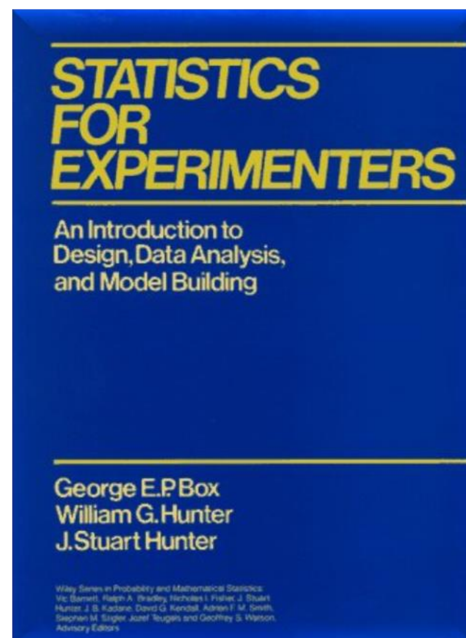




How we learn





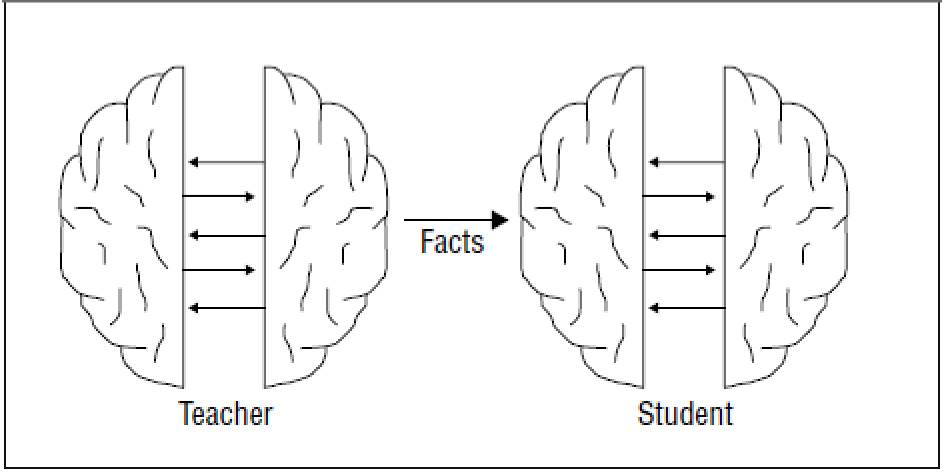


Chapter 16: Mechanistic Model Building

Chapter 17: Study of Variation

The Theory of Applied Statistics *(á la Box)*

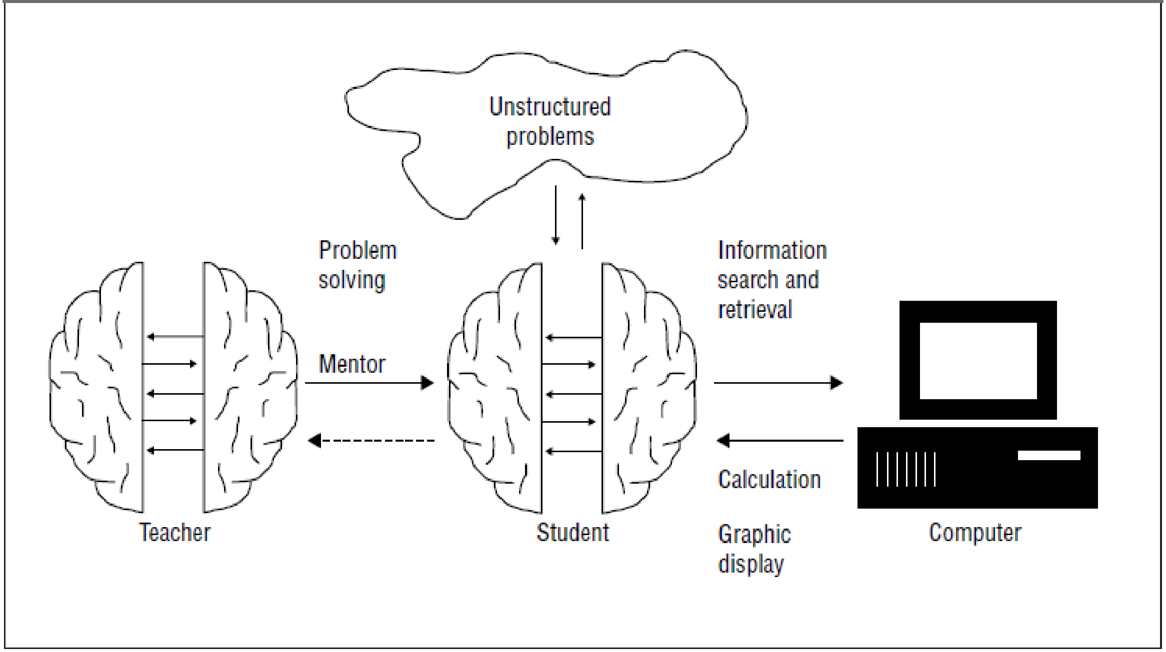
Figure 2. Traditional Method of Teaching



George Box (1997), Scientific Method: The Generation of Knowledge and Quality, *Quality Progress*, January, pp. 47-50.

How we teach

Figure 3. A Model for Modern Teaching



David Banks

Statistics, Politics, and Policy

Volume 2, Issue 1 2011 *Article 4*

Reproducible Research: A Range of Response

David Banks, *Duke University*

As a former editor of the Journal of the American Statistical Association, my own sense is that very few applied papers are perfectly **reproducible**. Most do not come with code or data, and even if they did, I expect a careful check would and discrepancies from the published paper. The reasons are innocent: code written by graduate students is continually tweaked and has sketchy documentation.

The exact data cleaning procedures are not perfectly remembered when the final version of the paper is written, or may be muddled by miscommunication among multiple authors. And even if a conscientious researcher provided a full description of every cleaning step, every model fitting choice, and all aspects of variable selection, the resulting paper would be so long and tedious that no doubt the foolish editor would demand that it be shortened.

"Yoav's current interest is the **replicability** problem in science: too often, the results of studies gaining headlines cannot be replicated by other experimenters. Part of the problem is the use of statistical tools that fail to address the challenge of selective inference. He is trying to develop statistical tools that will aid researchers to cope with this problem, from the areas of Medicine, Epidemiology, Genomics, Bioinformatics, Neuroscience and behavior."

http://en.wikipedia.org/wiki/Yoav_Benjamini

The multi-family selective inference problem

We select interesting/significant/promising families

The uninteresting families loose importance

and are dropped/ignored from the reported results

(or hidden in the available database/online appendix)

We wish to infer on the selected families

- test hypotheses within
- set confidence interval
- estimate

Selection adjusted separate testing of families

Let P_i be the p-values for the hypotheses in family i,

$P=\{P_1,P_2,\ldots P_m\}$. $I=\{1,2,\ldots m\}$.

Any data based selection procedure of families yields $S(P)$ in I. Let $|S(P)|$ be the (random) number of families selected.

The control of error $E(C)$ (FDR, FWER, and others) on the average over the selected families means

$$E\left(\frac{\sum_{i\in S(P)} C_i}{|S(P)|}\right)\le q$$

Yoav Benjamini

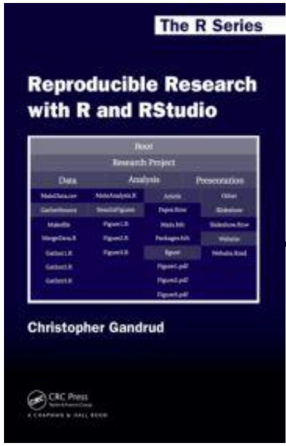
A variety of error-rates

Unadjusted inference	$E(V/m) \le \alpha$
False Excedance Rate	$\Pr(V/R \ge q) \le \alpha$
k- FDR	$E((V-k)_+/R) \le q$
False Discovery Rate	$E(V/R) \le \alpha = q$
k-FWER	$P(V \ge k) \le \alpha$
Strong control of FWER	$\Pr(V\ge 1) \le \alpha$
Per family Error-rate	$E(V) \le \alpha$

All above are of the form $E(C)$

But not Fdr = E(V)/E(R); local fdr(z); positive FDR

Presented at The San Francisco Chapter of ASA February 29, 2012



Christopher Gandrud

Sweave

knitr

Replicability is not Reproducibility: Nor is it Good Science

Chris Drummond

Proc. of the Evaluation Methods for Machine Learning Workshop at the 26 th ICML, Montreal, Canada, 2009.

Chris Drummond CHRIS.DRUMMOND@NRC-CNRC.GC.CA
Institute for Information Technology
National Research Council Canada
Ottawa, Ontario, Canada, K1A 0R6

"Reproducibility requires changes; replicability avoids them. A critical point of reproducing an experimental result is that irrelevant things are intentionally not replicated. One might say, one should replicate the result not the experiment."

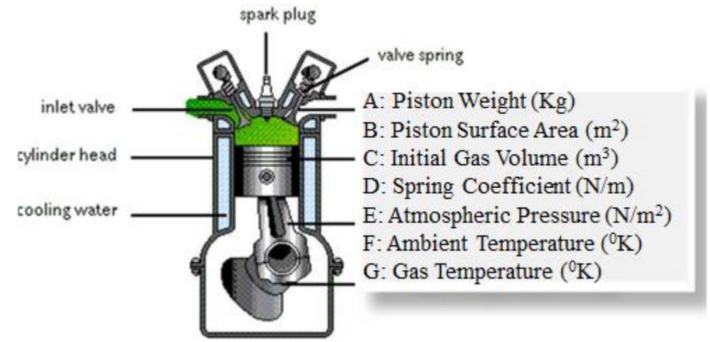
Is this reproducible?

Is this replicable?

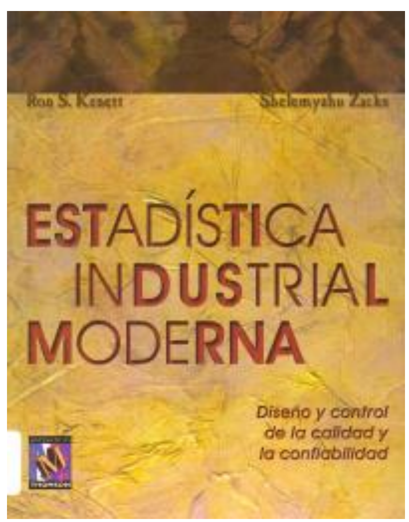
What did we learn $\neq$ *What did we do*

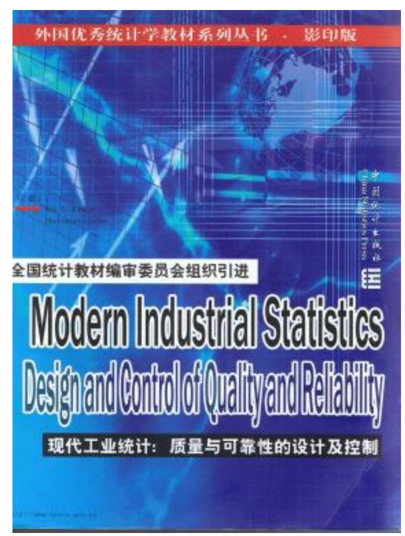
Information

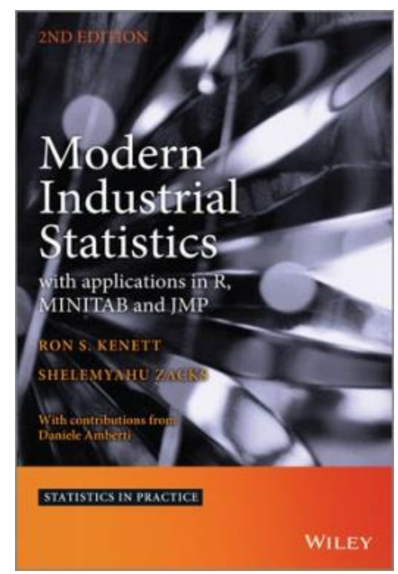
Closing the Gap: A Pedagogical Approach



- Understand why – the motivation
- Learn how to do it – using the computer
- Study the foundations – using mathematics
- Practice, practice, practice







Closing the Gap: A Pedagogical Approach

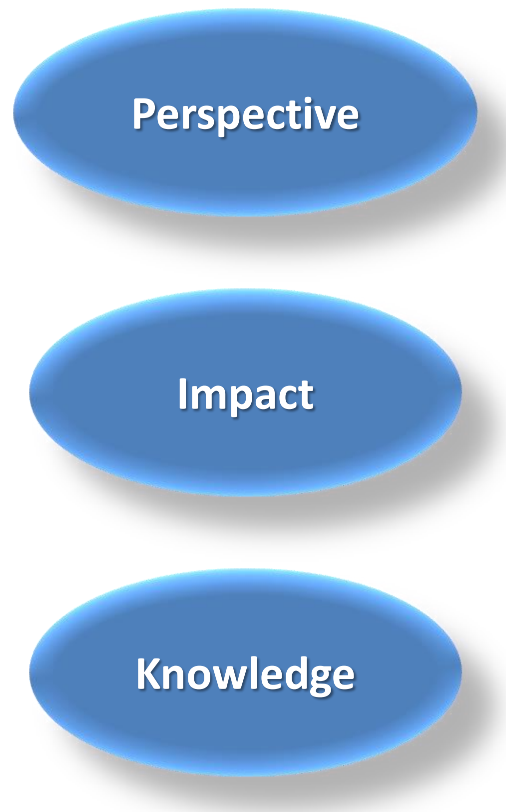
- The pedagogical structure of *Modern Industrial Statistics* combines a **practical** approach, with **theoretical** foundations and **computer** support. It is intended for students and instructors who have an interest in studying modern methods by combining these three elements.
- The first edition referred to S-Plus, MINITAB and compiled QuickBasic code. The second edition provides examples and procedures in the now popular R language and also refers to MINITAB and JMP. Each of these three computer platforms carries unique advantages. Focusing on only one or two of these is also possible.
- Exercises are provided at the end of each chapter in order to provide more opportunities to learn and test your knowledge.

From preface to second edition of Kenett and Zacks, *Modern Industrial Statistics with applications in R, MINITAB and JMP, Wiley 2014*

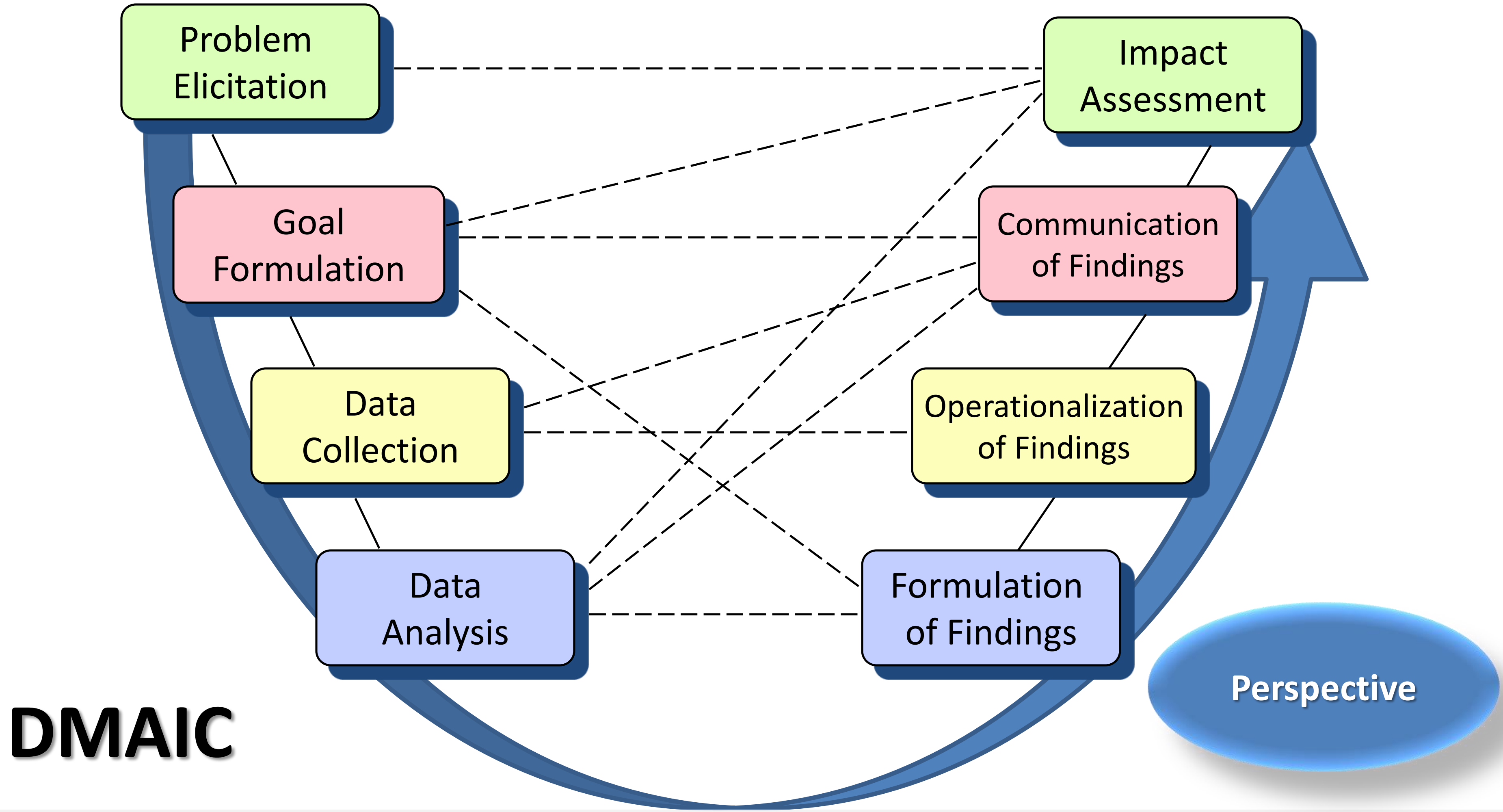
The Trilogy

The Trilogy of Applied Statistics

- Consider a life cycle view
- Assess impact
- Generate knowledge



Statistics: A life cycle view



Kenett, R.S. and Thyregod, P. (2006) Aspects of statistical consulting not taught by academia, *Statistica Neerlandica*, 60 (3): 396-412.

Problem Elicitation

Perspective

Greenfield, T. (1987) Consultant's cameos: A chapter of encounters. pp. 11-25 in Hand, D.J. and B.S. Everitt eds, *The statistical consultant in action*, Cambridge University Press

Assessing Impact

Impact

Practical Statistical Efficiency (PSE)

$$PSE = \text{function}[E\{R\}, T\{I\}, P\{I\}, V\{PS\}, P\{S\}, V\{P\}, V\{M\}, V\{D\}]$$

V{D} = value of the data actually collected

V{M} = value of the statistical method employed

V{P} = value of the problem to be solved

P{S} = probability that the problem actually gets solved

V{PS} = value of the problem being solved

P{I} = probability the solution is actually implemented

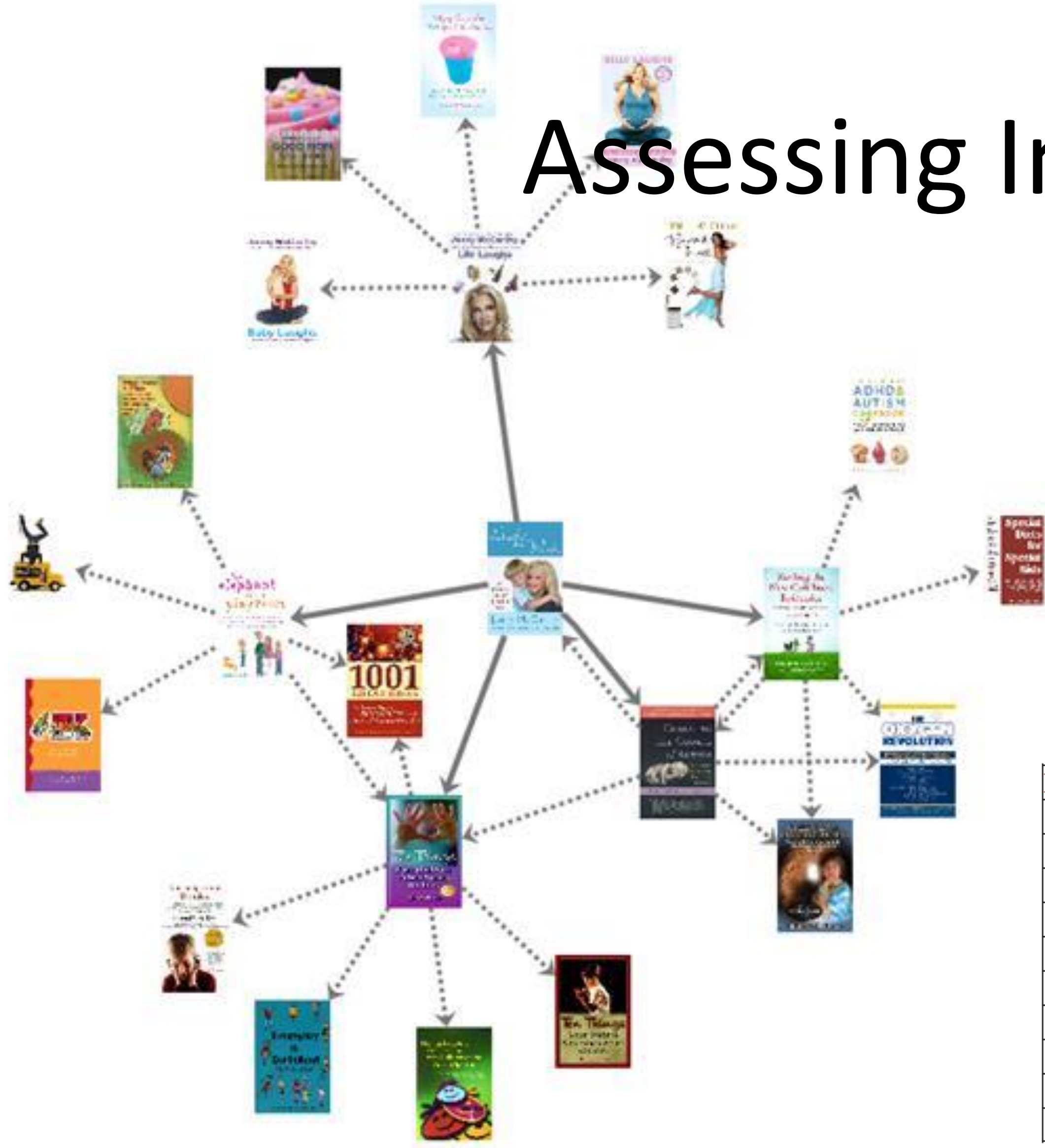
T{I} = time the solution stays implemented

E{R} = expected number of replications

Kenett, R.S., Coleman, S.Y. and Stewardson, D. (2003). Statistical Efficiency: The Practical Perspective, *Quality and Reliability Engineering International*, 19: 265-272.

Assessing Impact

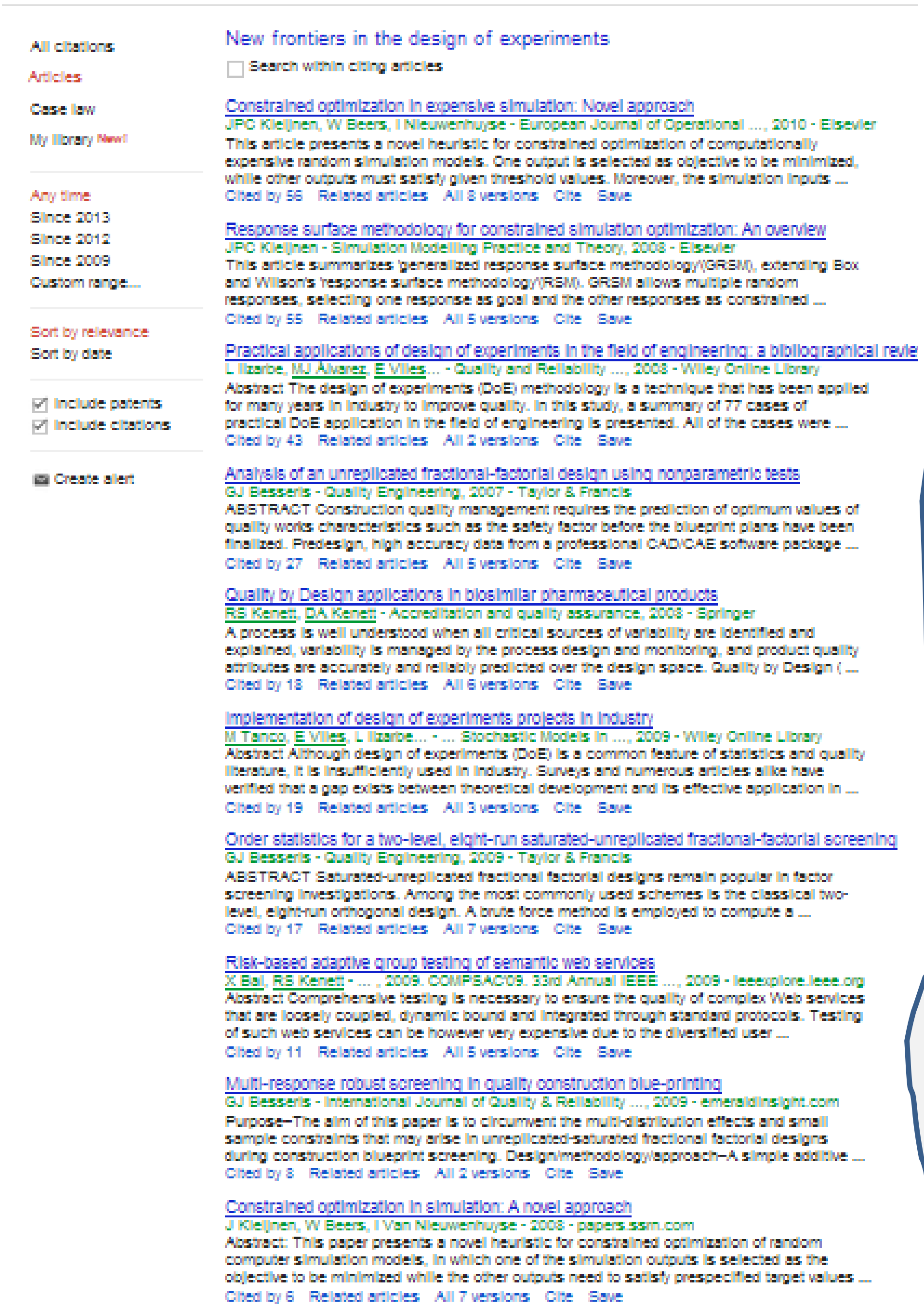
Impact



Papers produced by experiments; literature papers cited by those produced by experiments; and literature papers citing experiment papers. citations from experiment papers to literature papers with *0to1* and the citations received by experiment papers from literature papers *1to0*

Experiments	0to0	0to1	1to0	1to1	H-index	>500 cit
ALEPH	93	223	1068	13360	77	4
DELPHI	96	230	795	10949	66	4
L3	119	289	773	10522	63	4
OPAL	120	374	936	13181	79	4
CDF	839	1338	4005	34900	119	6
D0	831	953	3356	24843	85	3
ALICE	377	435	541	17345	34	1
ATLAS	3272	1070	6891	54648	78	4
CMS	1708	1018	4918	55906	69	4
LHCb	243	325	592	22396	33	1

Gal Oestreicher-Singer and Arun Sundararajan (2012) Recommendation networks and the long tail of electronic commerce E, *MIS Quarterly Vol. 36 No. 1 pp. 65-83/March 2012*
www.eiburs.unimi.it Cost Benfit Analysis in the Research, Development and Innovation Sector.
Carrazza S. Ferrara A., Salini S. (2013) Research infrastructures in the LHC era: a scientiometric approach, EIB



All citations
Articles
Case law
My library New!

Any time
Since 2013
Since 2012
Since 2009
Custom range...

Sort by relevance
Sort by date

Include patents
Include citations

Create alert

New frontiers in the design of experiments

Search within citing articles

[Constrained optimization in expensive simulation: Novel approach]
JPC Kleijnen, W Beers, I Nieuwenhuyse - European Journal of Operational ..., 2010 - Elsevier
This article presents a novel heuristic for constrained optimization of computationally expensive random simulation models. One output is selected as objective to be minimized, while other outputs must satisfy given threshold values. Moreover, the simulation inputs ...
Cited by 56 Related articles All 8 versions Cite Save

[Response surface methodology for constrained simulation optimization: An overview]
JPC Kleijnen - Simulation Modelling Practice and Theory, 2008 - Elsevier
This article summarizes 'generalized response surface methodology'(GRSM), extending Box and Wilson's 'response surface methodology'(RSM). GRSM allows multiple random responses, selecting one response as goal and the other responses as constrained ...
Cited by 55 Related articles All 5 versions Cite Save

[Practical applications of design of experiments in the field of engineering: a bibliographical revie]
L Ilzarbe, MJ Álvarez, E Viles... - Quality and Reliability ..., 2008 - Wiley Online Library
Abstract The design of experiments (DoE) methodology is a technique that has been applied for many years in industry to improve quality. In this study, a summary of 77 cases of practical DoE application in the field of engineering is presented. All of the cases were ...
Cited by 43 Related articles All 2 versions Cite Save

[Analysis of an unreplicated fractional-factorial design using nonparametric tests]
GJ Besseris - Quality Engineering, 2007 - Taylor & Francis
ABSTRACT Construction quality management requires the prediction of optimum values of quality works characteristics such as the safety factor before the blueprint plans have been finalized. Predesign, high accuracy data from a professional CAD/CAE software package ...
Cited by 27 Related articles All 5 versions Cite Save

[Quality by Design applications in biosimilar pharmaceutical products]
RS Kenett, DA Kenett - Accreditation and quality assurance, 2008 - Springer
A process is well understood when all critical sources of variability are identified and explained, variability is managed by the process design and monitoring, and product quality attributes are accurately and reliably predicted over the design space. Quality by Design (...
Cited by 18 Related articles All 6 versions Cite Save

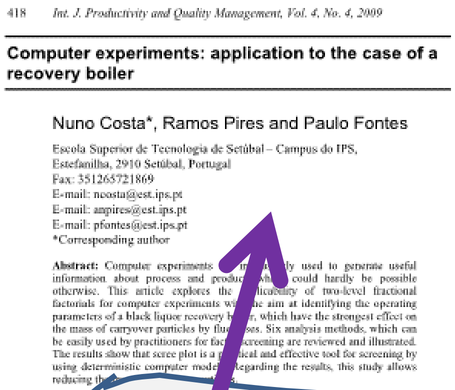
[Implementation of design of experiments projects in industry]
M Tanco, E Viles, L Ilzarbe... - ... Stochastic Models in ..., 2009 - Wiley Online Library
Abstract Although design of experiments (DoE) is a common feature of statistics and quality literature, it is insufficiently used in industry. Surveys and numerous articles alike have verified that a gap exists between theoretical development and its effective application in ...
Cited by 19 Related articles All 3 versions Cite Save

[Order statistics for a two-level, eight-run saturated-unreplicated fractional-factorial screening]
GJ Besseris - Quality Engineering, 2009 - Taylor & Francis
ABSTRACT Saturated-unreplicated fractional factorial designs remain popular in factor screening investigations. Among the most commonly used schemes is the classical two-level, eight-run orthogonal design. A brute force method is employed to compute a ...
Cited by 17 Related articles All 7 versions Cite Save

[Risk-based adaptive group testing of semantic web services]
X Bai, RS Kenett - ... , 2009. COMPSAC'09. 33rd Annual IEEE ..., 2009 - ieeexplore.ieee.org
Abstract Comprehensive testing is necessary to ensure the quality of complex Web services that are loosely coupled, dynamic bound and integrated through standard protocols. Testing of such web services can be however very expensive due to the diversified user ...
Cited by 11 Related articles All 5 versions Cite Save

[Multi-response robust screening in quality construction blue-printing]
GJ Besseris - International Journal of Quality & Reliability ..., 2009 - emeraldinsight.com
Purpose–The aim of this paper is to circumvent the multi-distribution effects and small sample constraints that may arise in unreplicated-saturated fractional factorial designs during construction blueprint screening. Design/methodology/approach–A simple additive ...
Cited by 8 Related articles All 2 versions Cite Save

[Constrained optimization in simulation: A novel approach]
J Kleijnen, W Beers, I Van Nieuwenhuyse - 2008 - papers.ssrn.com
Abstract: This paper presents a novel heuristic for constrained optimization of random computer simulation models, in which one of the simulation outputs is selected as the objective to be minimized while the other outputs need to satisfy prespecified target values ...
Cited by 6 Related articles All 7 versions Cite Save



418 Int. J. Productivity and Quality Management, Vol. 4, No. 4, 2009

Computer experiments: application to the case of a recovery boiler

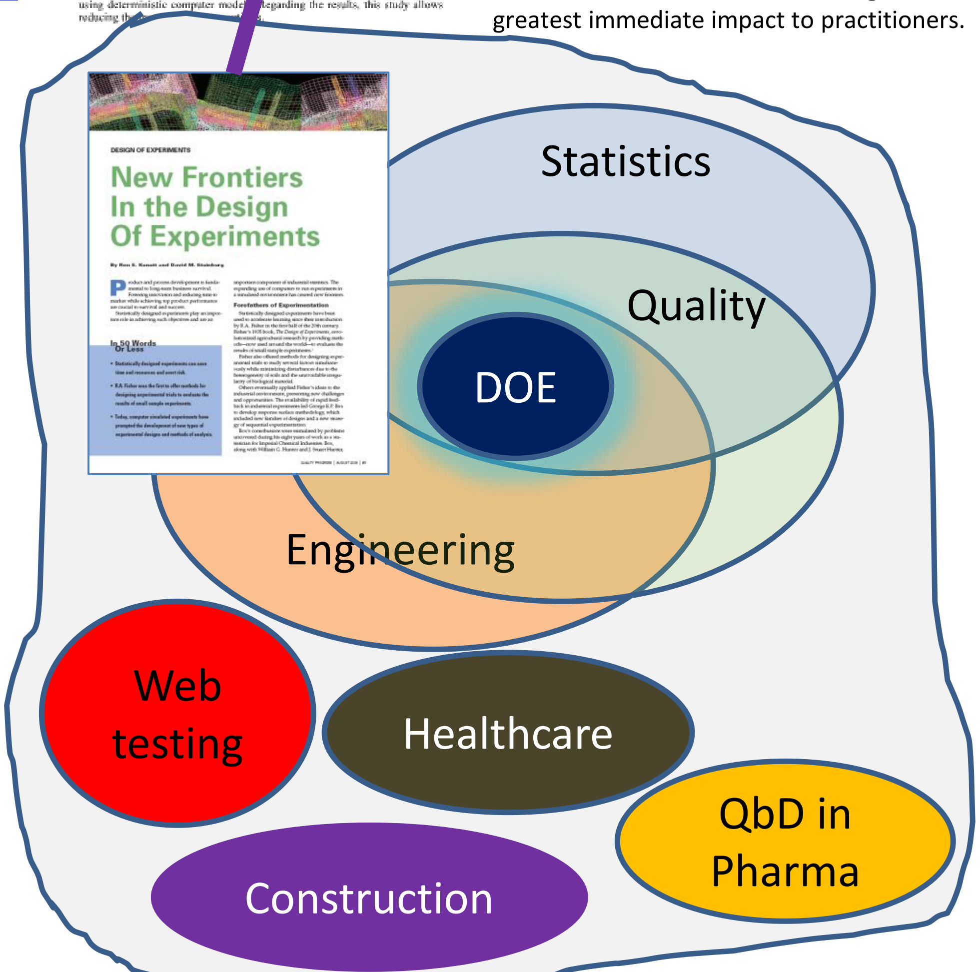
Nuno Costa*, Ramos Pires and Paulo Fontes

Escola Superior de Tecnologia de Setúbal – Campus do IPS, Estefanilha, 2910 Setúbal, Portugal
Fax: 351265721869
E-mail: ncosta@est.ips.pt
E-mail: anpires@est.ips.pt
E-mail: pfontes@est.ips.pt
*Corresponding author

Abstract: Computer experiments ... ly used to generate useful information about process and produc... could hardly be possible otherwise. This article explores the ...licability of two-level fractional factorials for computer experiments wi...he aim at identifying the operating parameters of a black liquor recovery b...r, which have the strongest effect on the mass of carryover particles by flu...ses. Six analysis methods, which can be easily used by practitioners for fac... screening are reviewed and illustrated. The results show that scree plot is a ...ical and effective tool for screening by using deterministic computer mode... Regarding the results, this study allows reducing th...

Impact

Brumbaugh Award for the paper making the largest contribution to industrial quality control.
Nelson Award for the article having the greatest immediate impact to practitioners.



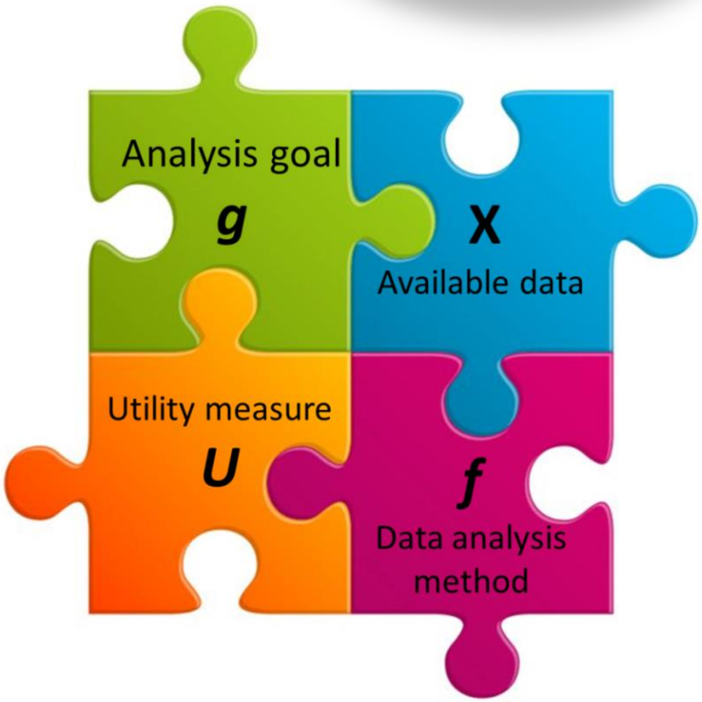
Kenett, R.S. and Steinberg, D. (2006), New Frontiers in Design of Experiments, *Quality Progress*, pp. 61-65, August 2006.

Information Quality

Knowledge

The potential of a particular dataset to achieve a particular goal using a given empirical analysis method

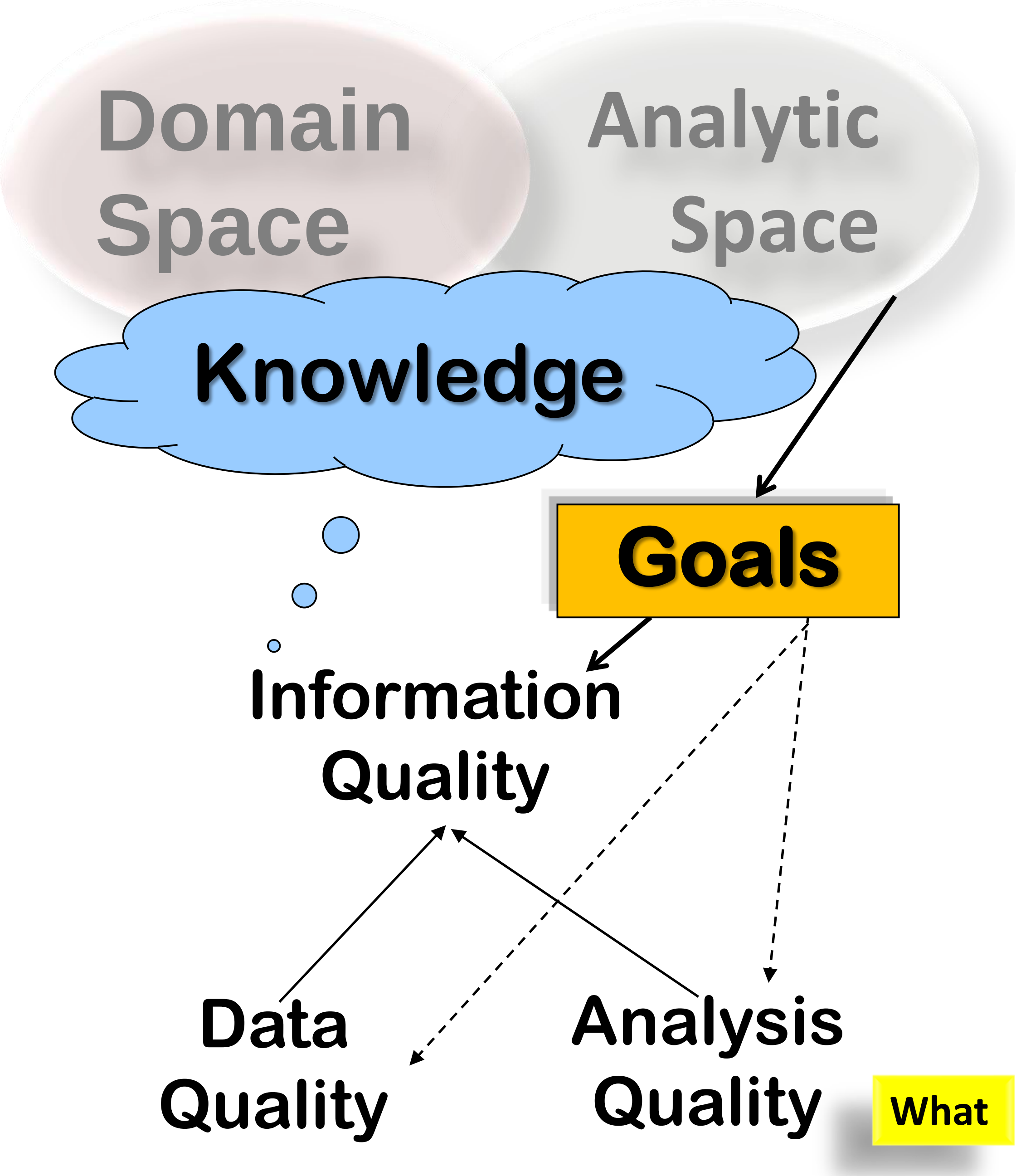
g	**A specific analysis goal**
X	**The available dataset**
f	**An empirical analysis method**
U	**A utility measure**



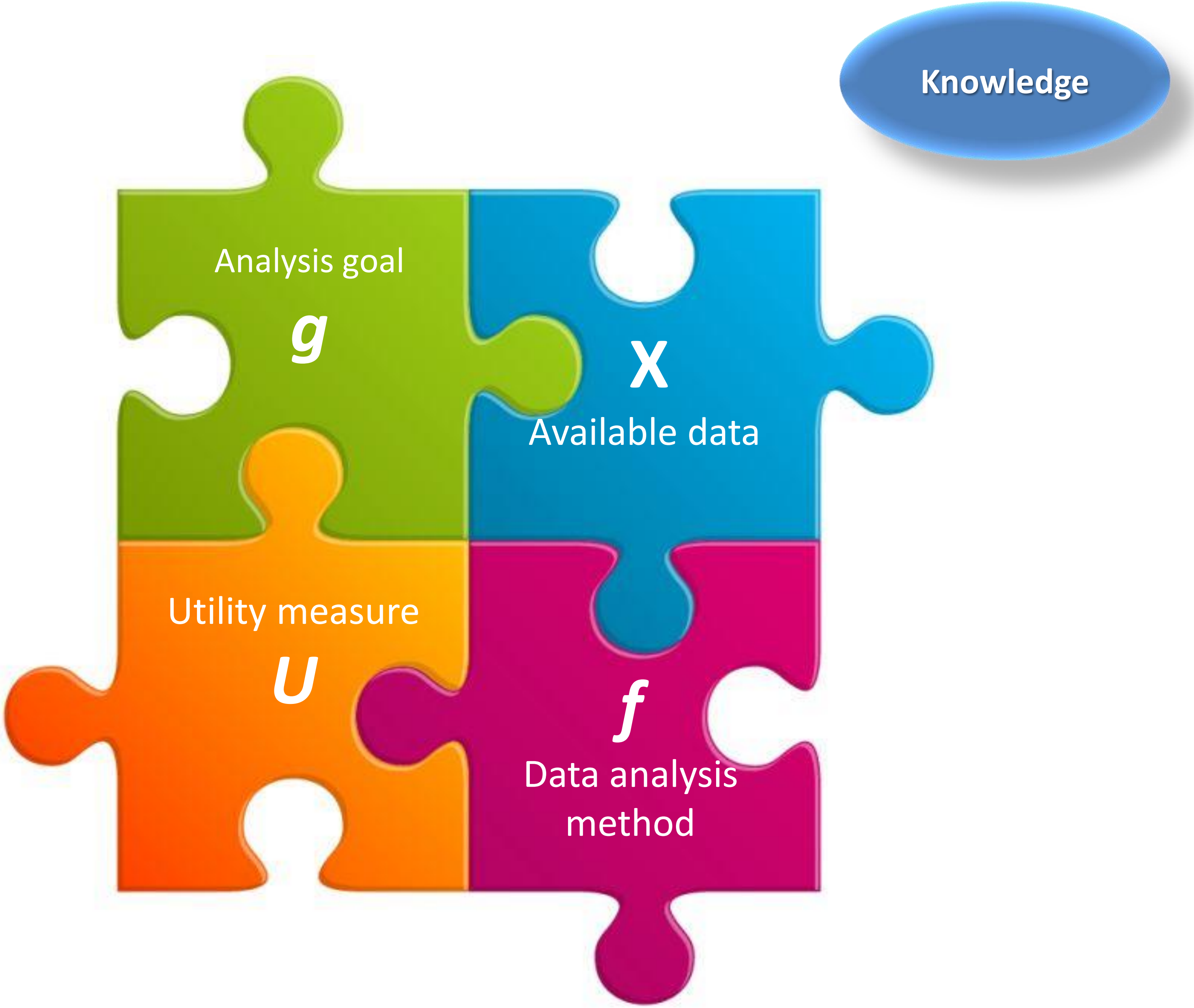
$$InfoQ(f,X,g) = U(\, f(X|g)\,)$$

Joint work with Galit Shmueli

Kenett, R.S. and Shmueli , G. (2014) On Information Quality, http://ssrn.com/abstract=1464444
Journal of the Royal Statistical Society, Series A (with discussion), 176(4).



$$InfoQ(f,X,g) = U(f(X|g))$$


Knowledge
Analysis goal
g
X
Available data
Utility measure
U
f
Data analysis method

Knowledge

Domain goal

What, why, when, where, how

-> Analysis goal

Explain, predict, describe
enumerative, analytic
exploratory, confirmatory

Quality of Goal Specification

- “error of the third kind” - giving the right answer to the wrong question – Kimball
- “Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise” - Tukey

Typical Goals of Customer Surveys

Goal 1. **Decide** where to launch improvement initiatives

Goal 2. **Highlight** drivers of overall satisfaction

Goal 3. **Detect** positive or negative trends in customer satisfaction

Goal 4. **Identify** best practices by comparing products

Goal 5. **Determine** strengths and weaknesses

Goal 6. **Set up** improvement goals

Goal 7. **Design** a balanced scorecard with customer inputs

Goal 8. **Communicate** the results using graphics

Goal 9. **Assess** the reliability of the questionnaire

Goal 10. **Improve** the questionnaire for future use

Knowledge

Data Source

- Primary, secondary
- Observational, experiment
- Single, multiple sources
- Collection instrument, protocol

Data Size and Dimension

- # observations
- # variables

Data Type

- Continuous, categorical, mix
- Structured, un-, semi-structured
- Cross-sectional, time series, panel, network, geographical

Data Quality *U(X|g)*

- "Zeroth Problem - How do the data relate to the problem, and what other data might be relevant?" - Mallows
- MIS/Database - usefulness of queried data to person querying it.
- *Quality of Statistical Data* (IMF, OECD) - usefulness of summary statistics for a particular goal (7 dimensions)

Knowledge

Statistical models and methods

- Parametric, semi-, non-parametric
- Classic, Bayesian

Data mining algorithms
Graphical methods
Operations research methods

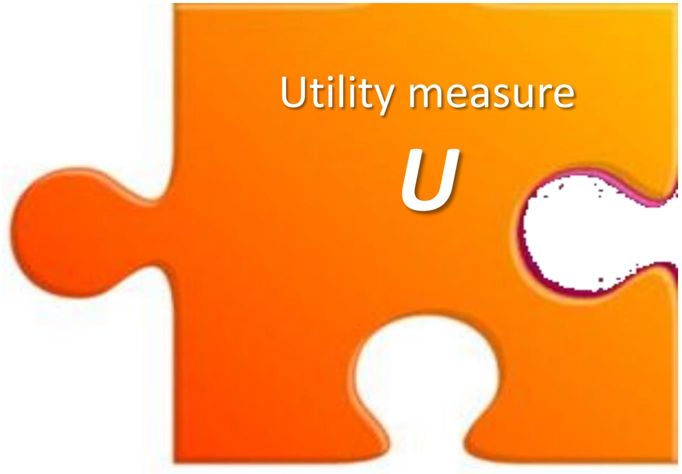
Analysis Quality

- "poor models and poor analysis techniques, or even analyzing the data in a totally incorrect way." - Godfrey
- Analyst expertise
- Software availability
- The focus of statistics education

Domain goal -> Analysis goal

Knowledge

- Predictive accuracy, lift
- Goodness-of-fit
- Statistical power, statistical significance
- Strength-of-fit
- Expected costs, gains
- Bias reduction, bias-variance tradeoff



Quality of Utility Measure

- Adequate metric from analysis standpoint (R^2, holdout data)
- Adequate metric from domain standpoint

Approaches for Increasing InfoQ

Knowledge

Study Design (Pre-Data)

- DOE
- Clinical trials
- Survey sampling
- Computer experiments

Randomization, Stratification, Blinding, Placebo, Blocking, Replication, Sampling frame, Link data collection protocol with appropriate design

Post-Data-Collection

- Data cleaning and preprocessing
- Re-weighting, bias adjustment
- Meta analysis

Recovering "real data" vs. "cleaning for the goal"
Handling missing values, outlier detection, re-weighting, combining results

Assessing InfoQ

Knowledge

InfoQ dimensions

1. Data resolution
2. Data structure
3. Data integration
4. Temporal relevance
5. Chronology of data and goal
6. Generalizability
7. Operationalization
8. Communication

"Quality of Statistical Data"
(Eurostat, OECD, NCSES,…)

- Relevance
- Accuracy
- Timeliness and punctuality
- Accessibility
- Interpretability
- Coherence
- Credibility

3 V's of *Big Data*

- Volume
- Variety
- Velocity

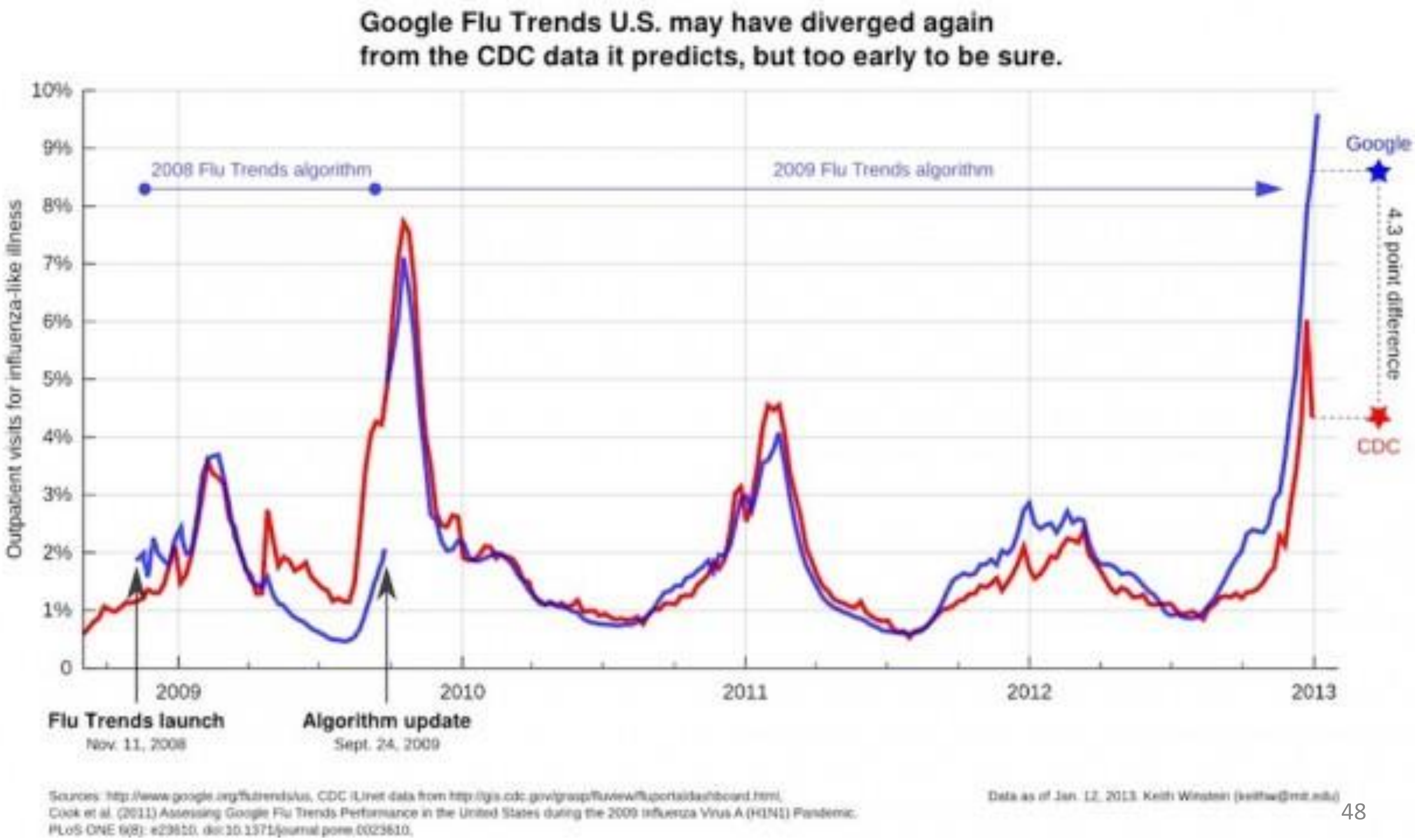
4 V's of *Big Data*

- Volume
- Variety
- Velocity
- Veracity

Marketing Research

- Recency
- Accuracy
- Availability
- Relevance

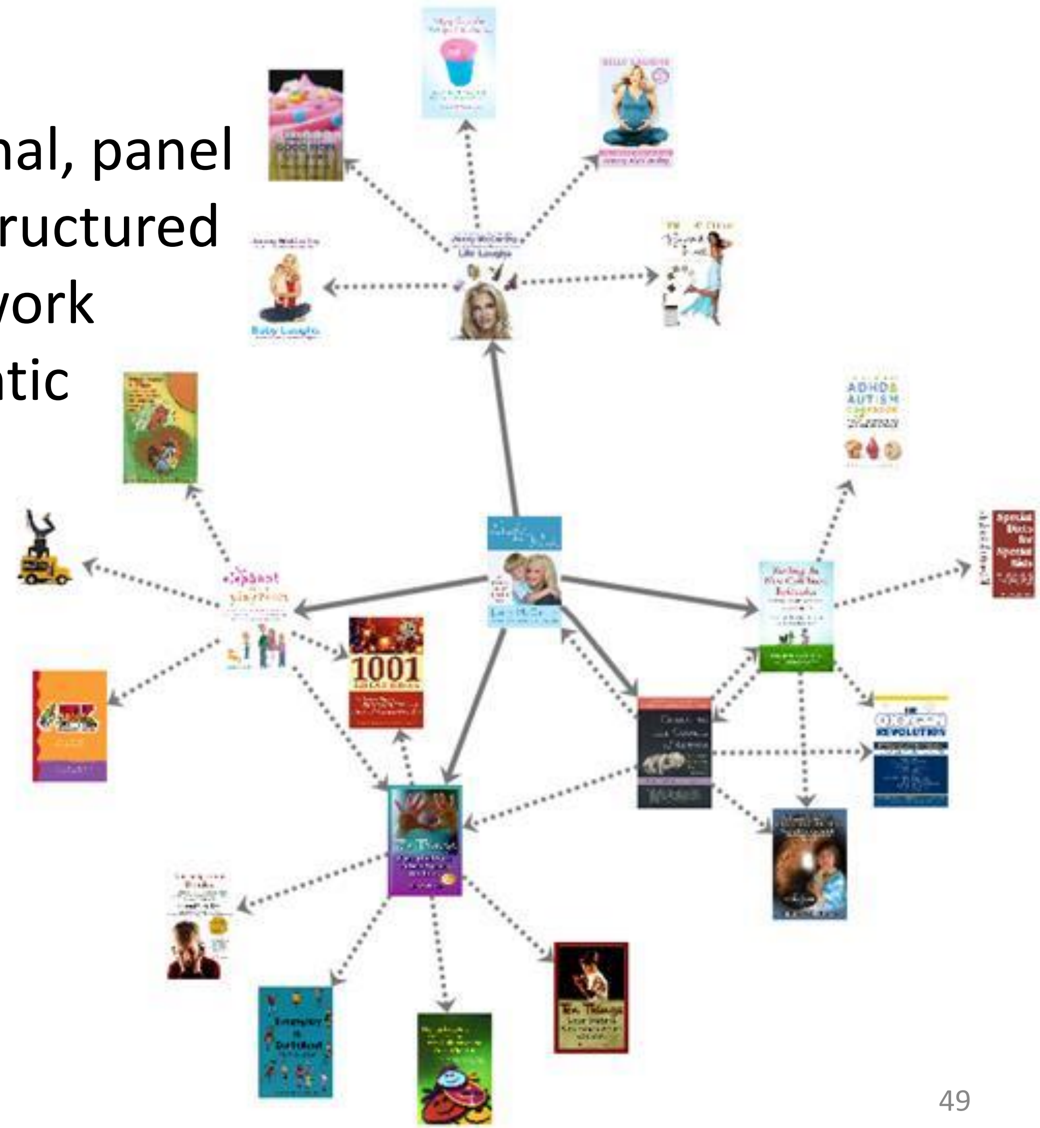
#1 Data Resolution



#2 Data Structure

Data Types

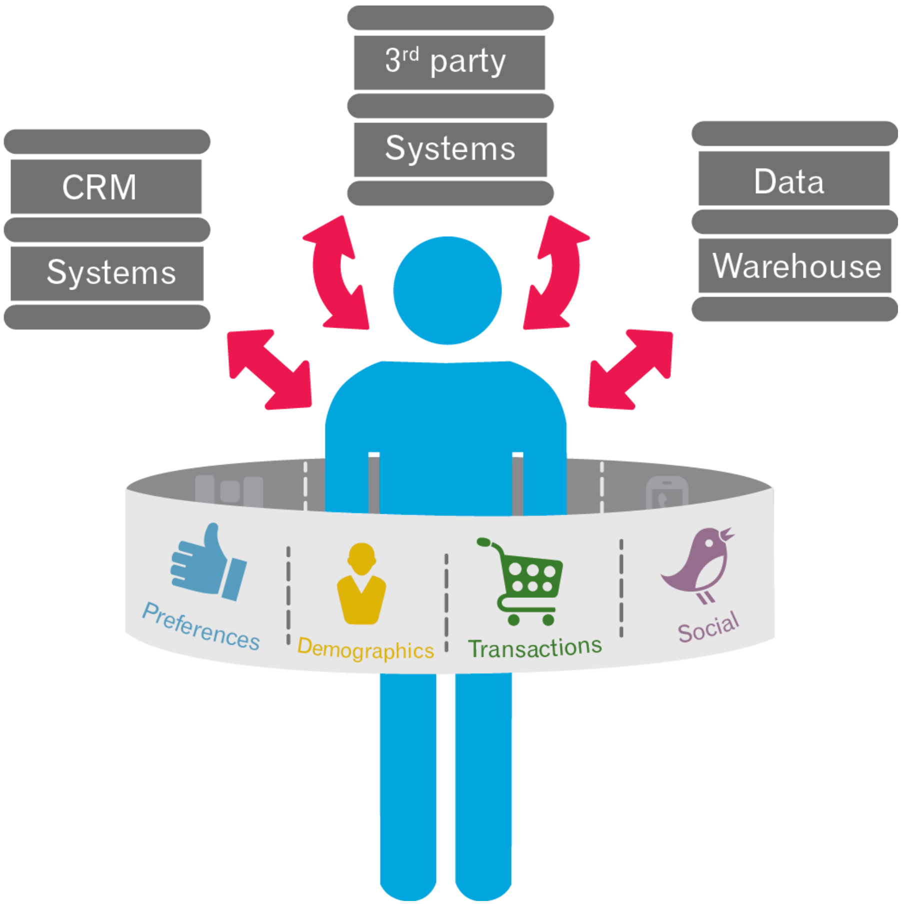
- Time series, cross-sectional, panel
- Structured, semi-, non-structured
- Geographic, spatial, network
- Text, audio, video, semantic
- Discrete, continuous



Data Characteristics

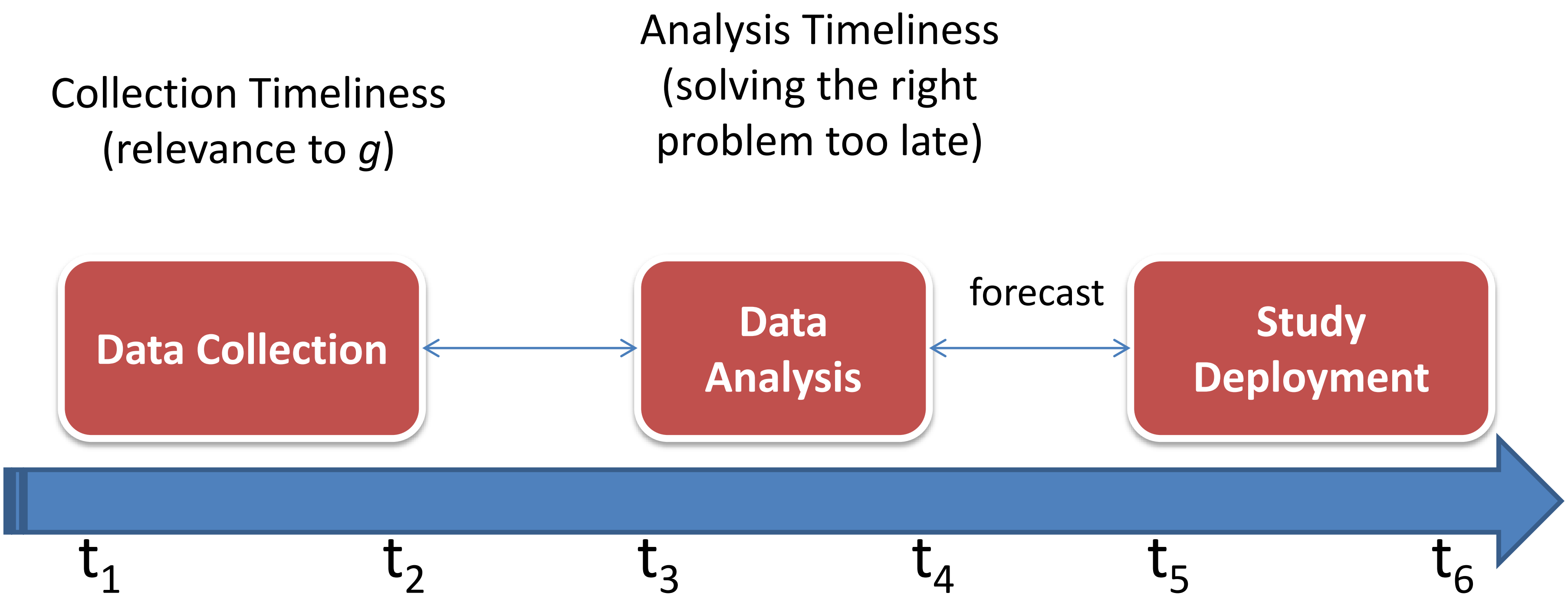
Corrupted and missing values due to study design or data collection mechanism

#3 Data Integration



Linkage, privacy-preserving methods: Increase or decrease InfoQ?

#4 Temporal Relevance



g: Prospective vs. retrospective; longitudinal vs. snapshot
Nature of X, complexity of *f*

#5 Chronology of Data & Goal

AQI
AIR QUALITY INDEX

Air Quality Index (AQI) Values	Levels of Health Concern
0 to 50	Good
51-100	Moderate
101-150	Unhealthy for Sensitive Groups
151-200	Unhealthy
201-300	Very Unhealthy
301 to 500	Hazardous

Data: Daily AQI in a city

g_1: Reverse-engineer AQI

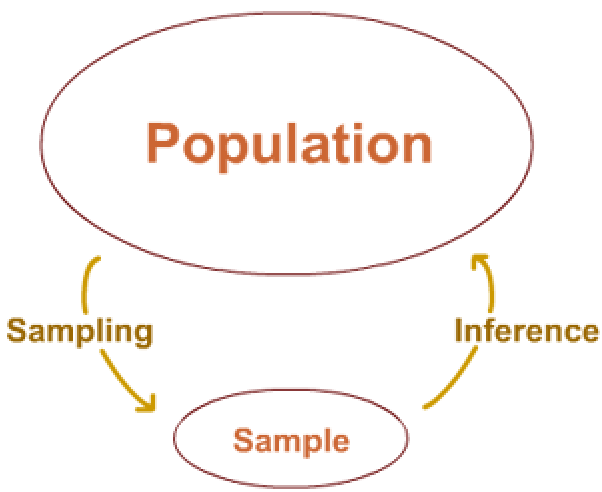
g_2: Forecast AQI

Retrospective/prospective
Ex-post availability
Endogeneity

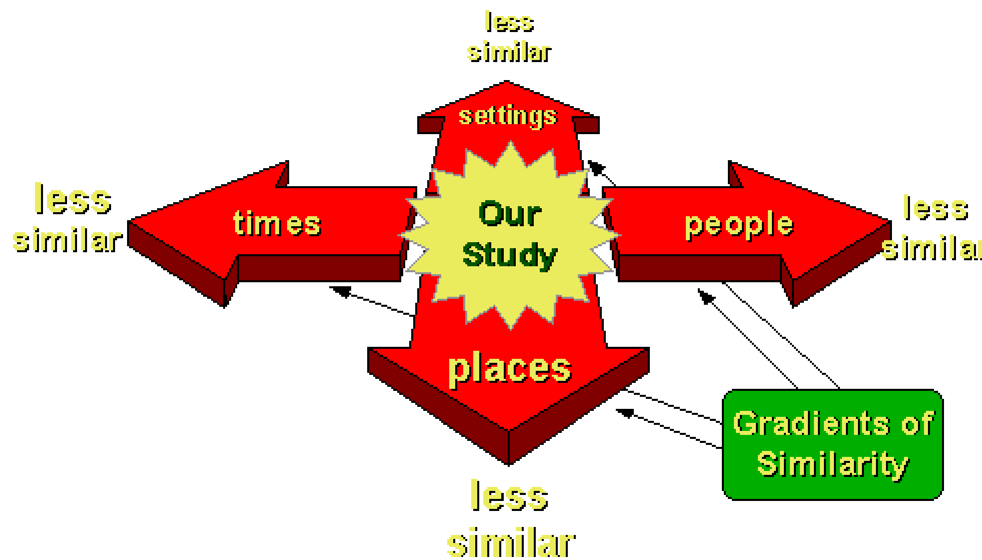
http://www.airnow.gov/?action=aqibasics.aqi

#6 Generalizability

Statistical generalizability



Scientific generalizability



Definition of *g*

Choice of X, *f*, *U*

#6 Generalizability

Judea Pearl stated that “Science is about generalization, and generalization requires ***transportability***. Conclusions that are obtained in a laboratory setting are transported and applied elsewhere, in an environment that differs in many aspects from that of the laboratory.”

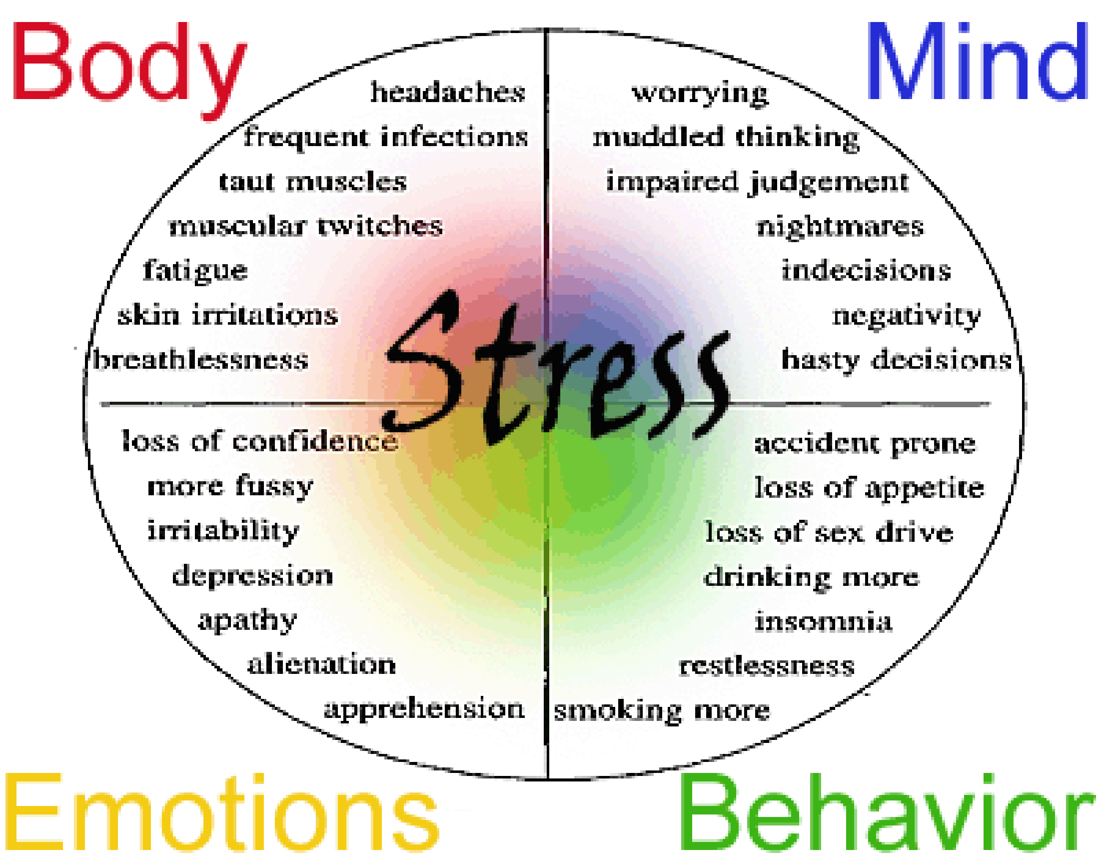
- Pearl, J. (2013), Transportability across studies: A formal approach, R-372

Georg Rasch used the term ***specific objectivity*** to describe that case essential to measurement in which "comparisons between individuals become independent of which particular instruments -- tests or items or other stimuli -- have been used. Symmetrically, it thought to be possible to compare stimuli belonging to the same class -- measuring the same thing -- independent of which particular individuals, within a class considered, were instrumental for comparison.“ The term ***general objectivity*** is reserved for the case in which absolute measures (i.e., amounts) are independent of which instrument (within a class considered) is employed, and no other object is required. By "absolute" we mean the measure "is not dependent on, or without reference to, anything else; not relative"

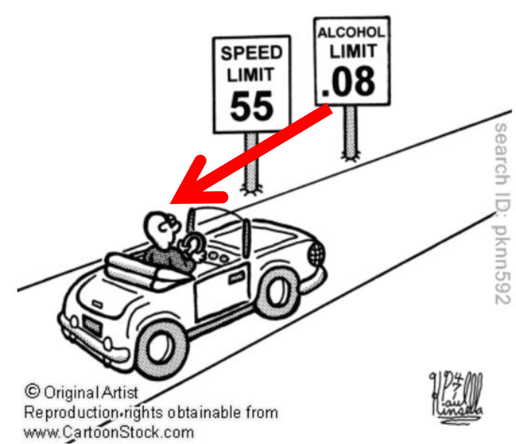
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology, pp. 321–334 in Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, IV. Berkeley: University of Chicago Press, 1980.
- Rasch, G. (1977). On Specific Objectivity: An attempt at formalizing the request for generality and validity of scientific statements. The Danish Yearbook of Philosophy, 14, 58-93.

#7 (Construct) Operationalization

χ construct
$X = \theta(\chi)$ operationalization (measurable)



- Causal explanation vs. prediction, description
- Theory vs. data
- Data: Questionnaire, physio measurement



#7 (Action) Operationalization

In the pre-publication drafts of *Quality, Productivity, and Competitive Position* Dr. Deming wrote:

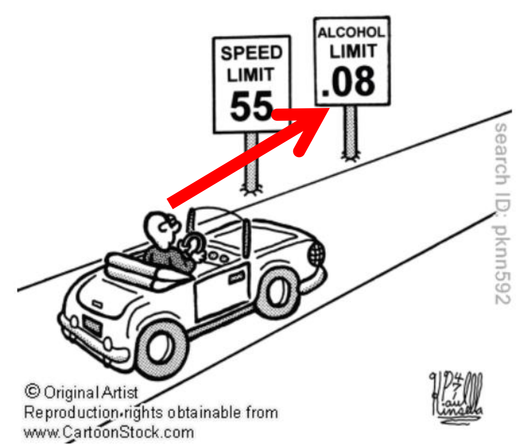
> "An operational definition consists of (1) a criterion to be applied to an object or a group of objects, (2) a test of compliance for the object or group, and (3) a decision rule for interpreting the test results as to whether the object or group is, or is not, in compliance."

In Dr. Deming's own conversations, when individuals would start telling him about what they or their organization were planning to do, he would invariably have one of two responses for them: "By what method?" or "How will you know?" Either one of these questions would generally end the conversation since the individual would have no answer. After discerning this pattern to Dr. Deming's responses, it finally occurred to me that these two questions corresponded to the last two parts of an operational definition. This realization, in turn, resulted in a generalization of an operational definition to become:

(1) What do you want to accomplish?

(2) By what method will you accomplish it?

(3) How will you know when you have accomplished it?

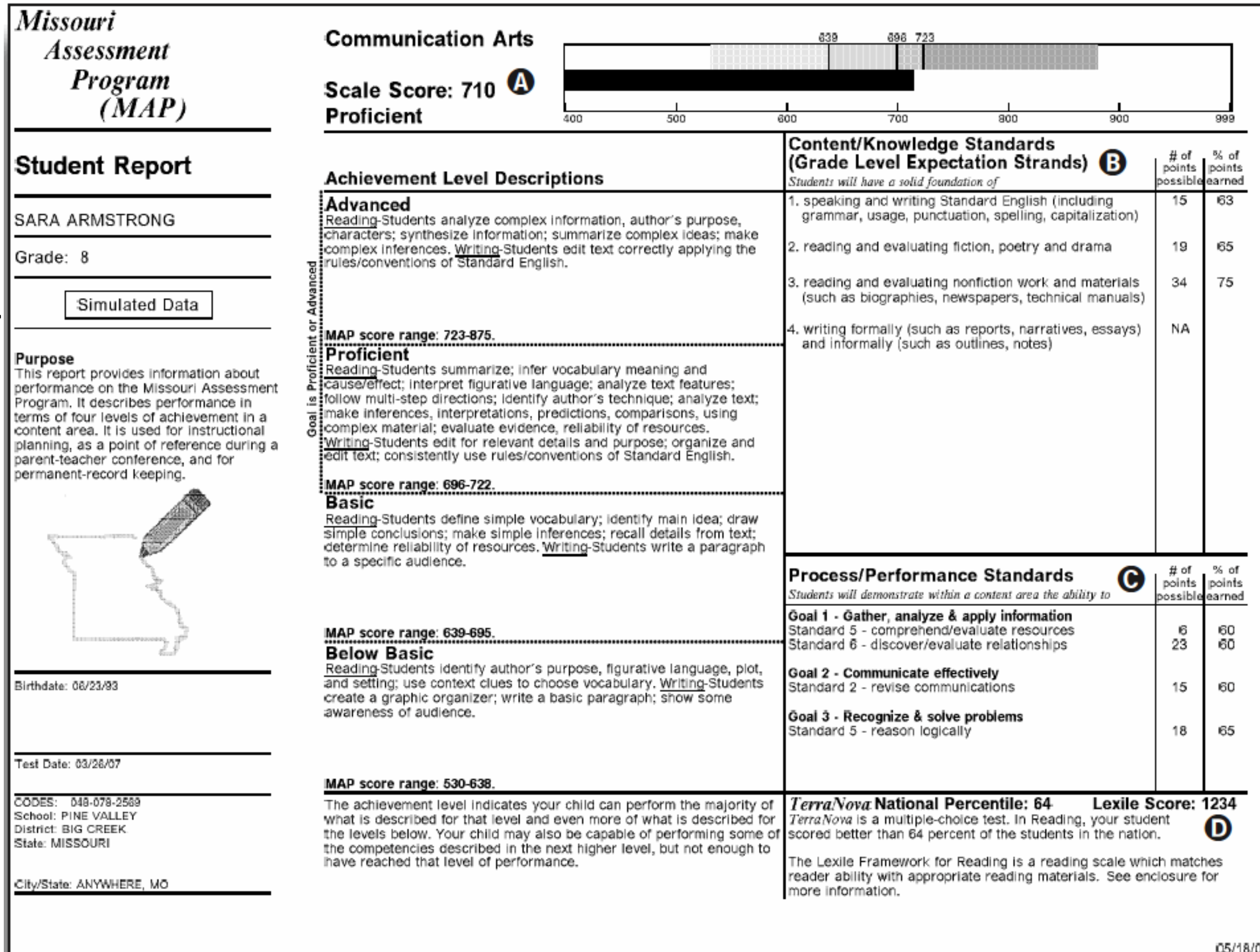


http://www.spcpress.com/pdf/DJW187.pdf

#7 Operationalization

National Education Goals Panel (NEGP) recommended that states answer four questions on their student reports:

1. How did my child do?
2. What types of skills or knowledge does his or her performance reflect?
3. How did my child perform in comparison to other students in the school, district, state, and, if available, the nation?
4. What can I do to help my child improve?



Missouri Assessment Program (MAP)

Student Report

SARA ARMSTRONG

Grade: 8

Simulated Data

Purpose
This report provides information about performance on the Missouri Assessment Program. It describes performance in terms of four levels of achievement in a content area. It is used for instructional planning, as a point of reference during a parent-teacher conference, and for permanent-record keeping.

Birthdate: 08/23/93

Test Date: 03/26/07

CODES: 048-078-2569
School: PINE VALLEY
District: BIG CREEK
State: MISSOURI

City/State: ANYWHERE, MO

Communication Arts

Scale Score: 710 Ⓐ
Proficient

639 696 723
400 500 600 700 800 900 999

Achievement Level Descriptions

Goal is Proficient or Advanced

Advanced
Reading-Students analyze complex information, author's purpose, characters; synthesize information; summarize complex ideas; make complex inferences. Writing-Students edit text correctly applying the rules/conventions of Standard English.

MAP score range: 723-875.

Proficient
Reading-Students summarize; infer vocabulary meaning and cause/effect; interpret figurative language; analyze text features; follow multi-step directions; identify author's technique; analyze text; make inferences, interpretations, predictions, comparisons, using complex material; evaluate evidence, reliability of resources. Writing-Students edit for relevant details and purpose; organize and edit text; consistently use rules/conventions of Standard English.

MAP score range: 696-722.

Basic
Reading-Students define simple vocabulary; identify main idea; draw simple conclusions; make simple inferences; recall details from text; determine reliability of resources. Writing-Students write a paragraph to a specific audience.

MAP score range: 639-695.

Below Basic
Reading-Students identify author's purpose, figurative language, plot, and setting; use context clues to choose vocabulary. Writing-Students create a graphic organizer; write a basic paragraph; show some awareness of audience.

MAP score range: 530-638.

The achievement level indicates your child can perform the majority of what is described for that level and even more of what is described for the levels below. Your child may also be capable of performing some of the competencies described in the next higher level, but not enough to have reached that level of performance.

Content/Knowledge Standards (Grade Level Expectation Strands) Ⓑ *Students will have a solid foundation of*	# of points possible	% of points earned
1. speaking and writing Standard English (including grammar, usage, punctuation, spelling, capitalization)	15	63
2. reading and evaluating fiction, poetry and drama	19	65
3. reading and evaluating nonfiction work and materials (such as biographies, newspapers, technical manuals)	34	75
4. writing formally (such as reports, narratives, essays) and informally (such as outlines, notes)	NA	

Process/Performance Standards Ⓒ *Students will demonstrate within a content area the ability to*	# of points possible	% of points earned
Goal 1 - Gather, analyze & apply information		
Standard 5 - comprehend/evaluate resources	6	60
Standard 6 - discover/evaluate relationships	23	60
Goal 2 - Communicate effectively		
Standard 2 - revise communications	15	60
Goal 3 - Recognize & solve problems		
Standard 5 - reason logically	18	65

***TerraNova* National Percentile: 64** **Lexile Score: 1234** Ⓓ
TerraNova is a multiple-choice test. In Reading, your student scored better than 64 percent of the students in the nation.

The Lexile Framework for Reading is a reading scale which matches reader ability with appropriate reading materials. See enclosure for more information.

05/18/07

Goodman , D. and. Hambleton , R.(2004). Student Test Score Reports and Interpretive Guides: Review of Current Practices and Suggestions for Future Research, *Applied Measurement in Education*, 17:2, 145-220

#7 Operationalization

http://sat.collegeboard.org/practice/sat-skills-insight/writing/band/200

UNDERSTANDING THE MAP | CRITICAL READING | MATHEMATICS | WRITING | PRINT

200-300 | 300-390 | 400-490 | 500-590 | 600-690 | 700-800

Using the Skills Map

CRITICAL READING SKILL GROUPS: 1 2 3 4 5

PREVIOUS

1. Determining the Meaning of Words

NEXT

Academic Skills

A typical student in this score band can do the following:

- **SKILL 1:** Use the *context* of a sentence or larger section of text to determine the meaning of unknown words or to differentiate among multiple possible meanings of words.
- **SKILL 2:** Understand how *syntax* (the arrangement of words and phrases in a sentence) influences the relationship among words and ideas within a sentence.
- **SKILL 3:** Demonstrate increased *comprehension* of specialized vocabulary.

Suggestions for Improvement

To advance to a higher score band, focus on the following skills:

- As you read a text about a topic with which you are unfamiliar, look for words that you know to help you determine what any unknown words might mean.
- When you encounter an unknown word or difficult word in your reading, look it up in a dictionary that provides information on the word's origins and history.
- When you encounter a difficult section of text in your reading, break down the ideas in it sentence by sentence and even within sentences. Think about how the ideas work together.

Skill Examples

1 2 3 4 5 6 7 8

#8 Communication

1992 NAEP Executive Summary Report

National Overall Average Mathematics Proficiency and Achievement Levels, Grades 4, 8, and 12

2 16 43

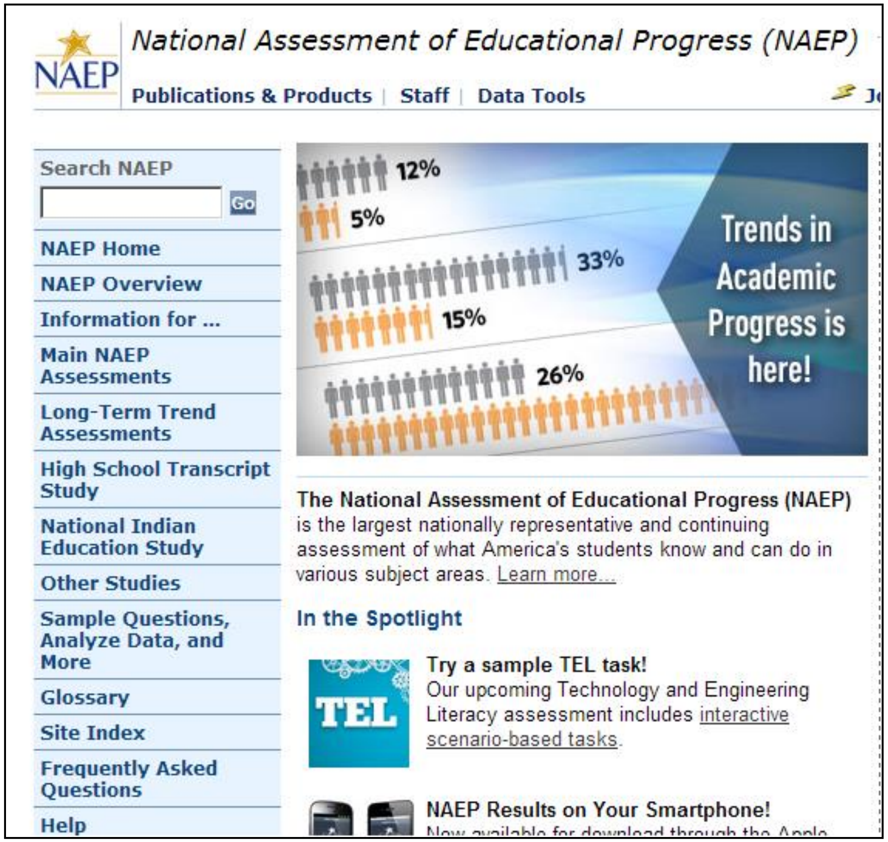
			Percentage of Students At or Above			
Grades	**Assessment Years**	**Average Proficiency**	**Advanced**	**Proficient**	**Basic**	**Percentage Below Basic**
4	1992 1990	218(0.7)> 213(0.9)	2(0.3) 1(0.4)	18(1.0)> 13(1.1)	61(1.0)> 54(1.4)	39(1.0)< 46(1.4)
8	1992 1990	268(0.9)> 263(1.3)	4(0.4) 2(0.4)	25(1.0)> 20(1.1)	63(1.1)> 58(1.4)	37(1.1)< 42(1.4)
12	1992 1990	299(0.9)> 294(1.1)	2(0.3) 2(0.3)	16(0.9) 13(1.0)	64(1.2)> 59(1.5)	36(1.2)< 41(1.5)

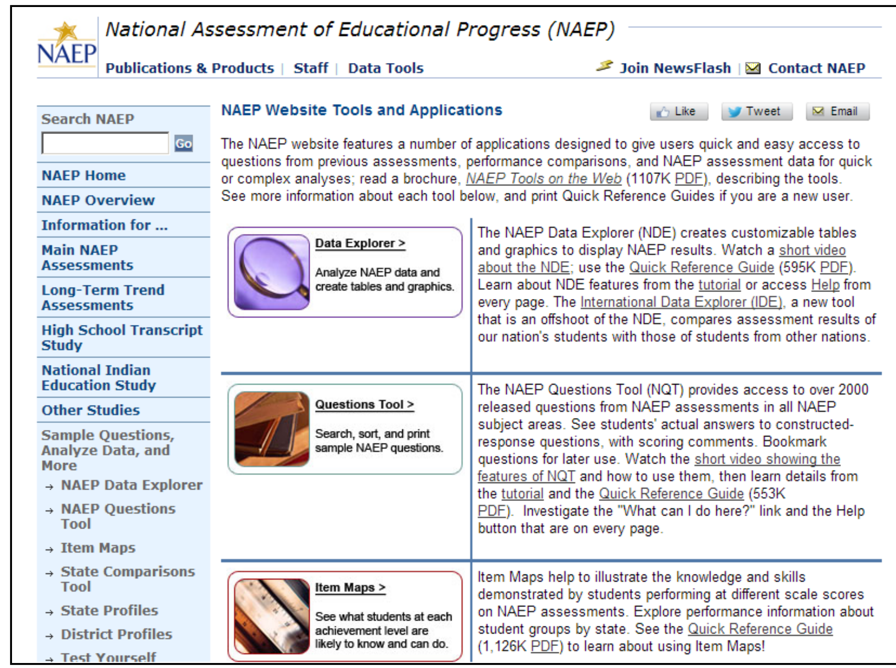
> The value for 1992 was significantly higher than the value for 1990 at about the 95 percent confidence level. < The value for 1992 was significantly lower than the value for 1990 at about the 95 percent confidence level. The standard errors of the estimated percentages and proficiencies appear in parentheses. It can be said with 95 percent confidence that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix for details).

When asked what the 18% in line 1 meant,
53% of the policy makers responded incorrectly

1 2 3 4 5 6 7 8

#8 Communication

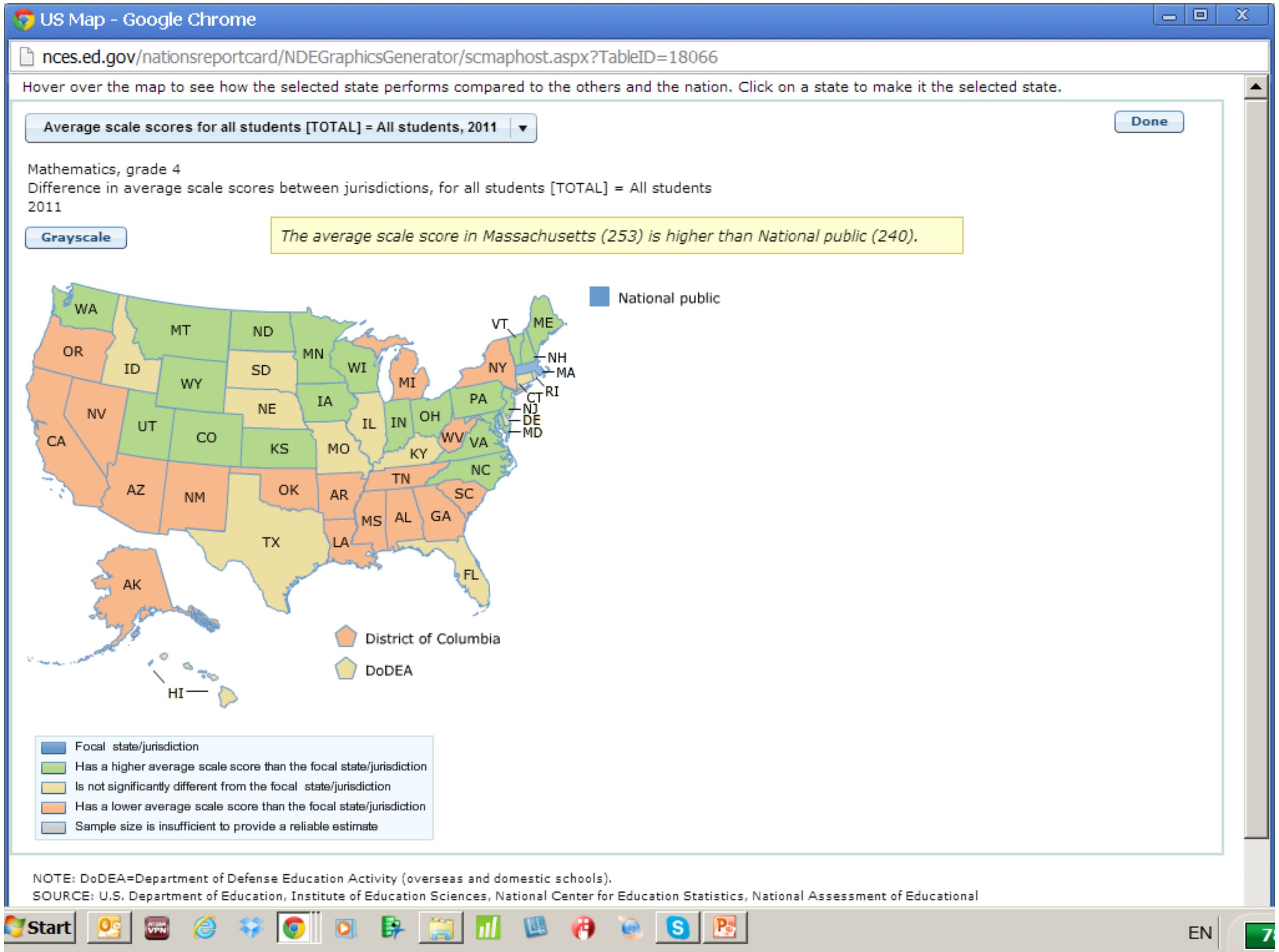




http://nces.ed.gov/nationsreportcard/itemmaps/index.asp

1 2 3 4 5 6 7 8

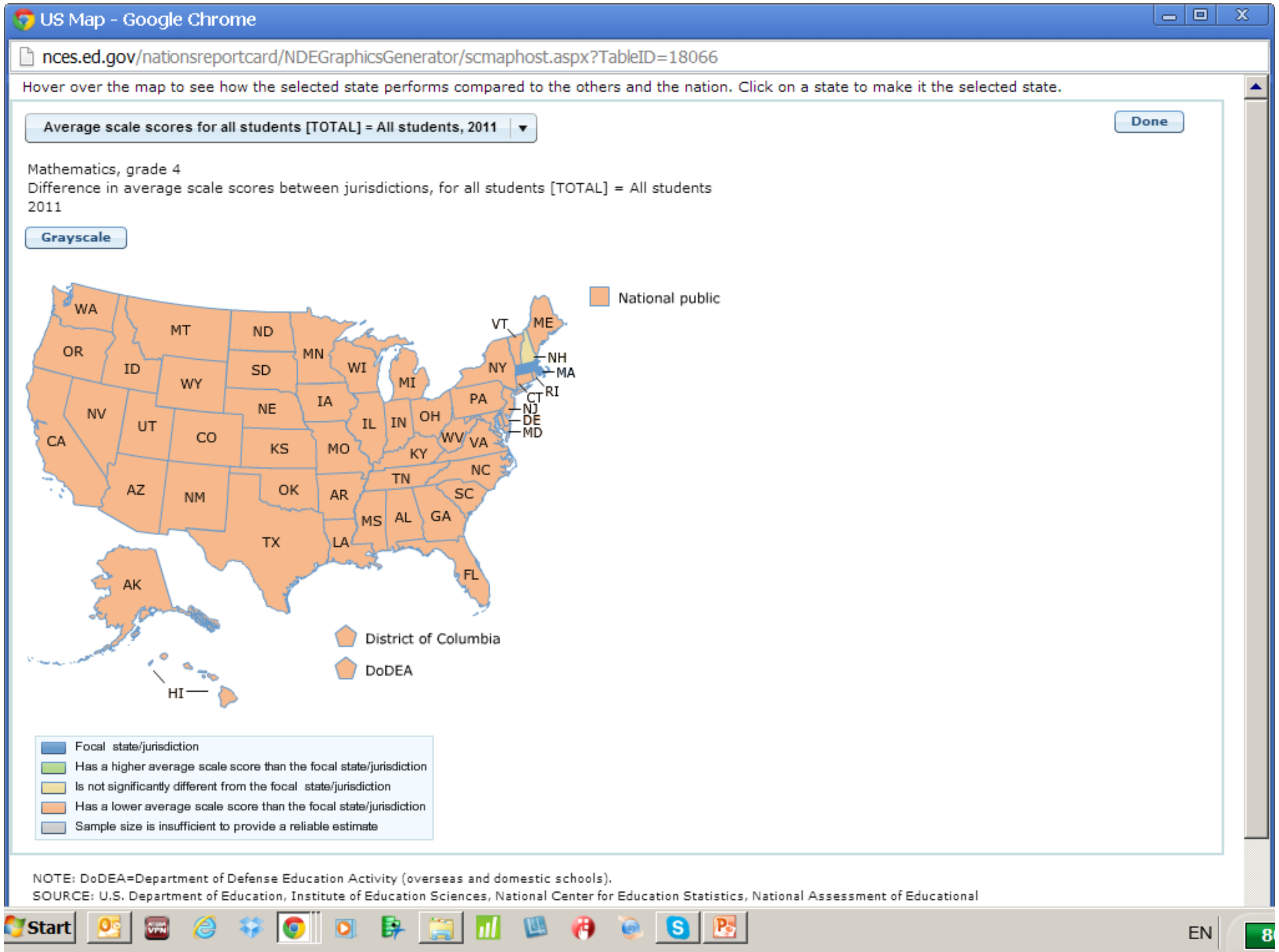
#8 Communication



http://nces.ed.gov/nationsreportcard/itemmaps/index.asp

1 2 3 4 5 6 7 8

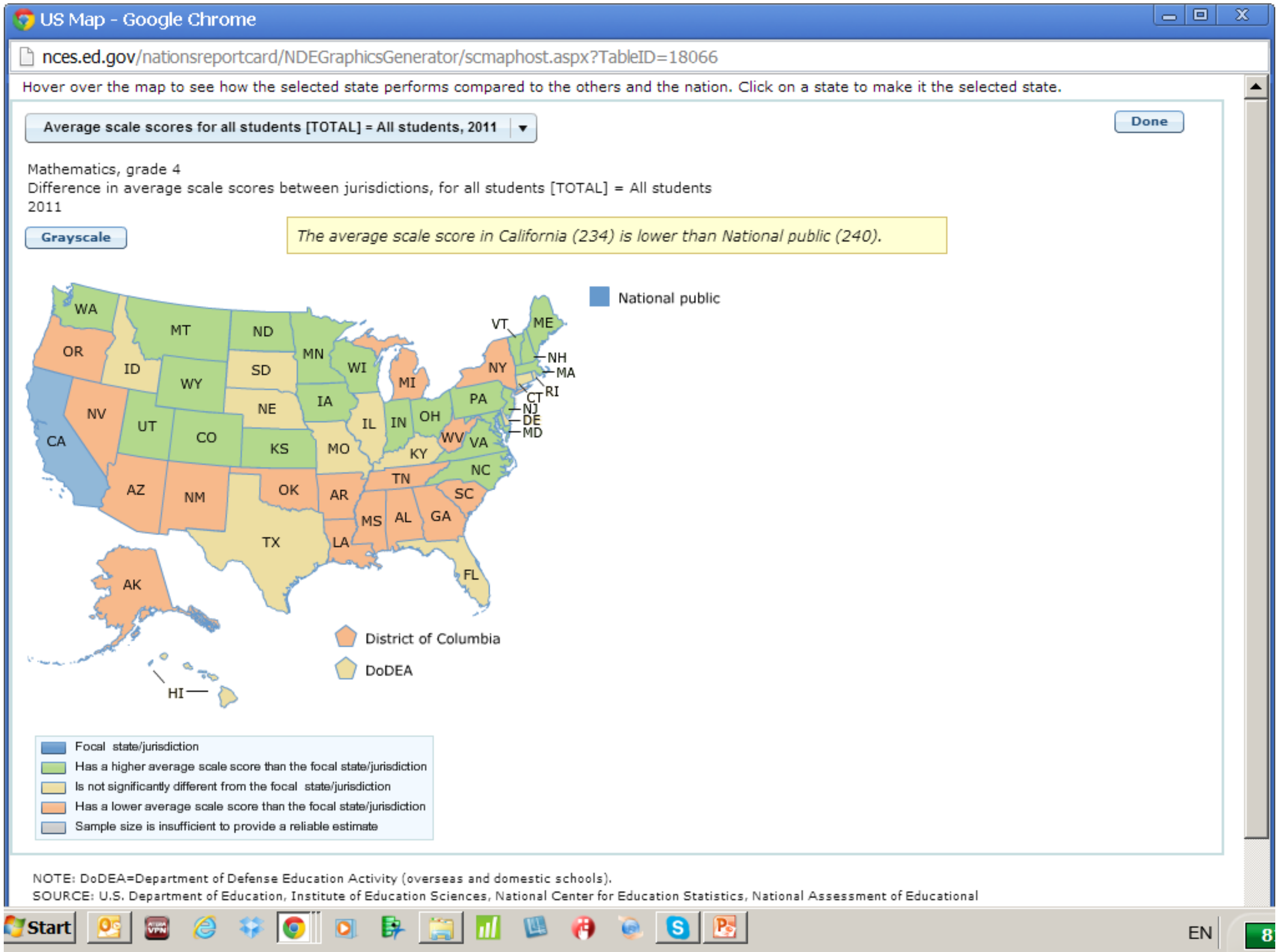
#8 Communication



http://nces.ed.gov/nationsreportcard/itemmaps/index.asp

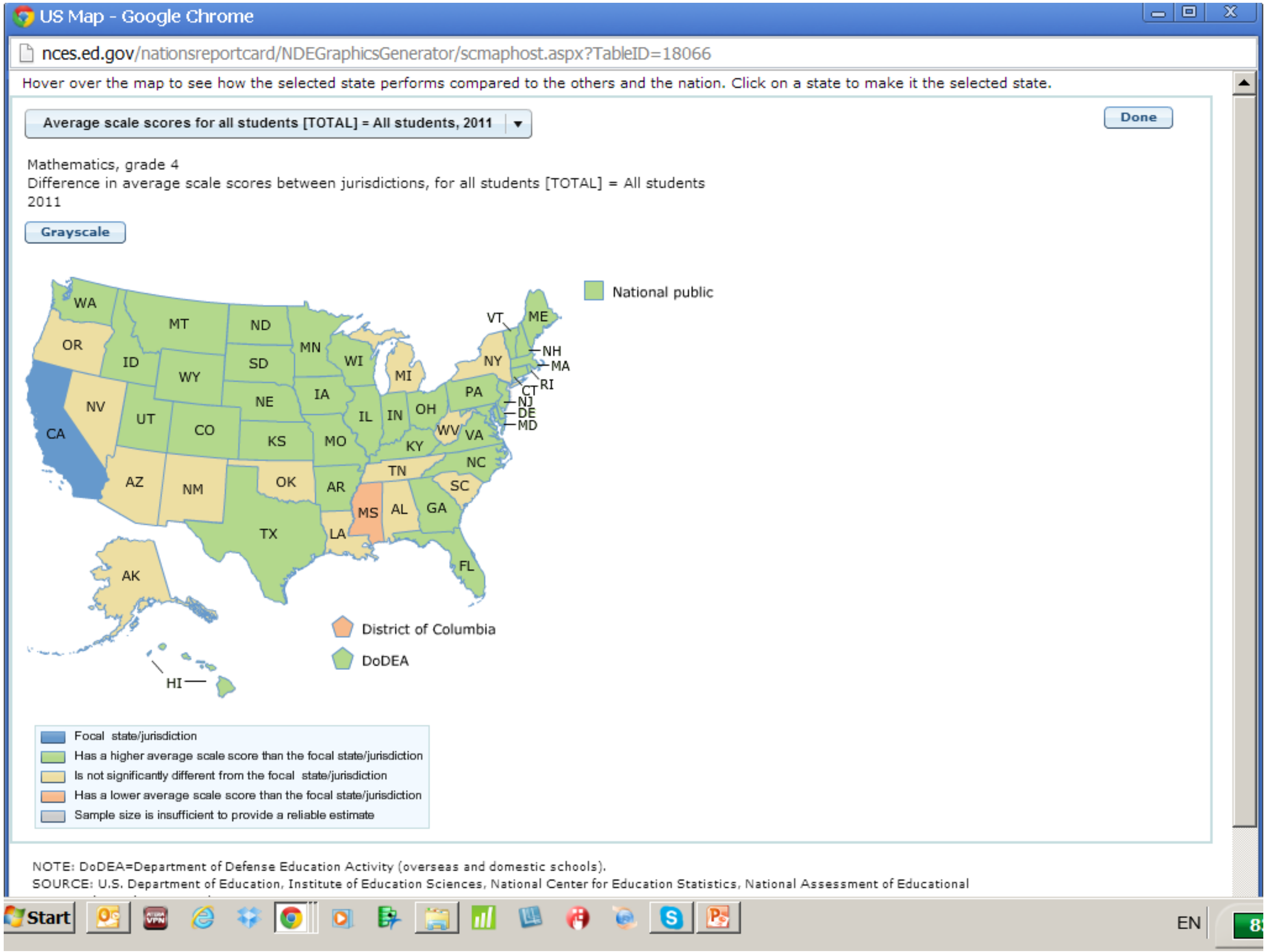
1 2 3 4 5 6 7 8

#8 Communication



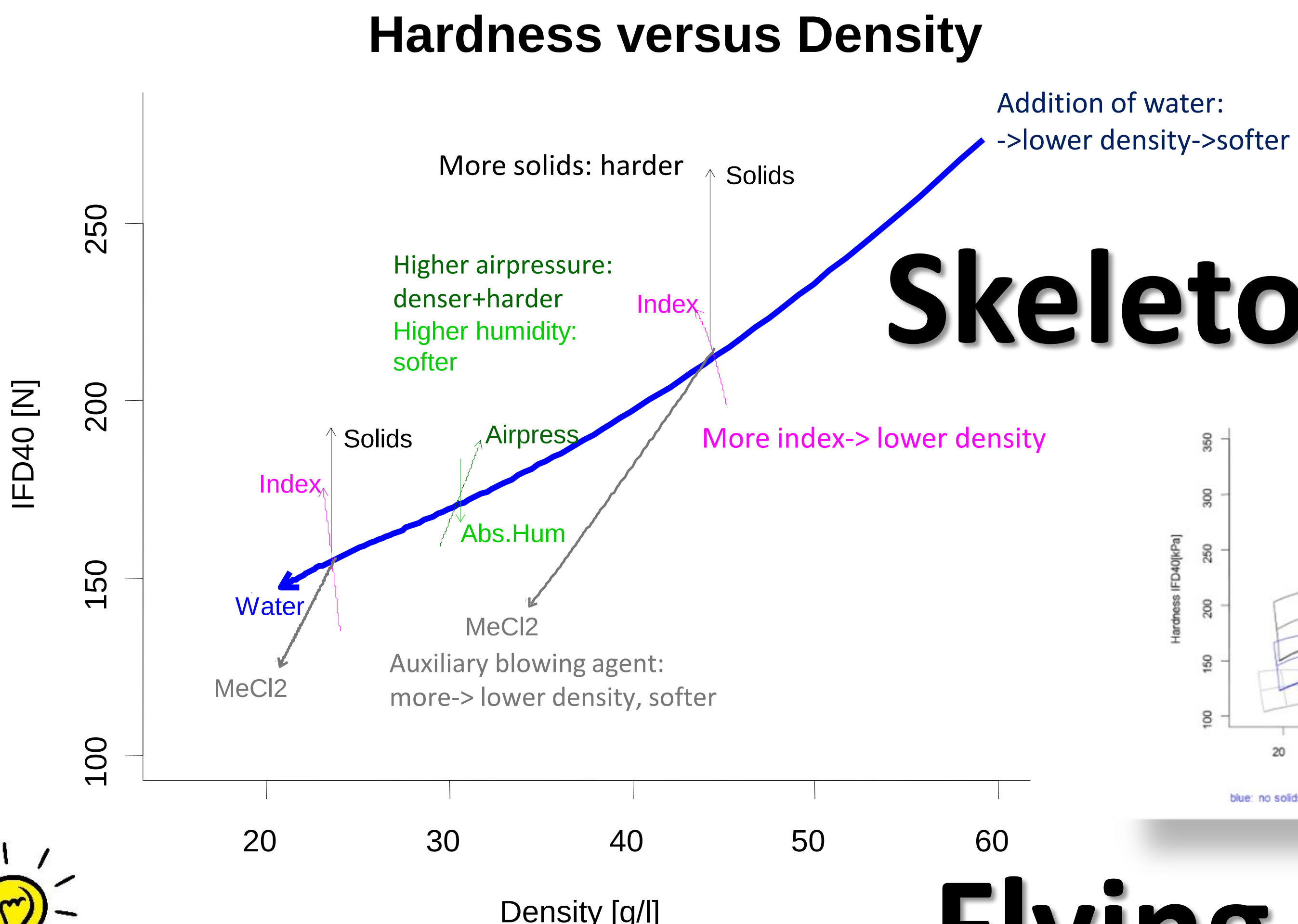
http://nces.ed.gov/nationsreportcard/itemmaps/index.asp

#8 Communication

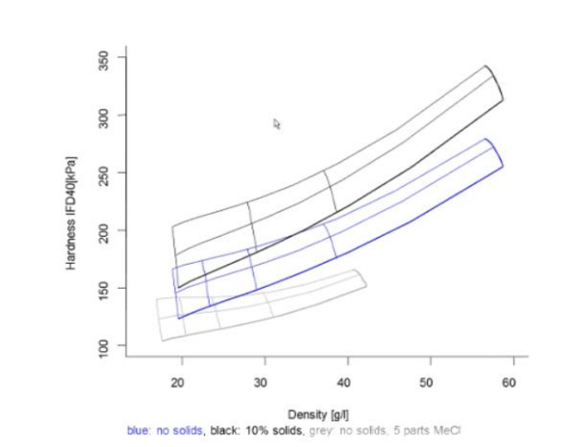


http://nces.ed.gov/nationsreportcard/itemmaps/index.asp

#8 Communication



Skeletons on



Flying Carpets

Christian Ritter

Assessing InfoQ in Practice

Rating-based assessment

1-5 scale on each dimension:

#	Dimension	Note	Value	Index
1	Data resolution		5	1.0000
2	Data structure		4	0.7500
3	Data integration		5	1.0000
4	Temporal relevance		5	1.0000
5	Generalizability		3	0.5000
6	Chronology of data and goal		5	1.0000
7	Concept operationalization		2	0.2500
8	Communication		3	0.5000
	InfoQ Score = 0.68			

$$InfoQ\ Score = [d_1(Y_1)\ \ d_2(Y_2)\ \ \ldots\ \ d_8(Y_8)]^{1/8}$$

Experience from two research methods courses

- Preparing a PhD research proposal (U Ljubljana, 50 students, goo.gl/f6bIA)
- Post-hoc evaluation of five completed studies (CMU, 16 students, goo.gl/erNPF)

Knowledge

Information Quality

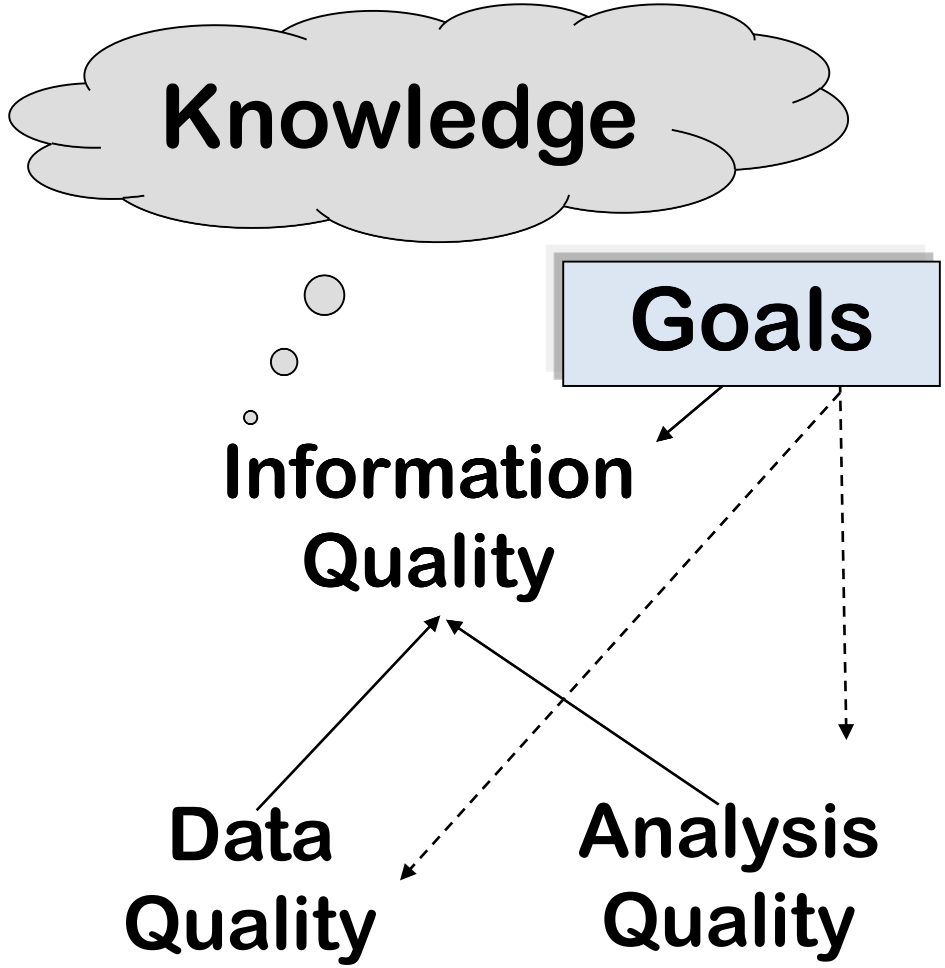
$$InfoQ(f,X,g) = U(f(X|g))$$

g	A specific analysis goal
X	The available dataset
f	An empirical analysis method
U	A utility measure

What

How

1. Data resolution
2. Data structure
3. Data integration
4. Temporal relevance
5. Chronology of data and goal
6. Generalizability
7. Operationalization
8. Communication



Primary Data
- Experimental
- Observational

Secondary Data
- Experimental
- Observational

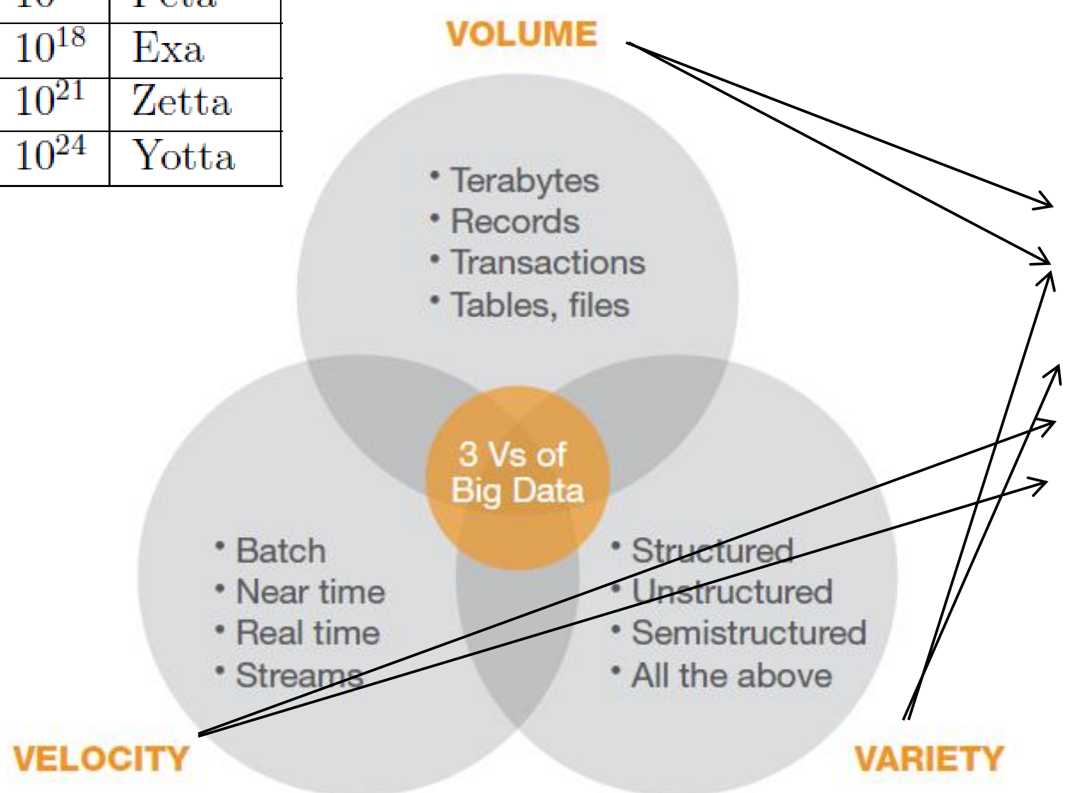
Big Data

Knowledge

Big Data Analytics

Power	Prefix
10^9	Giga
10^{12}	Tera
10^{15}	Peta
10^{18}	Exa
10^{21}	Zetta
10^{24}	Yotta

V V V



1. Data resolution
2. Data structure
3. Data integration
4. Temporal relevance
5. Chronology of data and goal
6. Generalizability
7. Operationalization
8. Communication

InfoQ

Russom, P., Big Data Analytics, TDWI Best Practices Report, Q4 2011

The Theory of Applied Statistics

1. Background

1.1. Statistics as a Mathematical Discipline

1.2. The Role of Case Studies in the Development of Statistics

1.3. Main Achievements in 100 Years of Statistics

1.4. New Challenges

2. Applied Statistics as a Discipline: Some Examples

2.1. Surveys

2.2. Clinical Trials

2.3. Industrial Statistics

2.4. Quality and Reliability

2.5. Risk Analysis

Kenett, R.S. (2012) A Note on the Theory of Applied Statistics . Available at SSRN: http://ssrn.com/abstract=2171179

The Theory of Applied Statistics

3. Tools of Applied Statistics
3.1. Cognitive Science and Psychology
3.2. Concept Science and Knowledge Management
3.3. Visualization Methods, Static and Dynamic
3.4. ETL and Data warehouses
3.5. Ontologies and Unstructured Data
3.6. Statistics in Management Science and Computer Science
4. Towards a Theory of Applied Statistics
4.1. Problem Elicitation
4.2. Communicating with other Disciplines
4.3. Formulation and Presentation of Findings
4.4. Education of Statistical Concepts (not techniques)
4.5. Evaluating Impact (Practical Statistical Efficiency)
4.6. Evaluating Value Added (Information Quality)
4.7. Designing a Strategy for Expanding the Role of Statistics (The Statistical Efficiency Conjecture and Integrated Models).

Who, how and what

Who is doing it

Certified Analytics Professional (CAP™)

BENEFITS OF CERTIFICATION

- Advances your career potential by setting you apart from the competition
- Drives personal satisfaction of accomplishing a key career milestone
- Helps improve your overall job performance by stressing continuing professional development
- Recognizes that you have invested in your analytics career by pursuing this rigorous credential
- Boosts your salary potential by being viewed as experienced analytics professional
- Shows competence in the principles and practices of analytics

DOMAINS OF ANALYTICS PRACTICE

Domain	Description	Weight*
I	Business Problem (Question) Framing	15%
II	Analytics Problem Framing	17%
III	Data	22%
IV	Methodology (Approach) Selection	15%
V	Model Building	16%
VI	Deployment	9%
VII	Life Cycle Management	6%
	**Percentage of questions in exam*	100%

Who is not doing it

Guidelines for Voluntary Professional Accreditation
by the American Statistical Association
4/16/10 (revised 2/1/11)

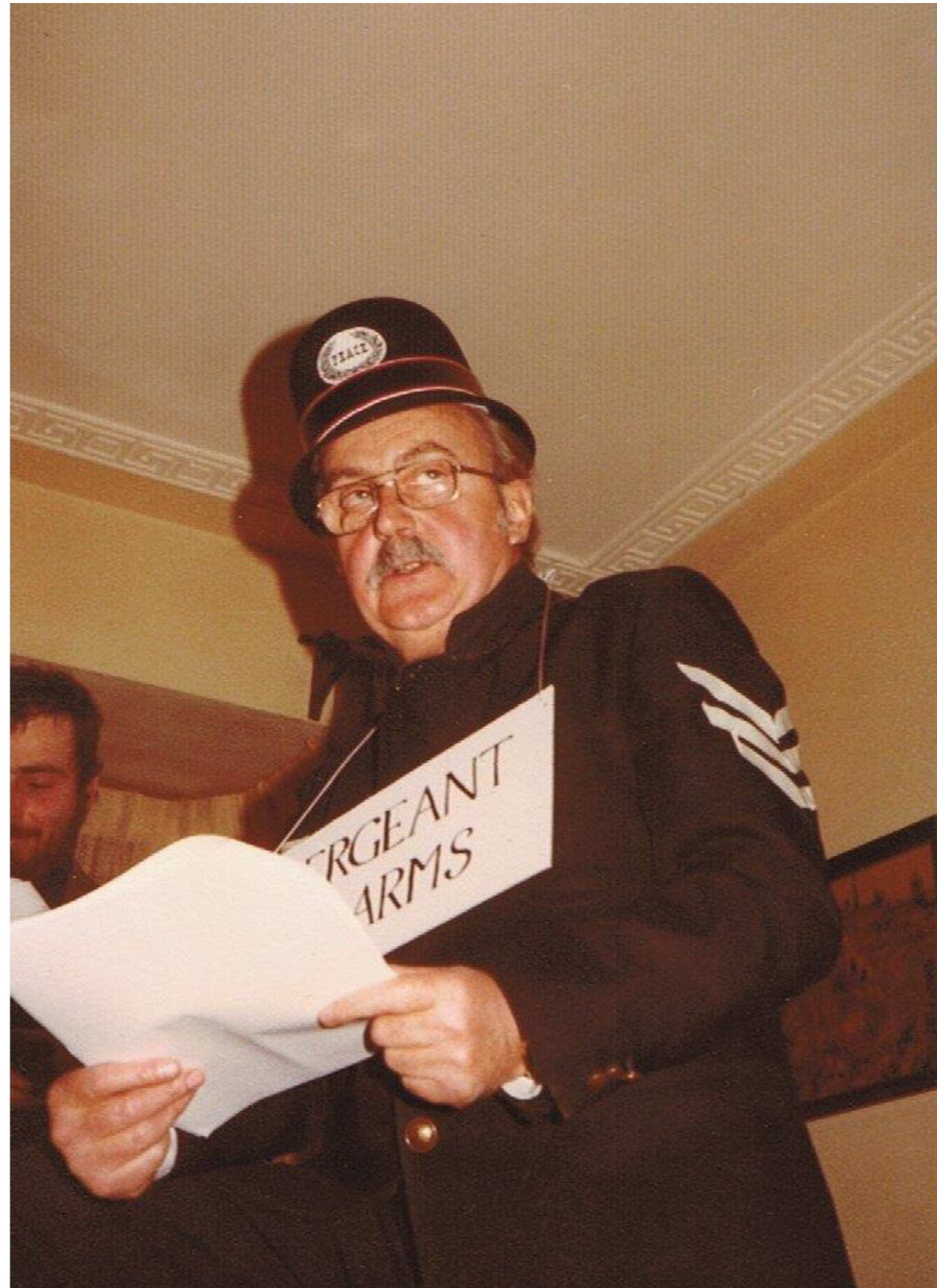
1. Introduction

This document, approved by the ASA Board of Directors on April 16, 2010, provides the framework for voluntary professional accreditation. Additional details for applicants are found in "Instructions for Applicants' document maintained on the ASA accreditation website.

PStat® accreditation is offered by the American Statistical Association as a service to those of its members who find added value in a voluntarily obtained credential that provides recognition by peers that they have statistical training and knowledge, have experience in applying that expertise competently, maintain appropriate professional development, agree to abide by ethical standards of practice, and are able to communicate effectively. Not all statisticians will need or seek PStat® accreditation, and the lack of PStat® accreditation should never be construed by itself as evidence of lack of education, expertise, or competence as a statistician. However, holders of the PStat® credential have voluntarily applied for this status, have submitted materials that have been carefully reviewed by peers and found to be deserving of the credential, and must periodically undergo further review to maintain this status.

A Role Model

Xmas
party
1979



Experiment by Cole Porter
performed by Mabel Mercer, with Cy Walter and Stan Freeman

Before you leave these portals
to meet less fortunate mortals,
there's just one final message I would give to you .
You all have learned reliance
on the sacred teachings of science,
so I hope through life you never will decline,
in spite of philistine defiance,
to do what all good scientists do.

Experiment,
Make it your motto day and night.
Experiment,
And it will lead you to the light.
The apple on the top of the tree
is never too high to achieve.
So take an example from Eve
Experiment.

Be curious,
Though interfering friends may frown.
Get furious,
At each attempt to hold you down.
If this advice you'll only employ,
the future can offer you infinite joy
and merriment.
Experiment,
and you'll see.

Some key lessons

Statistics needs interactions with other disciplines

Good problems drive good Statistics

Teaching Statistics requires continuous investments in the learning environment

Fun should be part of doing and learning Statistics

Ask customers to assess the quality of your work

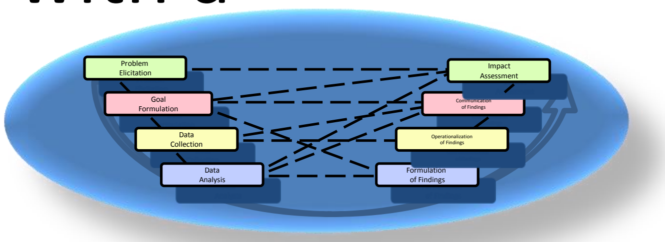
Statistical Engineering

Definition (Hoerl and Snee): The study of how to best utilize statistical concepts, methods, and tools and integrate them with information technology and other relevant sciences to generate improved results.

The **NIST Statistical Engineering Division (founded 1946)**, part of the NIST's Information Technology Laboratory, seeks to contribute to research in information technology, to catalyze scientific and industrial experimentation, and to improve communication of research results by working collaboratively with, and developing effective statistical methods for, NIST scientists and our partners in industry.
http://www.nist.gov/itl/sed/

The Theory of Applied Statistics
(The Trilogy)

- Develop models with a ***life cycle view***



- Design methodology for assessing ***impact***

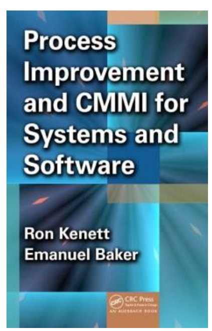
- Improve the generation of ***knowledge***


I am a statistician

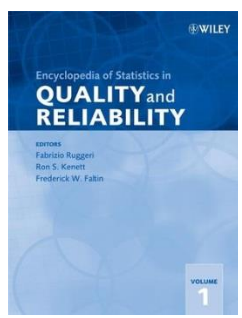
Thank you for your attention

Quality Control

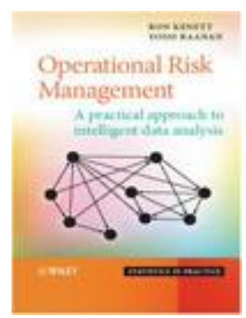
First Edition



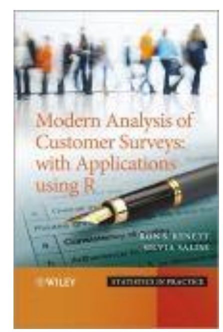
Software Engineering



Four Volumes

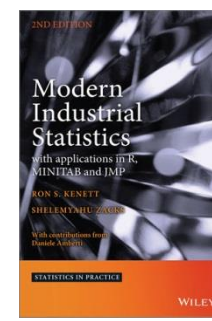


Operational Risks



Customer Surveys

Health Care



Second Edition