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."

David Cox (1981),
“...most statisticians seem to agree that statistics is becoming relatively less influential among the information sciences.”

Jerry Friedman (2001)
“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)
“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.”

Statisticians are not good data scientists

Jonathan Rosenblatt
Statistics postdoctoral at WIS

The place of statistically trained contestants in Kaggle is disappointing if not alarming:


Like • Comment (2) • Unfollow • 23 hours ago
Statisticians are not good data scientists

Jonathan Rosenblatt
Statistics postdoctoral at WIS
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.

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

How to close the gap

David Benjamini

Yoav Banks

David Draper

George Box

Science Businesses Industry Services

1890-1962
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
(à la Draper)

(1) An axiomatica 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 
(à 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 $p_l(A|B)$ — the plausibility (weight of evidence in favor) of (the truth of) $A$ given $B$ — should follow so that $p_l(A|B)$ 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., $p_l(A|B)$ is a function from propositions $A$ and $B$ to $\mathbb{R}$);
- 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
\(\text{(à la Draper)}\)

— 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 \(\{A\ \text{given that}\ B\ \text{is true}\}\) (or equivalently the plausibility of \(A\) and that of \(\{B\ \text{given that}\ A\ \text{is true}\}\)).

• 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
(à 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):
  \[
P(AB|C) = P(A|C) \cdot P(B|AC) = P(B|C) \cdot P(A|BC).
  \]

The proof (see, e.g., Jaynes (2003)) involves deriving two functional equations \( F[F(x,y),z] = F[x,F(y,z)] \) and \( xS\left[\frac{S(y)}{x}\right] = yS\left[\frac{S(x)}{y}\right] \) that \( pl(A|B) \) must satisfy and then solving those equations.
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.
The Theory of Applied Statistics (á la Box)
Sampling and Bayes’ Inference in Scientific Modelling and Decision-Making

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 statistician. 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 (à la Box)

Chapter 16: Mechanistic Model Building
Chapter 17: Study of Variation
The Theory of Applied Statistics (á la Box)


How we teach
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.”

---

**The multi-family selective inference problem**

We select interesting/significant/promising families

The uninteresting families lose 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 \( i \),

\[ P = \{ P_1, P_2, \ldots P_m \}. \quad l = \{ 1, 2, \ldots m \}. \]

Any data based selection procedure of families yields \( S(P) \) in \( l \). 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

\[ E \left( \frac{\sum_{i \in S(P)} C_i}{|S(P)|} \right) \leq q \]

---

**A variety of error-rates**

**Unadjusted inference** \( E(V/m) \leq \alpha \)

**False Exceedance Rate** \( Pr (V/R \geq q) \leq \alpha \)

**k-FDR** \( E( (V-k)/R ) \leq q \)

**False Discovery Rate** \( E( V/R ) \leq \alpha = q \)

**k-FWER** \( P(V \geq k) \leq \alpha \)

**Strong control of FWER** \( Pr(V \geq 1) \leq \alpha \)

**Per family Error-rate** \( E(V) \leq \alpha \)

---

**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
"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."

**What did we learn ≠ What did we do**
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
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.

The Trilogy
The Trilogy of Applied Statistics

• Consider a life cycle view
• Assess impact
• Generate knowledge
Problem Elicitation

Assessing 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

Assessing 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

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


www.eiburs.unimi.it Cost Benfit Analysis in the Research, Development and Innovation Sector.

New frontiers in the design of experiments

Constrained optimization in expensive simulation: Novel approach


This paper presents a novel heuristic for constrained optimization of computationally expensive random simulation models. One output is selected as objective, while the other outputs must satisfy given threshold values. Moreover, the simulation inputs are selected as the objective to be minimized while the other outputs need to satisfy predefined target values.

Response surface methodology for constrained simulation optimization: An overview

Jeroen Kleijnen - Simulation Modelling Practice and Theory, 2008 - Elsevier

This paper summarizes the use of response surface methodology (RSM) for simulation optimization. RSM is used to fit a response surface to the simulation outputs, which is then used for optimization. The paper also introduces a novel approach for constrained optimization using RSM.

Practical applications of design of experiments in the field of engineering: A bibliographical review

L. Lizarazu, M. Alvarez, B. Vives - Quality and Reliability, 2008 - Wiley Online Library

This paper provides a bibliographical review of the practical applications of design of experiments (DOE) in engineering. The paper covers various fields such as quality, reliability, and safety. The review is based on a comprehensive search of the literature, and the findings are presented in a structured manner.

Analysis of an unreplicated fractional-factorial design using nonparametric tests

D.J. Bessels - Quality Engineering, 2007 - Taylor & Francis

This paper presents an analysis of an unreplicated fractional-factorial design using nonparametric tests. The design is used to study the effects of various factors on a response variable. The paper also discusses the advantages and disadvantages of using nonparametric tests in this context.

Quality in design applications in biostatistical pharmaceutical products


This paper discusses the importance of quality in design applications in biostatistical pharmaceutical products. The paper highlights the role of quality in ensuring the safety and efficacy of these products, and the challenges and opportunities associated with achieving quality in this field.

Implementation of design of experiments projects in industry


This paper presents an implementation of design of experiments (DOE) projects in industry. The paper discusses the benefits and challenges of implementing DOE in different industrial sectors, and the strategies used to overcome these challenges.

Order statistics for a two-level, eight-run saturated-unreplicated fractional-factorial screening

D.J. Bessels - Quality Engineering, 2009 - Taylor & Francis

This paper presents an analysis of order statistics for a two-level, eight-run saturated-unreplicated fractional-factorial screening design. The design is used to study the effects of various factors on a response variable. The paper also discusses the advantages and disadvantages of using order statistics in this context.

Risk-based adaptive group testing of semantic web services

S.K. Harris, K. Harris - 2009 COMPUGO 33rd Annual IEEE Conference, 2009 - IEEE Computer Society

Abstract: Adaptive group testing is necessary to ensure the quality of compliant web services that are loosely coupled, can dynamically bind, and integrated through standards protocols. Testing of such web services can be very expensive due to the diversified user.

Multiple-response robust screening in quality construction blue-printing

D.J. Bessels - Journal of Quality & Reliability Management, 2008 - Springer

This paper presents a novel method for multiple-response robust screening in quality construction blue-printing. The method is used to study the effects of various factors on multiple response variables. The paper also discusses the advantages and disadvantages of using this method in this context.

Constrained optimization in simulation: A novel approach

Jeroen Kleijnen, W. Beets, M. Van Nieuwenhuyse - 2008 - papers.ssm.com

Abstract: This paper presents a novel heuristic for constrained optimization of random computer simulation models. One output is selected as objective, while the other outputs must satisfy given threshold values. Moreover, the simulation inputs are selected as the objective to be minimized while the other outputs need to satisfy predefined target values.
Information Quality

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

\[ \text{InfoQ}(f,X,g) = U( f(X|g) ) \]

- \( g \): A specific analysis goal
- \( X \): The available dataset
- \( f \): An empirical analysis method
- \( U \): A utility measure


Journal of the Royal Statistical Society, Series A (with discussion), 176(4).

Joint work with Galit Shmueli
Analysis goal

Available data

Data analysis method

Utility measure

Data Quality

Information Quality

Domain Space

Analytic Space

Knowledge

Goals

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

Available data $X$

Utility measure $U$

Data analysis method $f$

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
Available data

Data Source
• Primary, secondary
• Observational, experiment
• Single, multiple sources
• Collection instrument, protocol

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

Data Size and Dimension
• # observations
• # variables

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)
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

• 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

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

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

Google Flu Trends U.S. may have diverged again from the CDC data it predicts, but too early to be sure.

#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

Collection Timeliness (relevance to $g$)

Analysis Timeliness (solving the right problem too late)

Data Collection  $\rightarrow$  Data Analysis  $\rightarrow$  Study Deployment

$t_1$  $\rightarrow$  $t_2$  $\rightarrow$  $t_3$  $\rightarrow$  $t_4$  $\rightarrow$  $t_5$  $\rightarrow$  $t_6$

$g$: Prospective vs. retrospective; longitudinal vs. snapshot
Nature of X, complexity of $f$
#5 Chronology of Data & Goal

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
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"

#7 (Construct) Operationalization

\[ \chi \text{ construct} \]
\[ X = \theta(\chi) \text{ operationalization (measurable)} \]

- 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?

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?

#7 Operationalization

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

**Critical Reading Skill Groups: 1 2 3 4 5**

**1. Determining the Meaning of Words**

**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**

The example questions below demonstrate the Academic Skills found in this score band. Without looking at the answers, try out the question yourself. Then review the answers at the end of the guide to see how you did.
When asked what the 18% in line 1 meant, 53% of the policy makers responded incorrectly.
The National Assessment of Educational Progress (NAEP) is the largest nationally representative and continuing assessment of what America's students know and can do in various subject areas. Learn more...

In the Spotlight

Try a sample TEL task! Our upcoming Technology and Engineering Literacy assessment includes interactive scenario-based tasks.

NAEP Results on Your Smartphone! Now available for download through the Apps.

#8 Communication

#8 Communication

#8 Communication

#8 Communication

Hardness versus Density

Addition of water:
- > lower density -> softer

More solids: harder

Higher airpressure:
denser + harder
Higher humidity:
softer

More index -> lower density

Auxiliary blowing agent:
more -> lower density, softer

Skeletons on Flying Carpets

Christian Ritter
Assessing InfoQ in Practice

Rating-based assessment
1-5 scale on each dimension:

<table>
<thead>
<tr>
<th>#</th>
<th>Dimension</th>
<th>Note</th>
<th>Value</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data resolution</td>
<td>5</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Data structure</td>
<td>4</td>
<td>0.7500</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Data integration</td>
<td>5</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Temporal relevance</td>
<td>5</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Generalizability</td>
<td>3</td>
<td>0.5000</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Chronology of data and goal</td>
<td>5</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Concept operationalization</td>
<td>2</td>
<td>0.2500</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Communication</td>
<td>3</td>
<td>0.5000</td>
<td></td>
</tr>
</tbody>
</table>

InfoQ Score = 0.68


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

Experience from two research methods courses

- Preparing a PhD research proposal (U Ljubljana, 50 students, goo.gl/f6blA)
- Post-hoc evaluation of five completed studies (CMU, 16 students, goo.gl/erNPF)
Information Quality

\[ \text{InfoQ}(f, X, g) = U(f(X|g)) \]

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

**Goals**

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

**What**

**How**

- Primary Data
  - Experimental
  - Observational
- Secondary Data
  - Experimental
  - Observational
Big Data Analytics

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

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

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)
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

<table>
<thead>
<tr>
<th>Domain</th>
<th>Description</th>
<th>Weight*</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Business Problem (Question) Framing</td>
<td>15%</td>
</tr>
<tr>
<td>II</td>
<td>Analytics Problem Framing</td>
<td>17%</td>
</tr>
<tr>
<td>III</td>
<td>Data</td>
<td>22%</td>
</tr>
<tr>
<td>IV</td>
<td>Methodology (Approach) Selection</td>
<td>15%</td>
</tr>
<tr>
<td>V</td>
<td>Model Building</td>
<td>16%</td>
</tr>
<tr>
<td>VI</td>
<td>Deployment</td>
<td>9%</td>
</tr>
<tr>
<td>VII</td>
<td>Life Cycle Management</td>
<td>6%</td>
</tr>
</tbody>
</table>

*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
Quality Control
First Edition
Software Engineering
Four Volumes
Operational Risks
Customer Surveys
Health Care
Second Edition

Thank you for your attention