American Statistical Association Albuquerque Chapter Meeting

29 April 2011
Santa Fe Hilton 12:30 – 5:00 p.m.

URBAN SCIENCE

July 30–August 4
MIAMI BEACH, FLORIDA
JSM 2011

Statistical Engineering as Practiced at Urban Science

31 July 2011

James G. Wendelberger, Ph.D.
Director of Statistical Analysis

Abstract

- Several statistical engineering solutions are described.
- Included are the engineering constraints under which a statistical engineering solution was arrived at.
- These include the many considerations that are relevant in statistical engineering, such as,
    - Quantitative Theory
    - Technology
    - Management System
    - Statistical Tools
    - Legal Aspects
    - Political Aspects
    - Software Constraints
    - Data Availability
    - Cost (time, money, political, etc.)
    - Computational (Memory, speed, storage)
    - End Result (Report, PowerPoint, Verbal, Software, etc.)
    - Model Constraints (External restrictions)
    - Model Assumptions (External Tenability)
    - Client Constraints (May affect any of the above and possibly more)
    - Delivery Vehicle and Deliverables

Urban Science The Company

We are Urban Science.

We are retail performance experts that help our clients increase market share and improve profitability.

Conquering the toughest business challenges through our unique combination of:

- Cutting edge analytics
- Proprietary software
- Real-world experience

Motivation

This presentation is motivated by **the session on "Statistical Engineering: An Idea Whose Time Has Come? A Discussion in Honor of Gerald Hahn's 80th Birthday"** with

- Martha Gardner, GE Global Research
- Roger Hoerl, GE Global Research
- Bill Parr, China Europe International Business School
- Geoffrey Vining, Virginia Tech.
- Ronald Snee, Snee Associates

presented at the 2010 Joint Statistical meetings.

Outline

Selected ideas presented in that session:

1. Nuclear Statistics
2. Statistical Engineering
3. The Statistician as a Leader

Nuclear Statistics

The analogy of finding the elementary particles in physics in the late 19th to early 20th century to the base statistical methods of the 20th century.

Does this analogy hold today?

I say “No” (mostly).

Nuclear Statistics

The world, when the elementary particles were discovered, presumably has not changed in terms of natural physical laws.

This is very different in the field of statistics today where even the base statistical methods are constantly evolving.

Nuclear Statistics

The statistical world is constantly changing in terms of data availability and computational possibilities.

The statistical tools must be constantly evolving to keep up with these changes.

Nuclear Statistics

"Mostly"?

It may be the case that small sample size statistical tools with simple computations have matured to the extent that there will likely be no further statistical tool evolution in this arena but this is not true of the entire field.

Statistical Engineering

In recent years, Hoerl and Snee have written several articles on the topic of

statistical engineering,

which they define as

"the study of how to best use statistical concepts, methods and tools, and integrate them with IT and other relevant sciences to generate improved results."

Statistical Engineering

What are these

- statistical concepts
- methods and tools
- integration
- IT
- other relevant sciences
- improved results?

These are all constraints!

Statistical Engineering

Constraints

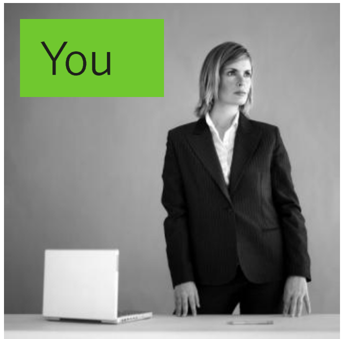
1. Quantitative Theory
2. Technology
3. Management System
4. Statistical Tools
5. Legal Aspects
6. Political Aspects
7. Software Constraints
8. Data Availability
9. Cost (time, money, political, etc.)
10. Computational (Memory, speed, storage)
11. End Result (Report, PowerPoint, Verbal, Software, etc.)
12. Model Constraints (External restrictions)
13. Model Assumptions (External Tenability)
14. Client Constraints (May affect any of above & possibly more)
15. Delivery Vehicle and Deliverables

Statistical Engineering

Issue – Problem – Opportunity
Time
Money
Data Availability
Client Constraints
Many Others

Statistical Engineering

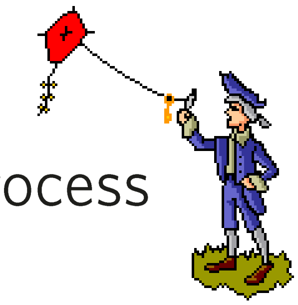
Technical Ability
Technical Knowledge
Technical Resources
Practical Considerations
Consulting Ability



Time
Societal Factors
Corporate Culture
Personal Obligations
Legal Counsel
Many Others

Statistical Engineering

- Time
- Money
- Data Availability
- Practical Constraints
- SM-14
- Statistical Modeling Process
- Many Others



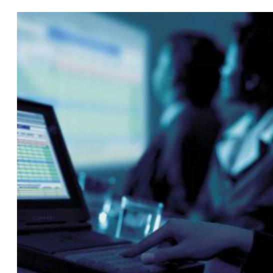
Scientific Method

For SM14 see:
http://www.scientificmethod.com/b_index.html

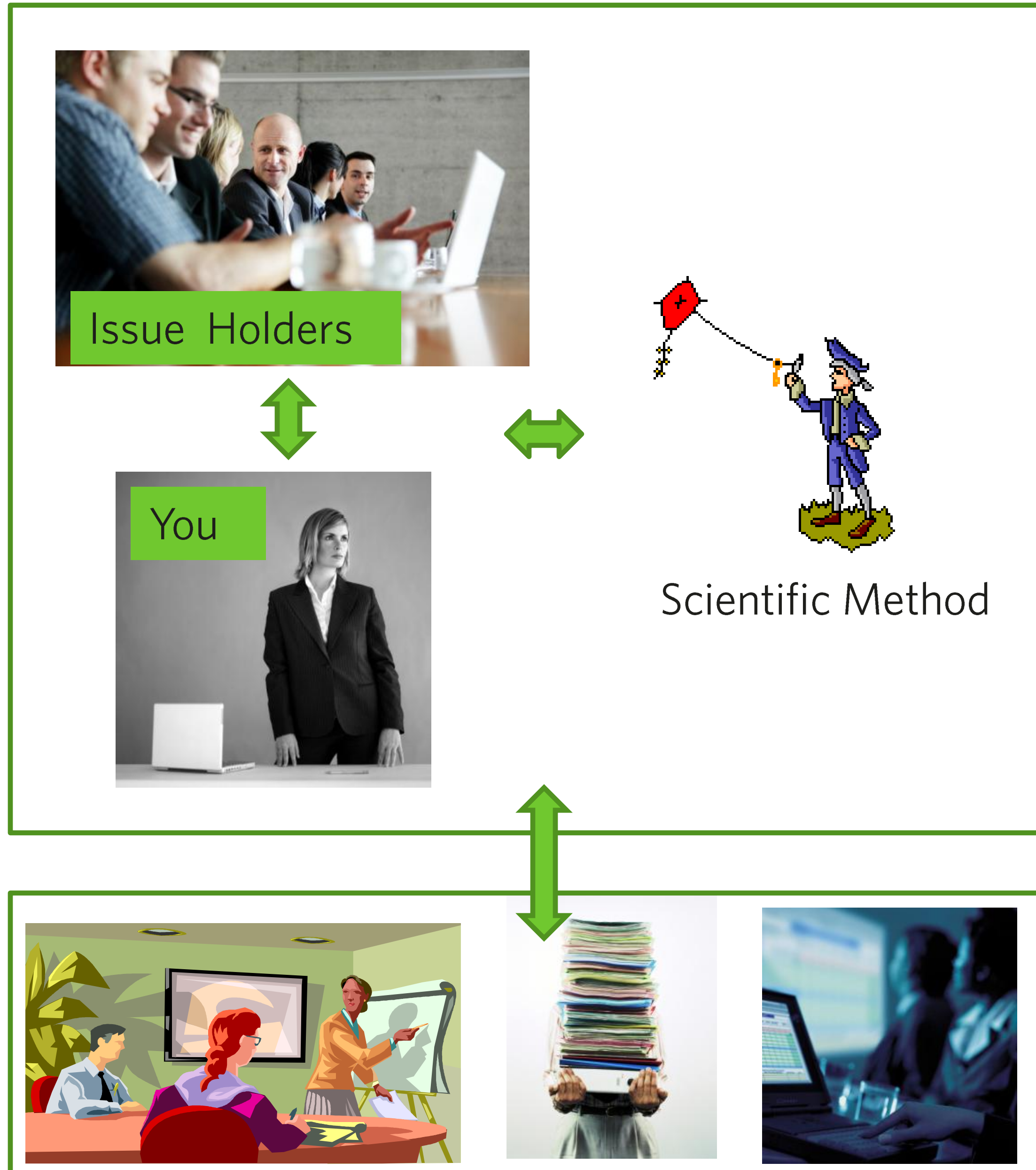
Statistical Engineering

Statistical/Scientific Solution to the Real World Problem
Deliverables
Results
Delivery Vehicle
Cost
Resources
Time
Practical Considerations
Many Others





Statistical Engineering



Quantitative Theory

Self versus Community

Statistical Engineering

Legal Aspects

Risk versus Reward.
Nuclear control room data monitoring.

Statistical Engineering

Management System

Self versus Redirect or Terminate
Fit within mission, vision or personal desire.

Scientific Method (SM14)

Urban Science Mission: the most relevant part of the Urban Science mission is to “provide state-of-the-art solutions to our customer’s problems worldwide utilizing scientific problem solving tools and computer technology driven by real world experiences.”

Statistical Engineering

Statistical Tools

Classical analytical process.
Books, journals, software, reports, etc.
Bayesian, Frequentist, etc.

Huge list.

Statistical Engineering

Political Aspects

External
"Not created here"
Global Cooling, Global Warming, Global Climate Change

Internal
"Home Office Syndrome"

Statistical Engineering

Software Constraints

Available or Unavailable
Self Constraints
Corporate Constraints
Client Constraints
Legal Constraints

Statistical Engineering

Data Availability

Temporal Availability
Source
Original versus Massaged

Is this really the data?

Knowledge about the data that may be used to provide the solution. This data may include: the usual numeric data, expert judgment, analog inputs, textural input or many other types of information. The data knowledge includes intricacies such as data validity, precision and accuracy.

Statistical Engineering

Cost (time, money, political, etc.)

Man hours (Self and Other)
Monetary (Data, Software, Man Hours)
Non quantifiable costs (legal, competitive, political, etc.)

Statistical Engineering

Computational (Memory, speed, storage)

Qualitative
Quantitative
Computational Resource Availability

- Storage
- Memory
- Speed
- Etc.

Statistical Engineering

End Result

Report
PowerPoint
Movie
Verbal
Software
Statement of Work
Interactive Media

Simple to Complex
Client Informed versus Self Informed

Statistical Engineering

Model Constraints (External Restrictions)

Prior Model Coherence
Green field

Client, Self, Corporate Induced

Model Assumptions (External Tenability)

Willingness to accept model assumptions –
Self, corporate, client, legal, etc.

Statistical Engineering

Client Constraints (May affect any of the above and possibly more)

Other client constraints may need to be considered as to their effect on the statistical project.

Within budget
With and without certain assumptions
With and without certain analytical methods
With and without specific software,
With and without certain people,
With and without whatever! When the client says “Jump!” we say “How high?”

Statistical Engineering

Delivery Vehicle

As specified in the project “statement-of-work”.

Statistical Engineering

Early work:

The standard statistical modeling process is divided into three phases. The three phases are:
the **setup phase**,
the **analysis phase** and the
reporting phase.
Each phase is divided into steps. The steps are described in detail in this report.

The Statistical Modeling Process

Urban Science Statistical Report Number: 15

Statistical Engineering

Urban Science Statistical Report # 15

The steps in this report are meant as a guide for those attempting to develop a statistical model. In practice, steps may be undertaken at different levels of involvement. However, it is important that all steps be considered. If any steps are omitted then the potential effect(s) of omission should be communicated to those responsible for the omission.

The statistical modeling process is an art as well as a science. There are many factors which may have not been accounted for by the procedures described in this report. A knowledgeable expert is needed to sort out the importance of considering these other factors.

Statistical Engineering

Deliverables

Document
Presentation
Algorithm
People
Support
Etc.

Statistical Engineering

This may not be an exhaustive list.

Notice the small footprint made by the statistical tools.

Statistical Engineering Examples

At Urban Science we have several examples of Statistical Engineering. Both of these were driven by our president, who has an engineering background.

- GainSmarts
- Network Planning

Statistical Engineering Examples

One of these statistical engineering examples is the GainSmarts data mining system. This system utilizes many statistical tools in a software package that is meant to be run by users of the results. The system design is to interface well with the linguistics of the application area. The system has embedded within it many statistical tools and utilizes them without requiring users to understand the intricacies of these statistical tools. It is made to solve real world problems and has won Knowledge and Data Discovery awards for its performance against other systems, groups and/or experts

Statistical Engineering Examples

GainSmarts

GainSmarts	
Quantitative Theory	Jacob Zahavi and Nissan Levin
Technology	State -of-the-Art
Management System	Technical verification
Statistical Tools	Best Methods - Verification and Testing Split
Legal Aspects	Who owns it?
Political Aspects	Will history detract?
Software Constraints	SAS
Data Availability	Client supplied.
Cost (time, money, political, etc.)	Huge initial investment.
Computational (Memory, speed, storage)	Substantial, even today.
End Result (Report, PowerPoint, Verbal, Software, etc.)	SAS output, log, tables and charts.
Model Constraints (External restrictions)	Restricted to Internal Statistical Models
Model Assumptions (External Tenability)	Internal versus external validity.
Client Constraints (May affect any of the above and possibly more)	Software literate and marketing left knowledgable.
Delivery Vehicle and Deliverables	Software and Support

Statistical Engineering Examples

A second statistical engineering example is our network planning methodology. This is a string of statistical tools that are used to determine the number and location of automotive dealership outlets in a market. This system has been defined formally and in great detail so that it can be faithfully utilized consistently throughout our organization and the world.

Statistical Engineering Examples

Network Planning

Network Planning	
Quantitative Theory	James Anderson
Technology	State -of-the-Art
Management System	Corporate
Statistical Tools	Effectiveness, Buyer Behavior, Franchise Share
Legal Aspects	Corporate
Political Aspects	Legal Case
Software Constraints	Internal
Data Availability	Client supplied.
Cost (time, money, political, etc.)	Huge initial investment.
Computational (Memory, speed, storage)	Normal except for Optimal Location
End Result (Report, PowerPoint, Verbal, Software, etc.)	Tables, maps and charts.
Model Constraints (External restrictions)	Corporate Supplied
Model Assumptions (External Tenability)	Client sSupplied
Client Constraints (May affect any of the above and possibly more)	Client Only Data and Segmentation
Delivery Vehicle and Deliverables	Software and Support

Statistician as a Leader

The ASA JSM 2010 session suggestion to "be a leader" is important advice for any statistician.

Considering all of the factors required for statistical engineering this is a necessary requirement for successful scientific/statistical collaboration.

Statistician as a Leader

Desirable Traits

- In depth knowledge of many state-of-the-art analytical methods.
- Professional exposure to state-of-the-art analytical methods.
- Professional development in the form of academic study, research and publications.
- Cursory knowledge of many application areas.
- Ability to perform the theoretical proofs needed to derive state-of-the art solutions.
- One must have direct contact with the individual that has a problem. The further removed one is from the problem the less likely that the proposed solution will in fact work for the problem at hand.
- A mediator between the client (internal or external) and the solution provider will cause more failures than direct contact will cause.
- Know what constitutes science and the scientific method. Knowledge of the 14 step scientific method is essential.

Statistician as a Leader

Desirable Traits

- One must be well versed in Science in general.
- Know which tools have been and may be successful in scientific solutions.
- Have the mathematical dexterity to create, evaluate and develop state-of-the-art analytical solutions.
- In depth knowledge of scientific tools.
- The ability to derive existing scientific and mathematical tools from first principles.
- The solution must be amenable to run on either corporate or client computers.
- Real world experience must be available.
- Historical knowledge of problems and solutions.
- Knowledge about what has worked and what has not worked in the past.
- Intimate knowledge of the analytical details of prior solutions.
- Knowledge about why some solutions work and others do not.

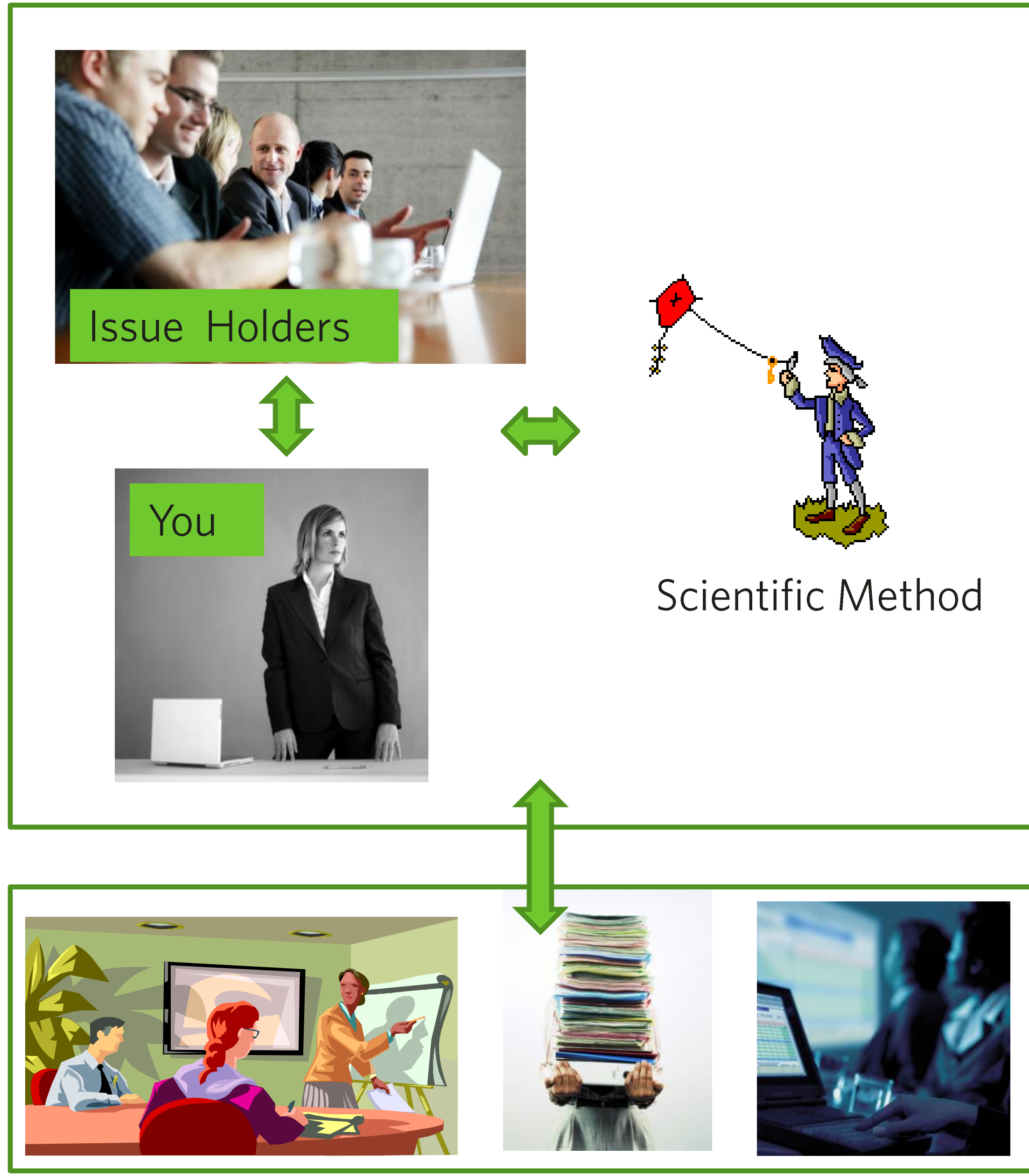
Statistician as a Leader

Desirable Traits

- Knowledge about the data that may be used to provide the solution. This data may include: the usual numeric data, expert judgment, analog inputs, textural input or many other types of information. The data knowledge includes intricacies such as data validity, precision and accuracy.

Statistical Engineering

The intention of this presentation is to provide evidence for and to help stimulate a rigorous establishment of the “Statistical Engineering” process.



References

- Easterling, Robert G., ‘Passion-Driven Statistics’, The American Statistician, Feb 2010, Vol. 64, No. 1: 1–5.
- Hoerl, Roger W. and Snee, Ronald D., ‘Moving the Statistics Profession Forward to the Next Level’, The American Statistician, Feb 2010, Vol. 64, No. 1: 10–14.
- Hoerl, Roger W. and Snee, Ron 'Statistical Thinking and Methods in Quality Improvement: A Look to the Future', Quality Engineering, 22:3, 119 - 129.
- Hoerl, Roger and Snee, Ronald 'Rejoinder: A Consensus that the Statistics Profession Must Change, Is Growing', Quality Engineering, 22:3, 137 - 139.
- Mast, Jeroen de and Does, Ronald J. M. M. 'Discussion of “Statistical Thinking and Methods in Quality Improvement: A Look to the Future”', Quality Engineering, 22:3, 130 – 132.
- Rotelli, Matthew 'Response to “Statistical Thinking and Methods in Quality Improvement: A Look to the Future” by Roger Hoerl and Ronald Snee', Quality Engineering, 22:3, 133 - 134 .
- Scinto, Philip R. 'Statistical Engineering Examples in the Engine Oil Additive Industry', Quality Engineering, 23:2, 125 - 133 .
- Snee, Ronald D., ‘W. Edwards Deming's “Making AnotherWorld”: A Holistic Approach to Performance Improvement and The Role of Statistics’, The American Statistician, Aug 2008, Vol. 62, No. 3: 251–255.
- Vining, Geoff, 'Discussion of “Statistical Thinking and Methods in Quality Improvement: A Look to the Future” by R. Hoerl and R. Snee', Quality Engineering, 22:3, 135 - 136 .
- Wendelberger, James G., ‘The Statistical Modeling Process’, Technical Report Number 15, Urban Science Applications, Inc., Suite 1200, 200 Renaissance Center, Detroit MI, 48243.

James Wendelberger
Urban Science Applications, Inc.
620 Los Pueblos
Los Alamos, NM 87544/2615

JGWendelberger@UrbanScience.com

1-505-662-2838

Thank You!

Dr. James G. Wendelberger

Director of Statistical Analysis