The Human Side of Statistical Consulting*

Bert Gunter
Genentech
April 2005

*With Apologies to Doug Zahn and James Boen, who authored a book with this title.

Disclaimer

- These are my personal views and do not necessarily reflect the opinions or policies of my department, my colleagues, Genentech or your local sponsors

Outline

- Psychology and Sociology
- Statistics
- Practice

-ology

- There are no statistical problems
  - Only engineering, scientific, medical, finance, … problems that require statistical design and data analysis
  - Someone else always owns the data or is responsible for doing the experiment or study
  - If you don't like science, you shouldn't be doing statistics

Consequences

- Learn more about the subject matter
  - It is **your** responsibility to understand what the important subject matter issues are and apply appropriate "statistical" approaches to them (which could and often does mean a good graph or two, as Bill Forrest also emphasizes).

- Success depends on:
  - Collaboration and communication
    - To define the essential issues
    - To determine what data might shed light on them
    - To determine useful analyses of the data
    - To communicate the results of those analyses
  - Imagination
    - Beware of the obvious solution
    - But also beware of reinventing wheels
- These are at least as important as mere technical knowledge

- The most important type of statistical error is not I or II, but III:
  - Right answer – wrong question

John Tukey: "An approximate answer to the right question is worth a great deal more than a precise answer to the wrong question"

George Box: "All models are wrong; but some are useful."

HR

- Everything is personal
  - Subject matter experts/investigators rarely fully understand the statistical issues
  - Hence, their acceptance of *your* methods for dealing with *their* problem and data is based on trust

You should ...

- Care
  - What is the context?
  - Can we do it better than the way it is "usually" done?
- Teach
  - At right level
  - Mostly informally
- Sell
  - the value of statistical methods

Share your enthusiasm. Statistics is not a spectator sport – get involved!

… (suggested by David Giltinan)

- Be wise – compromise!
- But if you must dig in, choose your battles wisely
- Find good people to work with – and nourish the working relationship

-ics

- All problems are statistical
  - Experimental design and data analysis are part of the warp and weft of science

All experiments are designed; the only question is whether well or poorly.

Consequences ...

- You are a professional, so ...
  - Be proactive in your collaboration
  - Help the investigator ask the right question
  - Emphasize importance of good design
  - Never accept data at face value
    - How were they obtained?
    - What systematic sources of variability might mislead?
  - Never assume
    - randomization
    - replication
      - at what levels of the variability hierarchy
    - relevance of past experience

In sum,

- Do not undervalue what you can contribute

  Statisticians are trained to understand and anticipate how variability, which is an inherent part of all natural phenomena, can affect observations of reality. This gives us powerful insight that many scientists do not possess. We need to use that insight to "help catalyze the scientific learning process." (George Box)

But ...

- Do not overvalue it either
  - Prior knowledge and experience matter even if they cannot be easily captured and quantified
  - All relevant information does not reside in the data at hand
    - Frequently, very little does.

Practice

- Collaborators, colleagues, co-workers
  - Not "clients"
  - Attitude makes a difference

Turf:

- Whenever possible, meet on theirs, not yours
  - CBWA: "Collaborating by Walking Around"
  - Touch the equipment, meet the subjects, observe the critters, try out the product, ...

Rules to Practice By

Rule 1:

- **NEVER** give advice over the phone.
  - It will always be bad

Rule 2

- **NEVER** answer when asked how many.
  - It is almost always the wrong question
  - The right question has to do with defining the goals of the experiment
    - Typically, you will have to help the investigator figure this out.

Rule 3

- Always interrogate the measurement
  - Systematic sources of measurement variability abound
  - Ivestigators rarely know how to quantify them
    - But they often can tell you what could be there

Rule 4 (suggested by Lisa Bernstein)

- Get the raw data
  - Many instruments, vendor-supplied/user-created software "preprocess" the data in ad hoc, crazy ways devised by folks with no statistical training
    - Producing irretrievable junk that no subsequent analysis can redeem
  - Often difficult to get "raw data" and difficult to deal with when you have them
    - e.g. images

Rule 5

- Thermodynamics: Disorder rules unless you work hard to defeat it. So...
  - Take nothing for granted
  - Provide explicit step by step instructions, data format specifications, ...
  - Remember Murphy

Some useful things I've learned

- Most science is about hypothesis generation, not hypothesis testing
- All scientists have strong priors
- All variation is caused
- Beware of the data that aren't there (suggested by Bill Forrest)
- All replicates are not created equal
- You can never know too much statistics
  - But most of what you learn is silly

A few more ...

- Listen more, talk less
- Always provide an "executive summary" of your results in a graph or two and a paragraph or so of text
- A little paranoia can be good thing
- It's better to lead the parade than sweep up after the elephants (good design is more important than fancy analysis)
- If you don't think it's ethical, don't do it
  - Your job is to speak for the data – integrity is everything!

And most important ...

- **HAVE FUN !**
  - (else why do it?)