Statistical Consultancy

David J. Hand

Imperial College
and
Winton Capital Management

Statistical consulting requires skills and expertise completely distinct from academic and technical expertise

'90% of UK academic statisticians have never done any statistics'

Is this true?

What does it mean?

- My background
- Consultancy environments
- The key attributes of a successful consultant
- The initial meeting
- Initial data analysis
- Analysing the data
- Time frames
- Your report
- Presentations
- Confidentiality, data security, and Chinese walls
- The contract
- How much should you charge?

My background

BA Maths
MSc Applied statistics
PhD Multivariate statistics

1977: Institute of Psychiatry
1984-91: Sigma X Ltd
1988: Open University
1999: Imperial College
2010: Two years leave of absence, commercial sector

IoP: consultant / collaboration role to researchers
ONS Methodology Advisory Committee, 2001-2009
GSK, AstraZeneca Advisory Committees

Private consultancy:

Banks, hedge funds, and other financial bodies (including Barclays Capital, Barclays Direct Loan Division, Barclaycard, TSB, Abbey National, Goldman Sachs, Bluecrest, Littlewoods, Fair Isaac, GMAC-RFC, British Credit Trust, Capital One, Alliance and Leicester, Lloyds, etc).

Pharmaceutical companies (including Sandoz, Pfizer, Upjohn, Lundbeck, Ethical Pharmaceuticals, Proctor and Gamble, Boots, Ciba-Geigy, Napp Pharmaceuticals, Keith McCormack, Unilever, AstraZeneca, and GlaxoSmithKline).

Governments (e.g. I am on the Methodology Advisory Committee for the UK Office for National Statistics, and have also advised the European Statistical Office, Eurostat)

A wide variety of other bodies (e.g. T-Mobile, Ted Bates Ltd, Forvus Computer Services, National Union of Journalists, RHR Management Consultants, Qinetiq, Shell, BP, GPG Plastics, National Institute of Social Work, Health Education Authority, the Wellcome Trust, BAE Systems, the Wellcome Trust, the Institute of Actuaries, the National Patient Safety Agency, The Royal Society, etc.)

Different consultancy environments

As part of a statistical service/advice unit,
within a university, corporation, gov dept, etc

- consultant to many projects, service unit (OU, IC)
- obliged to answer all who request advice
- collaborator or servant? Map out rel'n'p beforehand: co-author?

Salary paid: security
Interesting problems
Can have variety
Can become an expert in the area

Consultancy from an academic base (*Salary paid*)

- internal, to other staff and researchers seeking advice
  - interesting projects
  - you can choose
  - salary paid but extra work is unpaid
  - co-authorship
  - failure does not mean career end or starve
  - however, can eat into your time
    - ➔ OU and IC SAS

- external, to people and organisations outside the university
  - extra income
  - and extra responsibility and risk (law suit, reputational risk?)
  - what are your university's policies?

Freelance consultant

- income depends on success and reputation
- the risks of being self-employed
- promoting yourself to potential clients, networking
- when soliciting new work you are not earning money
- time spent invoicing clients, chasing slow payers
- professional indemnity insurance

Expert witness

- specialist area

- high reward, high stress !

- formal training courses
  (e.g. *Expert Witness Institute*, *UK Register of Expert Witnesses, The Academy of Experts, etc*)

- ***you are only an expert in your narrow area***
  - the Sally Clarke case
  - there will be expert statisticians on the other side

- the lawyers are experts at tying you up in knots

What you need to be a successful consultant

Affability

Availability

Ability

in decreasing order of importance!

For general commercial consulting, the technical aspects are not the most important thing

People relationships are the critical factor

Listening to the client and understanding

- the aims: what the client wants

- the constraints
  - from the client's perspective
    what does the client understand?
  - from a practical perspective
    what can be achieved
  - on resources (your time, their money, ...)
  - on the data (*and its shortcomings*)

The initial meeting

Getting a feel for the project
My strategy
- initial meeting, am I interested, am I expert enough?
- be prepared to say no (ethics, expertise, timeframe,...)

- Ask questions
- Take notes
- *Don't be afraid to say you don't know*
- *Listen* to the answers

I *always* sleep on a problem
Bounce things off more experienced colleague

Determine what the client wants to know: may not be what they say at first (they are not statisticians)

e.g. ‘is this medicine more effective than that?’

Be prepared to say ‘*you cannot answer that question with these data*’ (or risk misunderstanding, disgruntled client, expensive law suit)

Be aware of the client’s level of technical knowledge

- they won’t have your statistical expertise
- they will know about the problem and the data
- they are paying you
- they did not come in for a lecture on statistics

Translating the problem into statistical terms

- to a man who has a hammer....

Initial data analysis

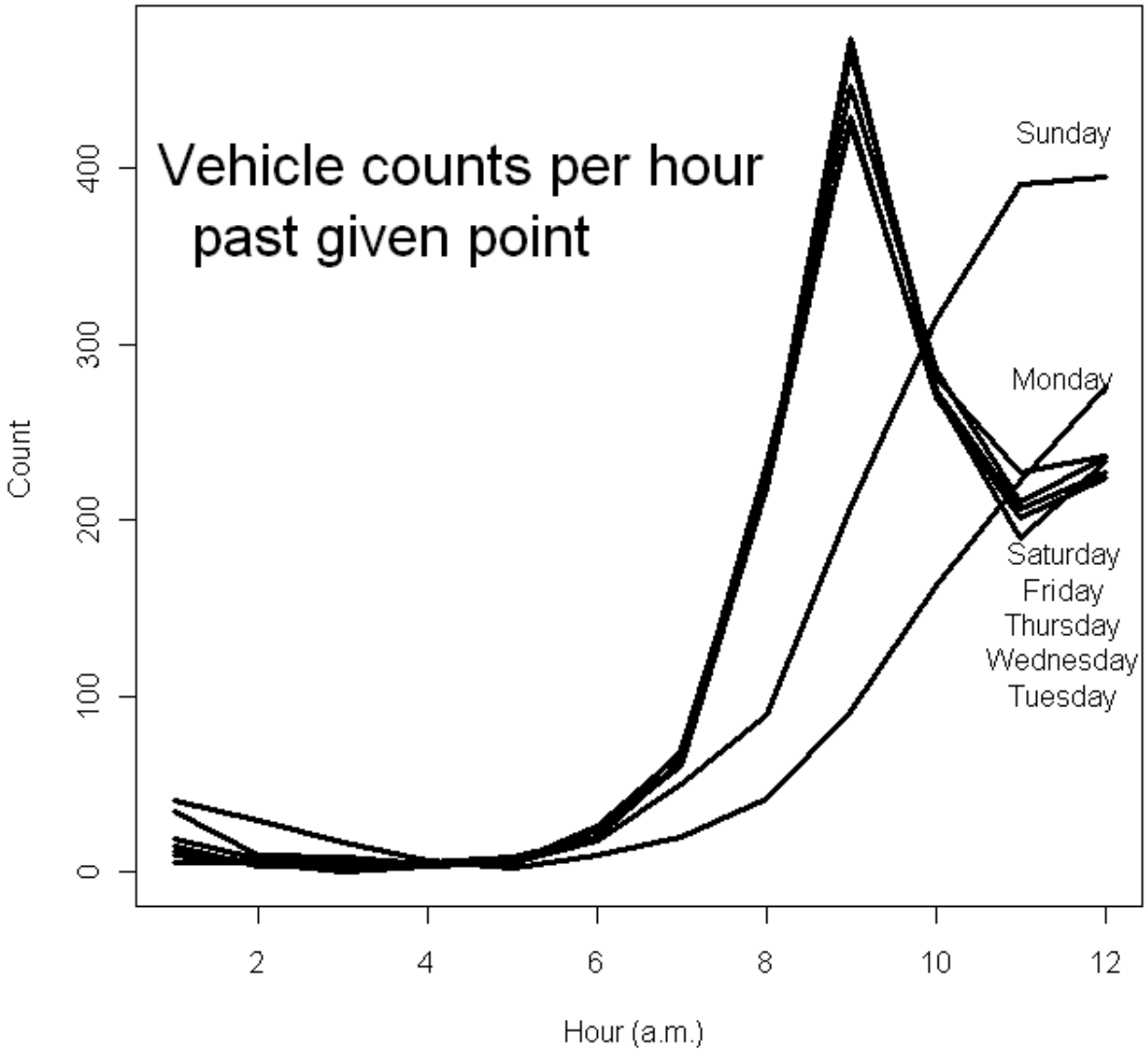
Reading the data, making sense of the data, understanding the data: *90% of data analysis is data cleaning*

- Do not take the data at face value
- Always undertake quality checks (pharma e.g.)
  Build this into your estimated timing

    e.g. How will you handle missing values? (*If there are no missing values then why not?*)

    e.g. If you are responsible for arranging data entry

	Day						
	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Hours							
0000-0100	41	10	12	5	15	19	35
0100-0200	29	4	5	5	3	8	10
0200-0300	17	3	3	0	4	6	9
0300-0400	7	5	5	3	5	3	4
0400-0500	2	6	9	6	7	6	6
0500-0600	10	24	20	20	20	26	18
0600-0700	20	60	63	62	70	68	50
0700-0800	42	219	232	220	225	222	89
0800-0900	91	429	425	473	447	466	207
0900-1000	163	270	274	272	285	281	314
1000-1100	222	201	206	190	210	227	391
1100-1200	276	224	227	233	235	236	395


Vehicle counts per hour
past given point
Count
0
100
200
300
400
2
4
6
8
10
12
Hour (a.m.)
Sunday
Monday
Saturday
Friday
Thursday
Wednesday
Tuesday

Analysing the data

- a cyclic process: interaction with the client

??? Frequentist (Fisherian, Neyman-Pearson, ...), Bayesian (subjective, logical, ...), Likelihood, or any of the many other schools of inference ???

Different approaches to inference answer different questions

Be pragmatic and eclectic not dogmatic: choose the right tool for the job

(Though note that the word 'Bayesian' has become a buzzword, and is applied to all sorts of analysis which are not Bayesian: the client might like that!)

Time frames

Often a source of difficulty

- Commercial time scales often an order of magnitude less than academic ones

Often an adequate solution will do

- Seldom need the best

In one business, we waited more than 20 months for a professor of statistics to come up with the "Cadillac" of scoring systems, while all the business needed was a "Chevrolet" that would work."

David Lawrence, Citicorp.

Your report

Spelling
punctuation
grammar
format
are crucial

if you aim to establish a reputation and repeat business

If you are not a native English speaker then (confidentiality permitting) get a native speaker to comment on the report

- *and learn from their comments*

Title page: (title, author, contact details, date)

Executive summary

Introduction

Methodology

Results

Conclusion

References

Appendix: (giving technical details: organise properly, not just a dump of computer output!)

Presentations

- Put yourself in the client's position
- Focus on results
  (Your aim is to answer the client's question, not show them how clever you are)
- Not a statistics seminar
- Keep it short

Rehearse, rehearse, rehearse !

Confidentiality

- Is often paramount
- Non-disclosure agreements
- Contravention could mean being sued

Data security

May be crucial

e.g. some of my work:

- all data and doc'ts to be encrypted and password protected
- all data to be kept in locked office
- not on a computer connected to the internet or any LAN
- always get prior agreement before sending anything via email or post (may be *discoverable*)

Chinese walls

May work on similar topics for competing organisations

Say so up front, even if it means you lose the contract

Behave ethically !

The contract

- describe what the client expects
- when the data will be delivered
- fee structure
- nature of expenses to be paid
- nature of deliverables (report, presentations, programs,..)
- when deliverables due
- payment terms, and penalties for late payment

- cyclical nature: supplementary report

Commercial timescales mean that not all negotiations result in a contract

How much should you charge?

- depends on length of project ➔ need to estimate
- different approaches
    - fixed sum for entire project
    - fee per hour or day
    - profit sharing
    - etc

Charge appropriately for what you are bringing to the table

Some of my projects

Clinical trials
Effectiveness of advertising
Underwater mine detection
Contaminated banknotes
NASA fault detection
Pharma manufacturing IP
IP on models for credit scoring
Scientific fraud
Banking fraud
Knife murder
Fake designer goods on eBay
Sex discrimination
Money laundering
Toothbrush design

Credit scoring
Forex
MoD
Hedge funds
Telecomms
Hedge funds
Phone charge rates
Job satisfaction
Car parts manufacture
etc.

Summary

- Money: supplementary income or main income
- Interesting problems
- Learning about real problems (90% of UK ac. statns...)
- Advances in statistical methodology (Fisher, Cox,...)

- 90% of analysis is data cleaning and quality issues
- Client relationship is the most important issue

- Put yourself in the client's shoes

- Behave professionally: ethics

Bibliography

Boen JR and Zahn DA (1982) *Human Side of Statistical Consulting*. Lifetime Learning.

Hand D.J. and Everitt B.S. (eds.) (1987) *The Statistical Consultant in Action*. Cambridge University Press. ***(including a bibliography of 68 articles on statistical consulting)***

Chatfield C (1995) *Problem Solving: A Statistician's Guide*. Chapman and Hall.

Derr J (2000) *Statistical Consulting: A Guide to Effective Communication*. Duxbury

Cabrera J and McDougall A (2002) *Statistical Consulting*. Springer.

Hahn G.J. and Doganaksoy N. (2008) *The Role of Statistics in Business and Industry*. Wiley.