Statistical Engineering

Tim Davis

Imperial College, February 17, 2010

The brief

Some questions to consider. In your professional experience

- Which are the frequently used statistical methods?
- Which methods did you want to use, but couldn't? Why not?
- What is the status of the Bayes/frequentist debate?
- How do you balance mathematical details with practical concerns?
- How do you balance state-of-the-art methods with more tried and tested methods?
- What are the common software tools?
- how important are computing skills?
- how important is it to continue to develop new computing skills?
- What issues arise communicating sophisticated statistical ideas;
  - to statistically weak colleagues and customers?
  - to senior management?
- When acting as a consultant,
  - How do you tease out the problem from the client?
  - what common problems and misunderstandings occur?
  - How do you give the client bad news (eg. The experiment does not give a significant result)
- How do you get in to the game?
- How do you get ahead?

Some of the topics we will discuss

- Induction and Deduction, and why it is important for statistical applications, particularly in industry
- Analytical & Enumerative studies
- Statistical Process Control
- Reliability & Failure Mode Avoidance
  - Mistake avoidance
  - Robustness improvement
- Experiments

Induction and Deduction

H=hypothesis ; D=data

- **Deduction: Pr(D|H).** This probability has a frequency interpretation - *aleatory* uncertainty.
- **Induction: Pr(H|D).** This probability has a degree of belief interpretation – *epistemic* uncertainty.

e.g. H= the coin is fair; D=45 heads in 100 tosses

Pr(D|H) is deductive → no enquiry necessary
→ **probability theory** → hypothesis testing

Pr(H|D) is inductive →enquiry necessary

→ **statistical science** → hypothesis *generation*

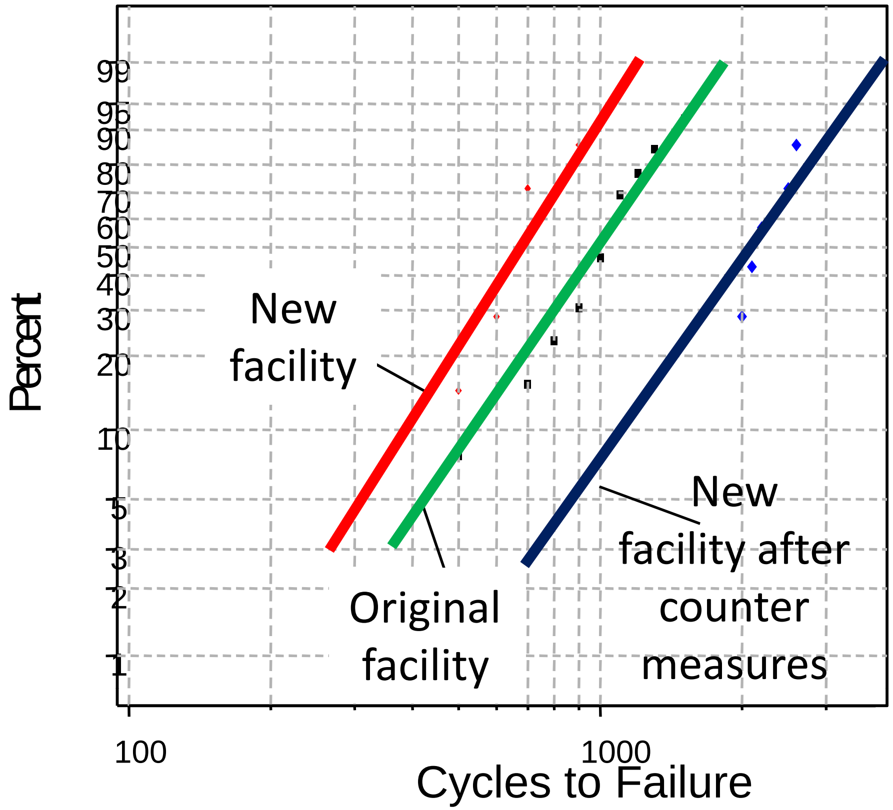
An engineering example

- An established vehicle design produced in a new manufacturing facility suffered an unusual, high severity structural welding failure 2/3 of the way through a durability test
- Subsequent lab test results (data=**T**) from samples of parts of the design produced in the two manufacturing facilities showed potentially inferior results for parts produced in the new facility.
- The hypothesis is that the reliability in the field of the product from the new facility will be the same as that from the original facility (hyp=**R**).
- In order to authorize production, do we need to evaluate Pr(**T**|**R**) or Pr(**R**|**T**)?

An engineering example - cont

<u>Hypothesis Testing Pr(**T**|**R**)</u>

- p-val=0.15
- Do not reject null hypothesis
- Ship product



<u>Hypothesis Generation Pr(**R**|**T**)</u>

- Investigate the differences between the 2 facilities
- Deploy counter measures
- Try for an order of magnitude improvement

We could say:-

Statistics is the science of making inferences through inductive logic and reasoning in the face of uncertainty.

Consequences of confusion

- Most problems in industry need inductive logic
- Many initiatives, supposedly aimed at quality improvement, such as Six Sigma & the D-M-A-I-C process have failed to teach the distinction between induction and deduction.
- Consequently, many practitioners use methods better aimed at deductive inference (e.g. significance tests) when trying to solve inductive problems.
- The probability you have measles given that you have spots is not the same as the probability that you have spots given that you have measles.

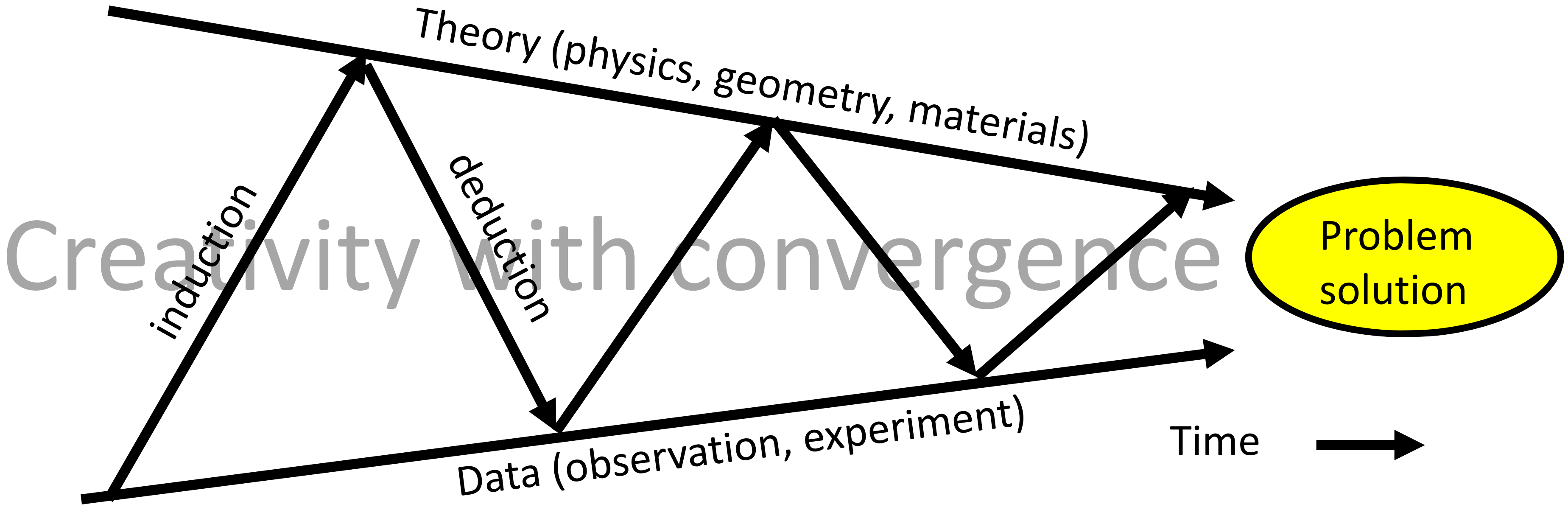
$$\text{i.e. } \Pr(\mathbf{D}|\mathbf{H}) \neq \Pr(\mathbf{H}|\mathbf{D})$$

Common mistakes in solving engineering problems

- “there are multiple root causes for this problem” – engineering equivalent of a conspiracy theory → it’s easy to make complicated theories fit the “facts”.
- Data thrown into Minitab → grope around in the output for significant “p-values”
- Lack of progress in solving the problem → too much data collection/analysis devoted to eliminating root causes that, through deduction, can be shown not to be true.

The iterative learning process

After George Box



- It is the job of the statistical investigator/ collaborator to ensure convergence
- Speed of this process determines what sort of statistical approach is required (industry usually quick)
- "Deduction is analysis, induction is science, synthesis of the two things is engineering" (Mischke)

A tool to aid convergence

- The "IS" / "IS NOT" Matrix
- Define some criteria
  - what is the defect?
  - when did we first observe the defect?
  - where did we first observe the defect?
  - what is the pattern or trend in the data?
  - etc...
- Ask what the problem "IS" relative to these criteria
- Then ask what the problem logically *could* be, but "IS NOT"
- Use the answers to these questions to filter the possible root cause theories
- Only experiment with theories that cannot be eliminated in this way

IS / IS NOT example

PROBLEM Vehicles suffer tire failure and roll over	What the problem IS	What the problem *could* be but IS NOT	*THEORY 1* There is a problem with the vehicle	*THEORY 2* There is a problem with the tire
What is the defect?	Tread Separation	Blow-out	+	+
What object has the defect?	Tire Brand A	Tire Brand B	-	+
When was the defect first observed?	3 years after vehicle on sale date	Immediately the vehicles went on sale	-/+	+
Where was the defect first observed?	In hot southern States of the US	In mild temperate states	-	+
What is the trend in the defects	Tires from Factory X have a higher failure rate than from Factory Y	Tires from each factory have the same failure rate	-	+
What is the nature of the failure rate?	IFR with time	CFR or DFR with time	-/+	+

Analytic vs. Enumerative studies

- Great emphasis placed on this by WE Deming
- Enumerative study – describes a known entity
  - e.g. How many defective parts are there in this particular batch of incoming material?
  - Requires us to construct a carefully selected random sub-sample that describes the entity. Action is taken on the entity.
- Analytical study – predicts the state of future entities
  - e.g. How many defective parts are there likely to be in future batches of incoming material not yet produced?
  - Requires us to make predictions about entities that don't yet exist. Action is taken on the process that produces the entities
- These two types of study present different methodological challenges

A word on Statistical Process Control

- Main tool – the Control Chart, due to Shewhart.
- Helps with analytical studies (change the future to make it more predictable).
- How? Provides an operational definition of when to treat problems as either *special* cause or *common* cause.
- Gaussian distribution ($\pm 3\sigma$ etc) not important for Control Charts to work.

Reliability

- *Probabilistic definitions*
  - Reliability is the **probability** that a unit will perform its intended function until a given point in time under *specified* usage conditions

    → $\mathbf{Pr[T>t|N_s]}$
  - Reliability is the **probability** that a unit will perform its intended function until a given point in time under *encountered* usage conditions

    → $\mathbf{\Sigma_i Pr[T>t|N_i]Pr[N_i]}$
  - These probabilities can only be estimated from enumerative studies, but are often treated as if they are analytical (predictive).

Reliability

- *Information based definition*
  - Reliability is **Failure Mode Avoidance** (unit of information is a counter measure for an identified potential failure mode) → **identify potential failure modes, engineer and evaluate counter measures against a range of conditions**
  - This is recognised as an analytical problem. Key tool is the FMEA (barely referenced in reliability textbooks)

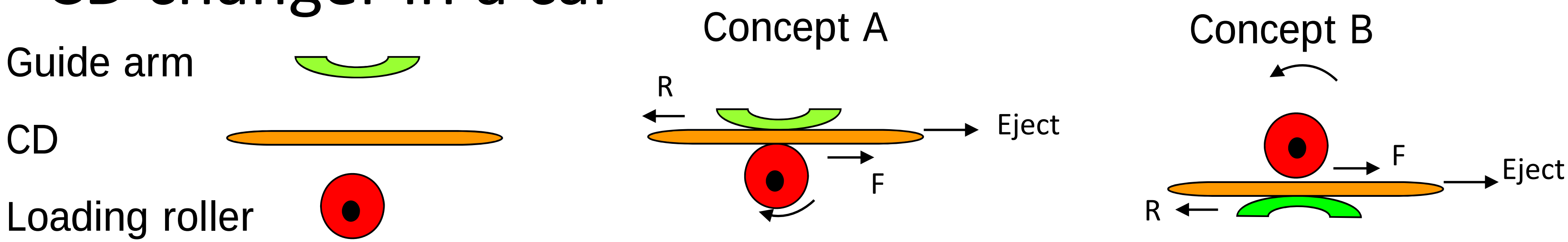
- We have to choose between an enumerative study or an analytical study – we can't do both!
- See Feynman's "inflamed appendix" - his report into the 1986 Challenger disaster.

Reliability

- Two causes of failure modes
  - Mistakes
  - Lack of robustness
- Prevention of mistakes is primarily a matter of vigilance
- Improvement of robustness needs a statistical approach.
- ***Failure Mode Avoidance*** provides a treatment for both situations

Mistake avoidance example

CD changer in a car



R = retention force F = ejection force

- In concept A, the addition of a paper label on the CD allows R>F. → CD sticks
- In concept B, *even with* a paper label, R<F always. → CD can't stick
- Hence choice of design concept A is a mistake. Reliability effort is best placed ensuring Concept B is chosen, rather than trying to predict how often Concept A will fail.
- The job of the engineer is to choose the design that will fail the least, not to predict how often the chosen design will fail.

Robustness

- Robustness = product & process performance that is insensitive to disturbances.
- Disturbances are called “noise factors” e.g.
  - i. Variation in product characteristics due to production rate.
  - ii. Variation in product characteristics due to usage.
  - iii. Customer usage profile (drives fast, drives slow, etc)
  - iv. Environment (hot, cold, etc)
  - v. System interfaces (vibration, heat transfer etc)

  (The five sources of noise)

- Two questions emerge
  1. How should we measure robustness?
  2. How should we search the design space for robust solutions?

Measuring robustness

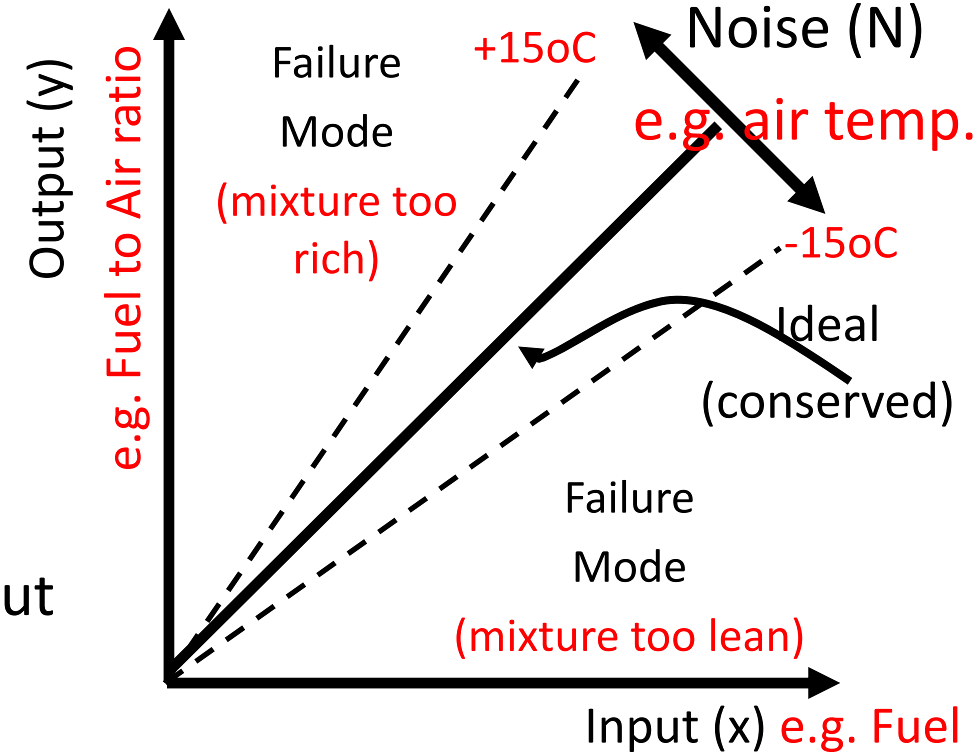
- To answer Q1, Taguchi used a signal to noise ratio:- S/N = log(μ/σ)

  μ=average product performance;

  σ = variation in performance induced by noises.
- Much controversy ensued in the statistical literature, in conferences, and 1-1 conversations ....

Engineering solution (with example)

- Engineering function is about transforming or transporting
  - Energy
  - Materials
  - Information
- Since these are conserved quantities, the basic transfer function between input & output will be linear



"Ideal" Function: $y=\alpha_0 x$. → "Noise Disturbed" Function: $y=\alpha_0(1+\alpha_1 N)x$.

Robustness is measured by α_1, a parameter in the transfer function.

Equivalent to Taguchi's S/N ratio: S/N = $\log(\alpha_0/[\alpha_0\alpha_1])=-\log(\alpha_1)$.

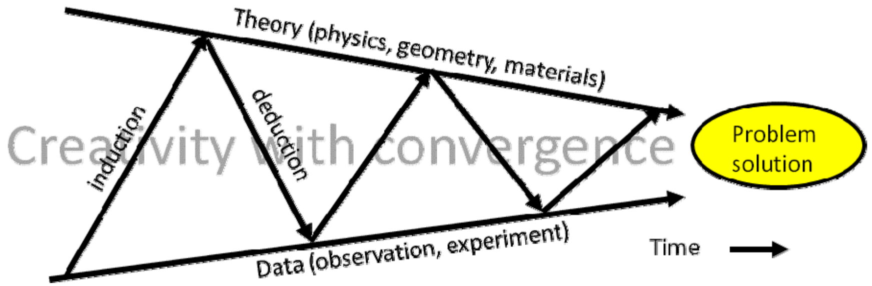
α_1 measures the "distance from the failure mode(s)"

Running engineering experiments

- With regards to Q2; some controversy introduced by Taguchi’s follower’s with regard to the treatment of interactions in experiments.
- Six Sigma training programs haven’t helped - too much emphasis on full factorials, ANOVA, and gauge capability at the expense of fractional factorials, graphical methods and hidden replication.
- If we get back to fundamentals, we can perhaps, start to overcome some of this poor teaching.
- Deficiencies in the skills required to run well planned experiments is a serious impediment to industrial effectiveness.

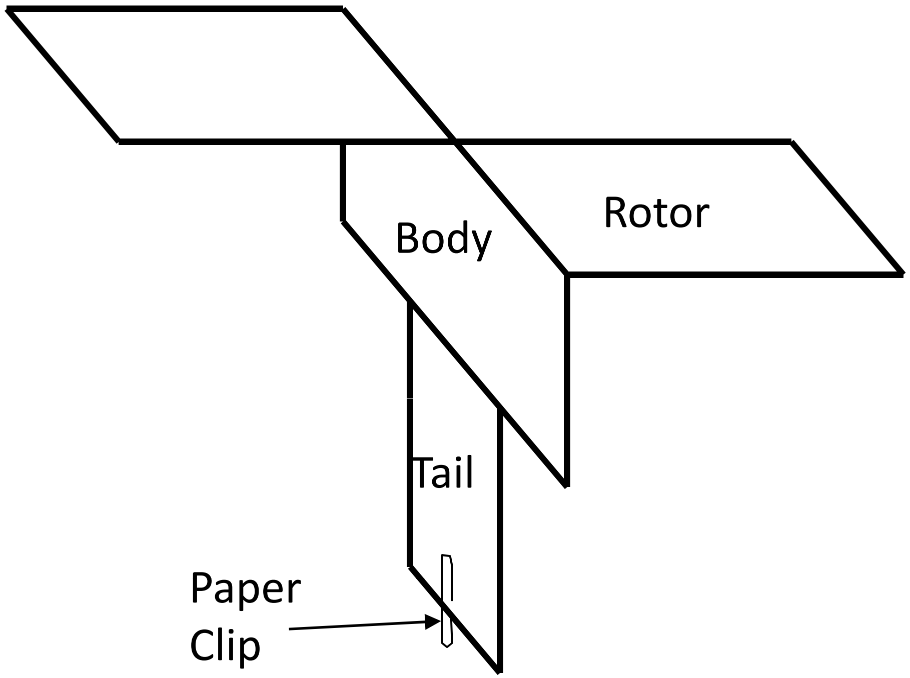
Dimensional Analysis

- Buckingham's Pi theorem: A functional relationship in *n* variables and *m* fundamental units can be rewritten in terms of $N \geq n-m$ dimensionless variables.
- This is an extremely useful theorem to drastically reduce the number of runs in an experiment.
- Requires some basic knowledge of the physics of the system being studied.
- Exemplifies the iterative nature of the deductive/ inductive learning process discussed earlier



Example – paper helicopter

Maximize the flight time, T, of the helicopter



Typical factors that might be used in a response surface experiment:

Rotor radius (x_R)

Tail length (x_L)

Tail width (x_W)

$T=f(x_R, x_L, x_W)$

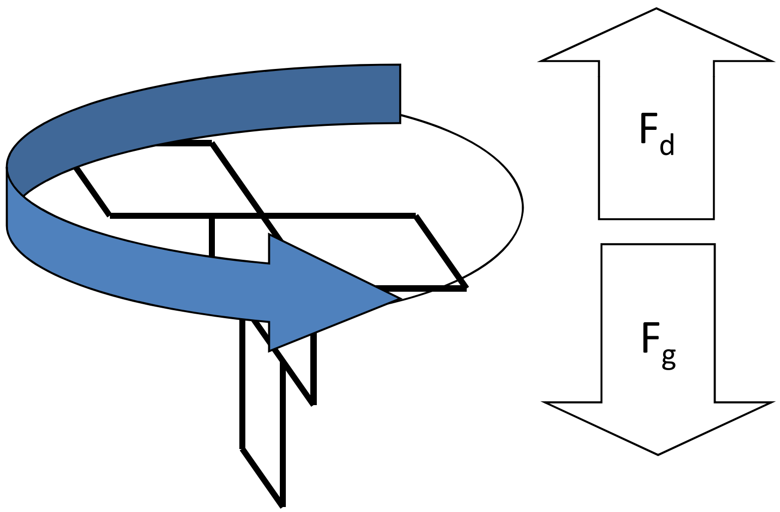
We could approximate this with a 2nd order response surface which would need ~15 runs to estimate.

$$T = -0.03\text{-}0.008\,x_L - 0.011\,x_W + 0.415\,x_R - 0.016\,x_R^2 - 0.002\,x_L x_W + 0.001\,x_L x_R + 0.001\,x_W x_R$$

(on the face of it) Dimensionally Inconsistent

Paper helicopter physics

- The helicopter very quickly comes to a steady state velocity (V_{ss})
- Time of flight (T) is determined by V_{ss} and the launch height (h)
- V_{ss} determined by the balance between the force of gravity F_g and drag F_d
- F_g is determined by the mass of the helicopter (M) and g
- F_d is determined by the area swept out by the rotor radius (R_R) and the density of air (ρ_{air}).



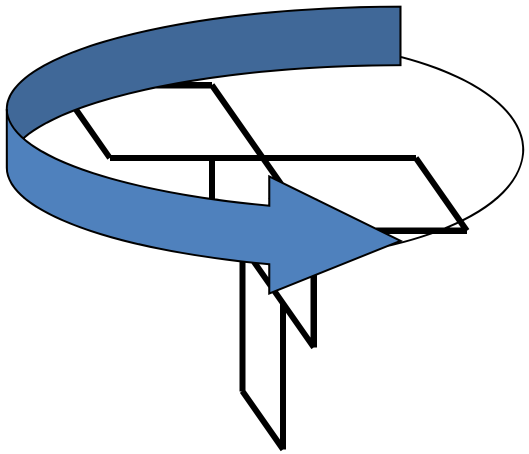
Without knowing the form of the relationship we can write down the important variables.

$T = F_1(M, g, \rho_{air}, R_R, h)$

Paper helicopter physics

- $T = F_1(M, g, \rho_{air}, A_R, h)$
- From the physics we already know exactly how T depends on h → $T = h/V_{ss}$
- So we are looking for an expression of the form

$$V_{ss} = F_2(M, g, \rho_{air}, A_R)$$



$V_{ss} = h/T$	m / s
M	kg
g	m/s^2
ρ_{air}	kg / m^3
R_R	m

- We have *n*=5 variables with *m*=3 fundamental units. Therefore we can express this in terms of 5-3=2 non-dimensional parameters.

Dimensional Analysis for the helicopter

- Define 2 core variables $\Phi_V \equiv V_{ss} R_R^a \rho_{\text{air}}^b g^c$

$$\Psi_M \equiv M R_R^d \rho_{\text{air}}^e g^f$$

- Analyze the dimensions of the core variables

$$[\Phi_V] \equiv \frac{m}{s}(m)^a \left(\frac{kg}{m^3}\right)^b \left(\frac{m}{s^2}\right)^c \quad [\Phi_M] \equiv kg(m)^d \left(\frac{kg}{m^3}\right)^e \left(\frac{m}{s^2}\right)^f$$

$$= m^{1+a-3b+c} kg^b s^{-1-2c} \qquad = m^{d-3e+f} kg^{1+e} s^{-2f}$$

- Enforce non-dimensionallity $a = -\frac{1}{2}; b = 0; c = -\frac{1}{2}; d = -3; e = -1; f = 0$

$$\Phi_V = \frac{V_{ss}}{\sqrt{gR_R}} = \frac{h}{T\sqrt{gR_R}}, \Psi_M = \frac{M}{\rho_{air} R_R^3}$$

Paper helicopter experiment

- We now need to fit a (dimensionless) equation of the form $\Phi_V = F_3(\Psi_M)$
- 3 experimental runs is the minimum that is needed to measure any curvature between Φ_V and Ψ_M.
- Change x_R, x_L, x_W, measure T, calculate Φ_V and Ψ_M.

Tail Length	Tail Width	Rotor Radius	Ψ_M	Φ_V
5	3.2	12	1.975	1.069
5	3.438	8.744	3.410	1.405
7	5.1	7.62	4.845	1.675

$$\Phi_V = 0.664 + 0.211\Psi_M$$

(Dimensionless)

Paper helicopter transfer function

- The non-dimensional form is converted back into original units and solved for T.

$$\Phi_V = 0.664 + 0.211\Psi_M$$

(Dimensionless)

$$\Downarrow$$

$$T = \frac{h}{\sqrt{gR_r}\left(0.664 + 0.211\frac{M}{\rho_{air}R_r^3}\right)}$$

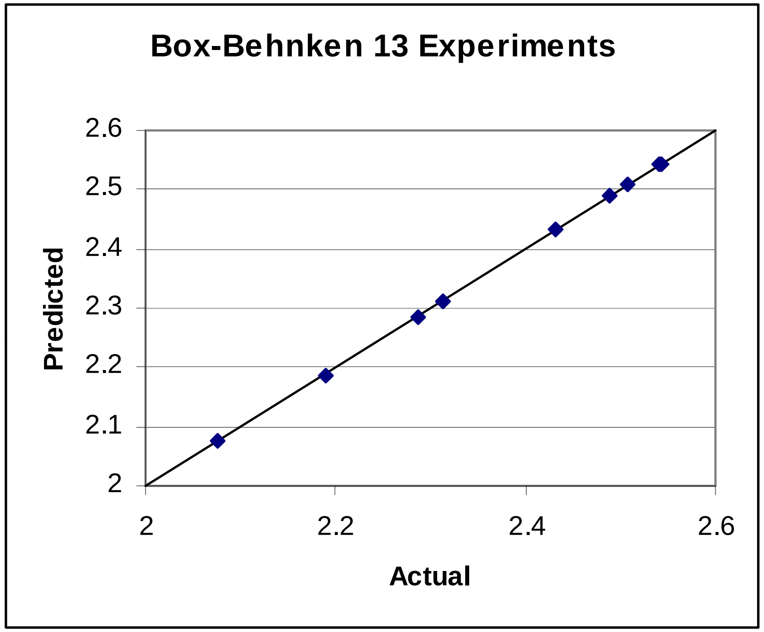
(Dimensionally consistent)

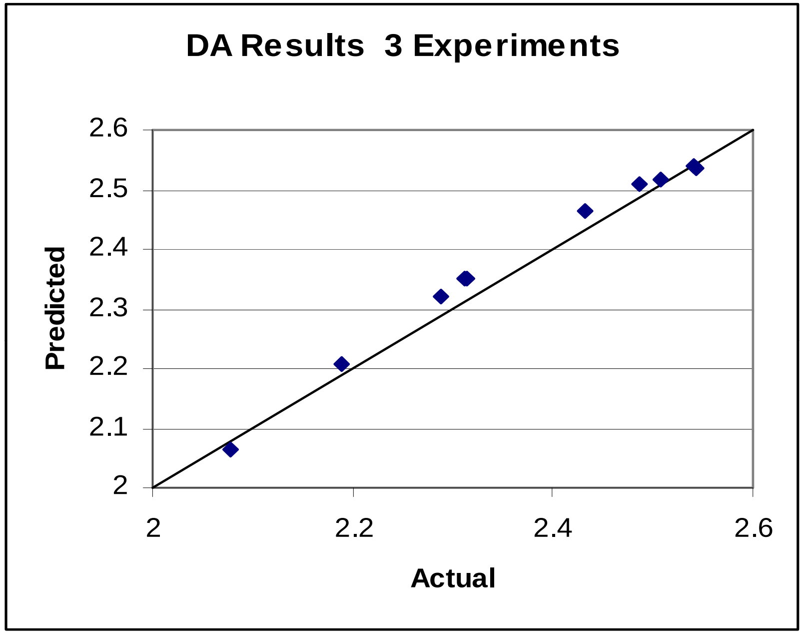
Paper Helicopter-Validation

- Perform some validation “experiments”

Box-Behnken Design
- 13 experiments
- 6 fitted parameters

Simple experiment Dim. Analysis
- 3 experiments
- 2 fitted parameters





- Buckingham’s Pi theorem is not cited in any of the well know texts on response surface design

Concluding remarks

- The application of statistical thinking and statistical methods is highly dependent on the nature of the problem to be solved.
- An understanding of the scientific context of the problem is crucial for statistics to be at its most productive, and most effective (this is much more important than any Bayesian/ frequentist argument – please don't get sidetracked).
- There is a difference between statistical mathematics and statistical science – make sure you know which is which, *and* know what you are or want to be.
- Unless you are very very good, specialize, don't generalize.
- The job of the scientist is to decide not which theory is true, but which theory is more *likely* to be true – make sure that you keep this at the forefront of your thinking.

Appendix: Tim Davis - Career

- *1981* – BSc Statistics, Univ. Of Wales → Dunlop Ltd.
- *1982* – Fellow, Royal Statistical Society (RSS)
- *1985* – Sumitomo Rubber Industries, Japan
- *1986* – Ford Motor Company
- *1988* – Captain's Player Ford Warley CC
- *1989* – Best Fielder Ford Warley CC
- *1991* – PhD (Competing Risks Survival Analysis)
- *1991* – Council member RSS (4 year term; VP '93-'95)
- *1992* – Book (*Engineering, Quality & Experimental Design*) with Dan Grove
- *1992* – Greenfield Industrial Medal, RSS
- *1994* – Chartered Statistician (C.Stat.)
- *1995* – Quality Manager, Ford Werke AG, Köln, Germany
- *1999* – Quality Director, Detroit, USA
- *2000* – Firestone Tire crisis
- *2001* – Henry Ford Technical Fellow for Quality Engineering
- *2004* – Fellow I.Mech.E, and Chartered Engineer (C.Eng.)
- *2005* – Donald Julius Groen Prize in reliability, I.Mech.E.
- *2007* – Quality Director and Board Member – Jaguar Land Rover
- *2010* – Council member RSS, 2^{nd} term