Statistical Engineering

Tim Davis

Royal Statistical Society International Conference
Brighton Centre, UK; Sept. 15, 2010.

What is engineering?

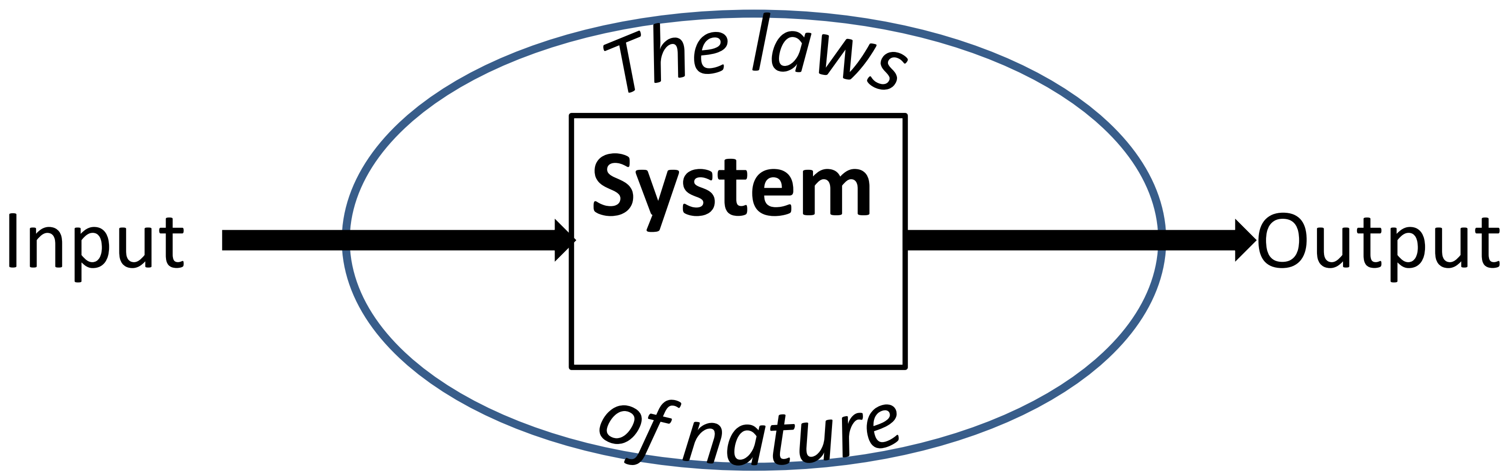
ABET (Accreditation Board of Engineering and Technology):

"Engineering is the profession in which a knowledge of the mathematical and natural sciences, gained by study, experience, and practice, is applied with judgment to develop ways to utilize, economically, the materials and forces of nature for the benefit of mankind."

This definition implicitly captures the interplay between **deduction** (the laws of the mathematical and natural sciences) with **induction** (the use of experience, practice and judgment).

A more formal model

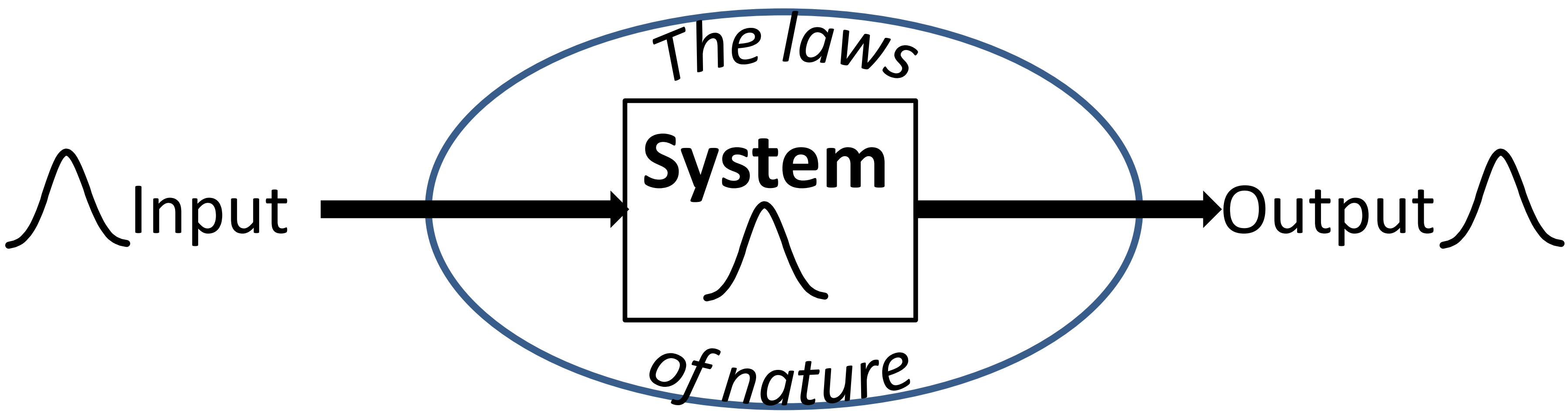
[due to CR Mischke, *Mathematical Model Building*, (1980)]



Given...	To find...	Skill needed	Name of the Game
System, Input, Laws	Output	Deduction	Analysis
System, Output, Laws	Input	Deduction	Reverse Analysis
Input, System, Output	Laws	Induction	Science
Input, Output, Laws	**System**	**Synthesis**	**Engineering**

A more formal model

[due to CR Mischke, *Mathematical Model Building*, (1980)]



Given...	To find...	Skill needed	Name of the Game
System, Input, Laws	Output	Deduction	Analysis
System, Output, Laws	Input	Deduction	Reverse Analysis
Input, System, Output	Laws	Induction	Science
Input, Output, Laws	**System**	**Synthesis**	***Statistical* Engineering**

Deduction & Induction

- **Deduction:** argues from a given theory's general principles to a specific case of expected data.
- **Induction:** argues in the opposite direction to deduction, from actual observed data, to an inferred model, from the specific to the general.

If we let **H=hypothesis** & **D=data**, then

- Deduction can be thought of **Pr(D|H)**; this probability has a ***frequency*** interpretation due to randomness.
- Induction can be thought of as **Pr(H|D)**; this probability has a ***degree of belief*** interpretation, due to lack of knowledge.

5

Deduction & Induction

e.g. H= the coin is fair; D=40 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*

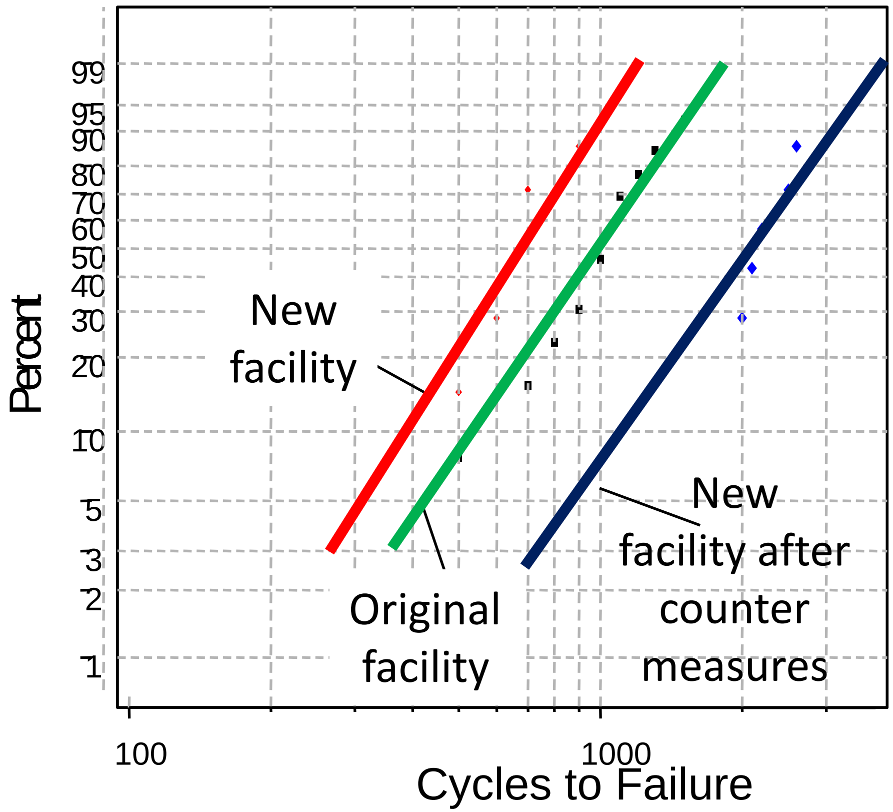
An engineering example

- An established vehicle design produced in a new manufacturing facility suffered an unusual, high severity structural welding failure ⅔ of the way through a vehicle durability test
- Subsequent lab test results (data=**D**) 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=**H**).
- In order to authorize production, do we need to evaluate $\Pr(\mathbf{D}|\mathbf{H})$ or $\Pr(\mathbf{H}|\mathbf{D})$?

An engineering example - cont

<u>Hypothesis Testing Pr(**D**|**H**)</u>

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



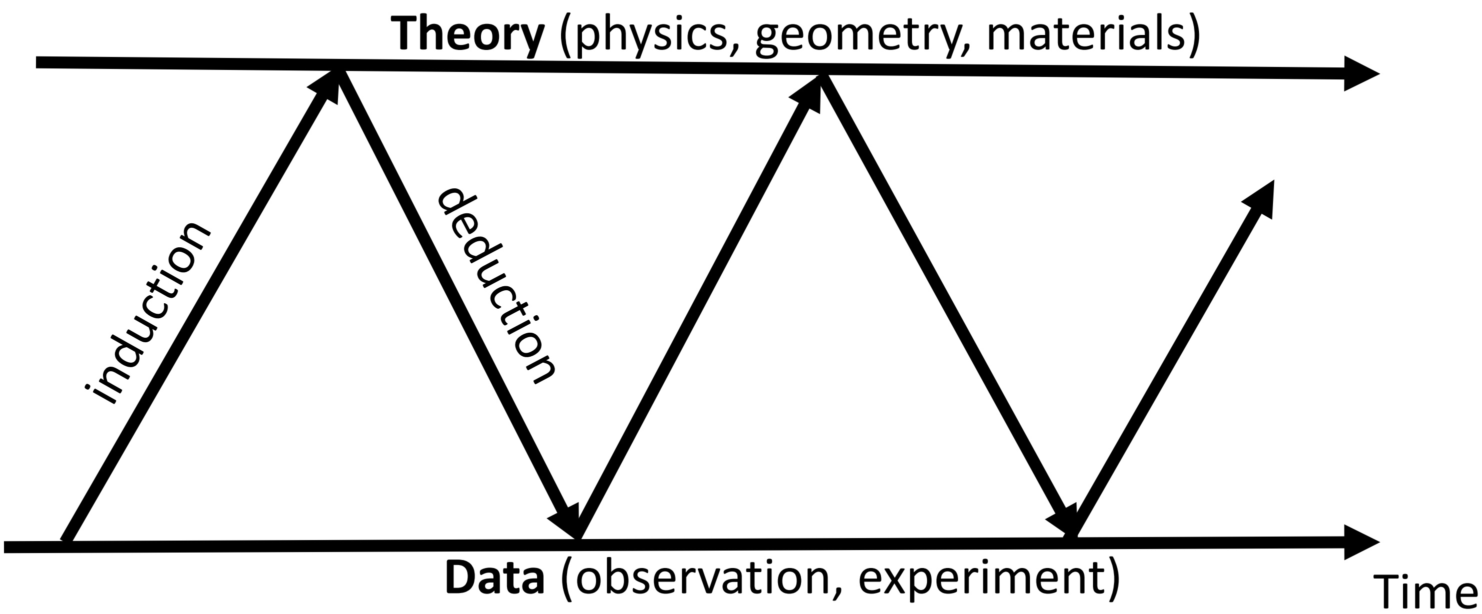
<u>Hypothesis Generation Pr(**H**|**D**)</u>

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

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

The iterative learning process

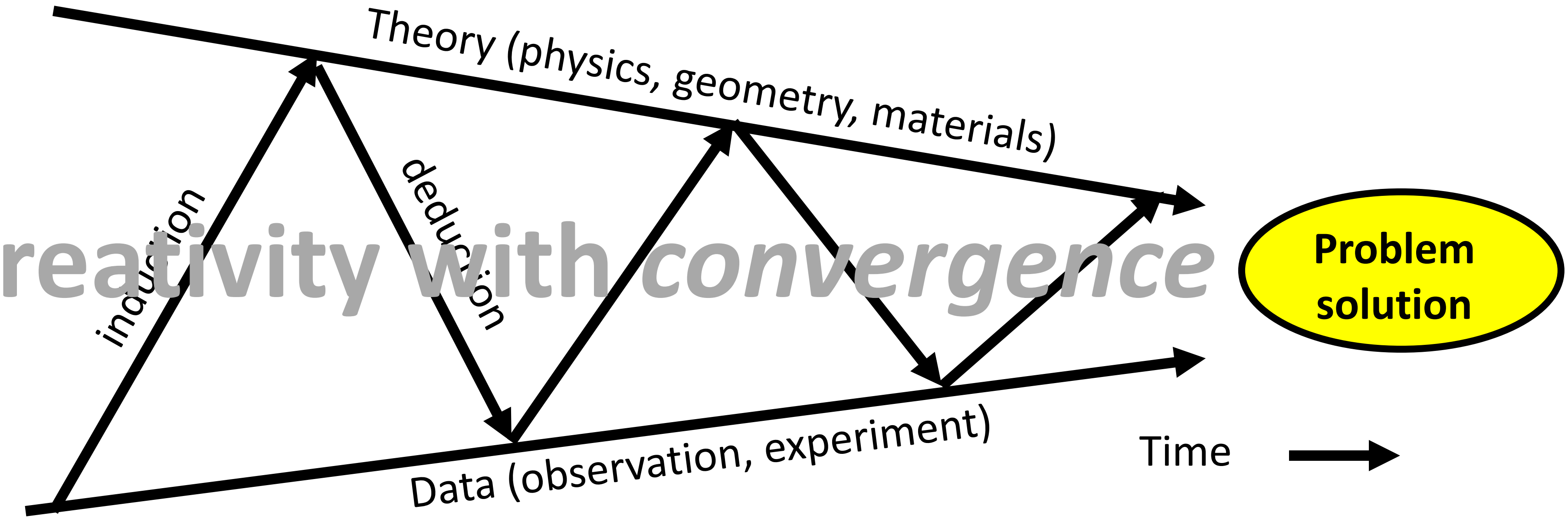
[GEP Box, *Science & Statistics* (1976)]



- The interplay between deduction and induction provides the synthesis for engineering to combine analysis with science.
- Speed & scale of this process determines what sort of statistical approach is required (engineering is usually quick).

The iterative learning process

[TP Davis, *Science, Engineering, & Statistics* (2006)]



- The interplay between deduction and induction provides the synthesis for engineering to combine analysis with science.
- Speed & scale of this process determines what sort of statistical approach is required (engineering is usually quick).
- **It is the job of the statistical investigator/collaborator to a) encourage creativity *and* b) to ensure convergence.**

Things that stand in the way of convergence

- Getting stuck on the deductive leg when an empirical approach is required and vice versa.
- Initiatives such as Six Sigma have not helped; they have failed to teach the distinction between induction and deduction.
- Data thrown into computer packages and groping around in the output looking for significant “p-values”
- Complicated theories put forward to fit the facts “There are multiple root causes for this problem” – the engineering equivalent of a conspiracy theory.
- Too much data collection/analysis devoted to eliminating “root cause theories” that, through deduction, can be shown not to be true, which slows down progress in solving the problem.

A key method to aid convergence

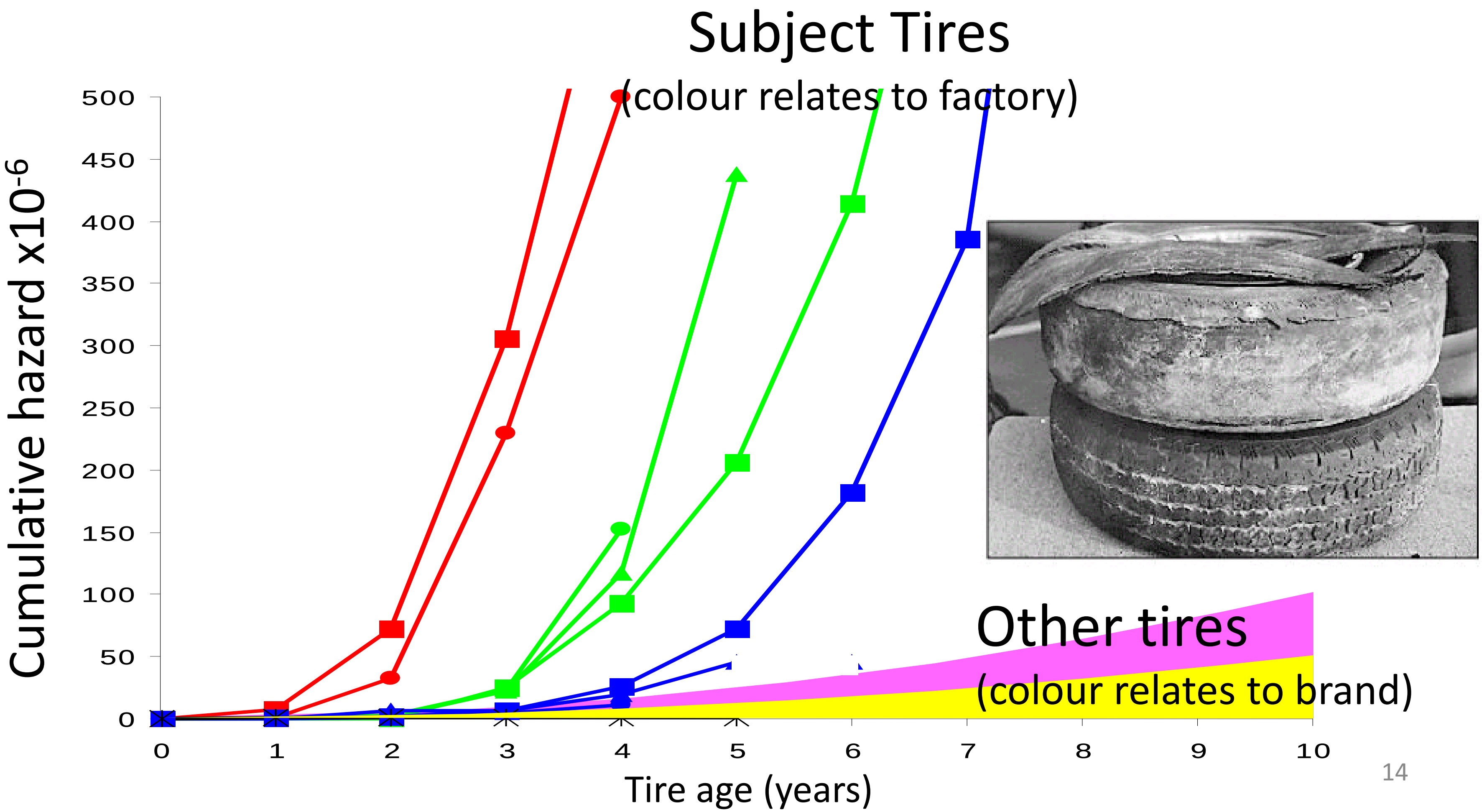
- The “IS” / “IS NOT” Matrix
- Define some criteria
  - what is the defect?
  - when did we first observe the defect?
  - where did we first observe the defect?
  - what is the pattern or trend in the data?
  - etc...
- Relative to these criteria, ask what the problem “IS”;
- Then ask what the problem logically *could have been*, but in fact “IS NOT”.
- Use these contrasts to *filter out* the candidate root cause theories that fail to explain the set of contrasts.
- Only test or experiment with theories that cannot be eliminated in this way

IS/IS NOT example – Firestone tire crisis

PROBLEM Vehicles suffer tire failure and then 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?	Certain Firestone Tires	Other Tire brands on the *same* vehicle	-	+
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 of the US	-	+
What is the trend in the defects	Tires from one factory have a higher failure rate than from another	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	?	+

Some statistical methods employed to converge to a solution

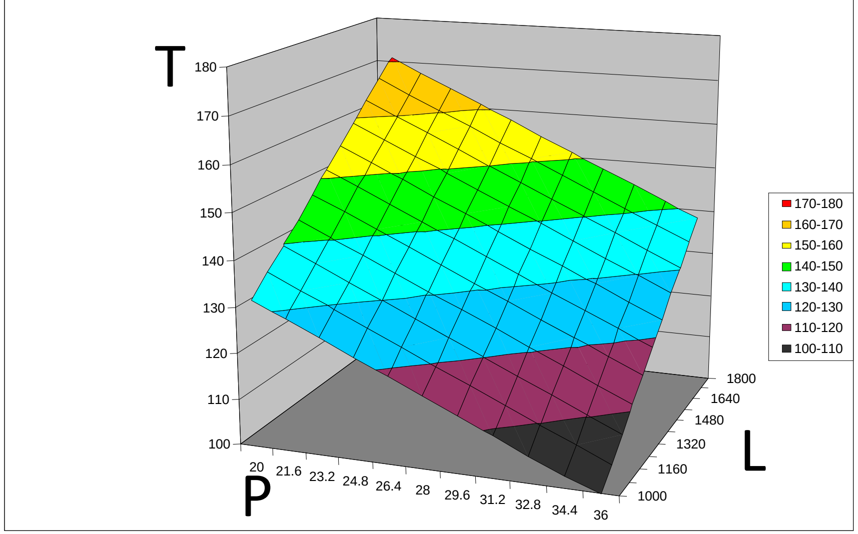
Cumulative hazard analysis → IFR. Note differences between factory of origin for the same tire type.



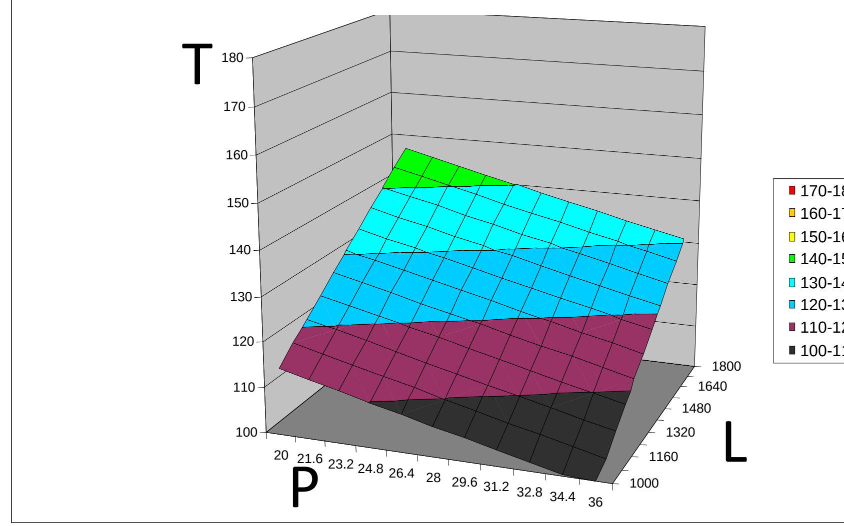
Some statistical methods employed to converge to a solution

Central Composite Response Surface design – to measure the temperature (T) robustness of the subject tires to pressure (P), load (L) and speed (S)

Subject Tires



Other Tires

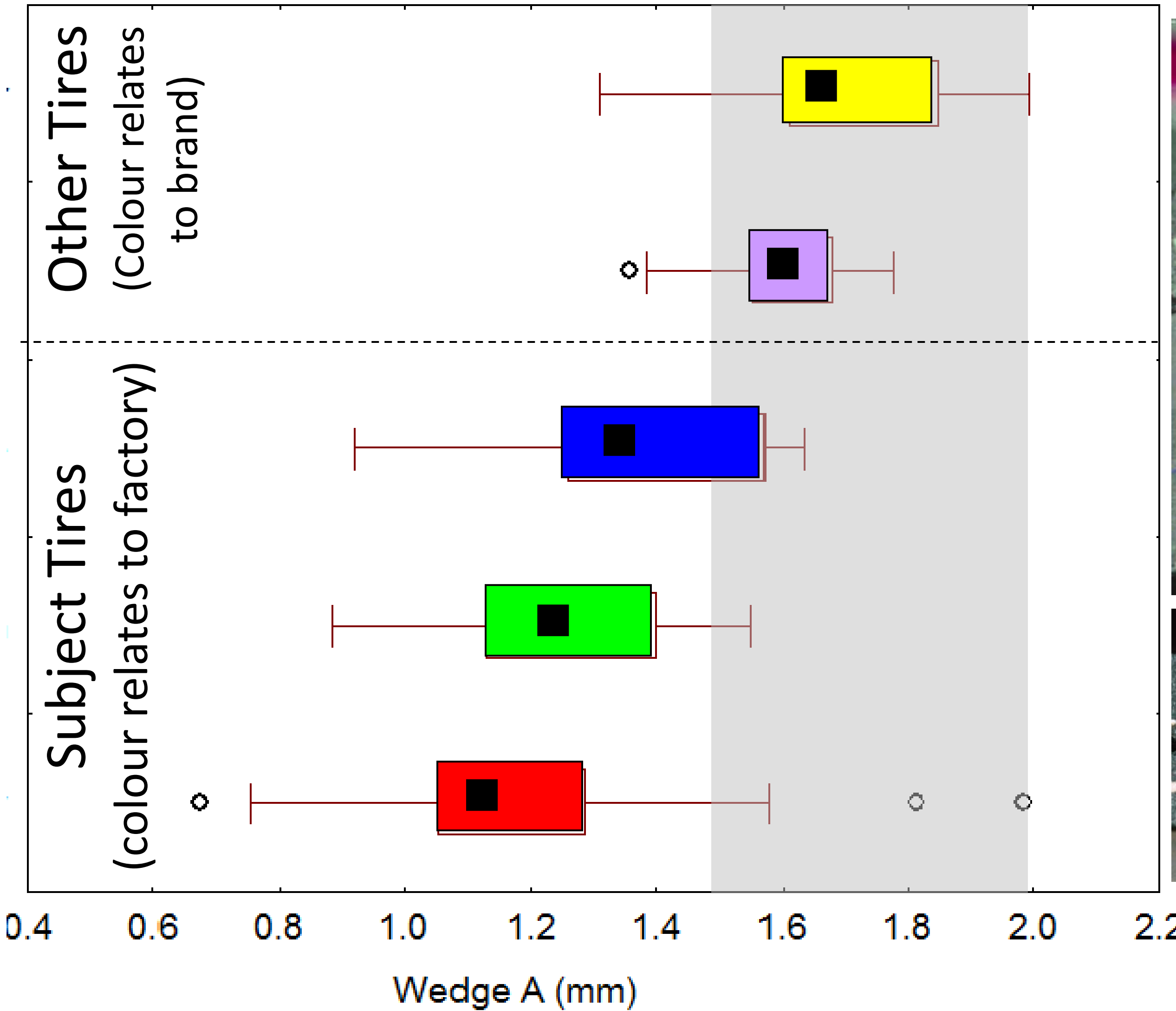


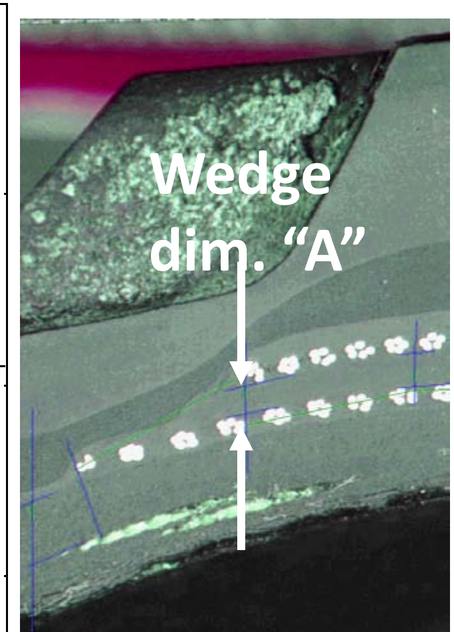
$\mathbf{T=144+8}\boldsymbol{S}\mathbf{+20}\boldsymbol{L}\mathbf{-14}\boldsymbol{P}$

$\mathbf{T=124+7}\boldsymbol{S}\mathbf{+16}\boldsymbol{L}\mathbf{-8}\boldsymbol{P}$

Some statistical methods employed to converge to a solution

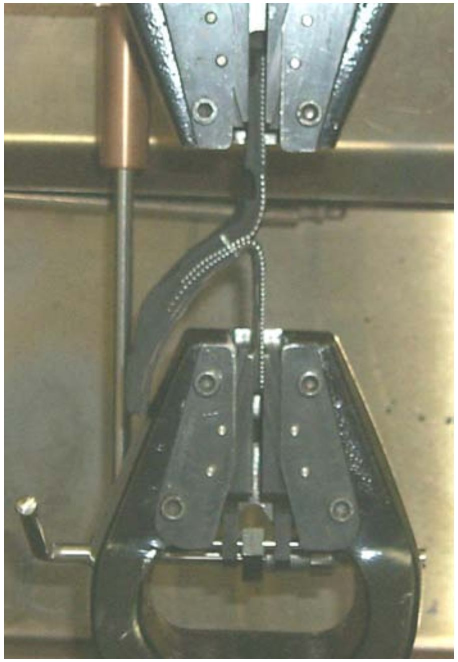
Exploratory Data Analysis with Box plots

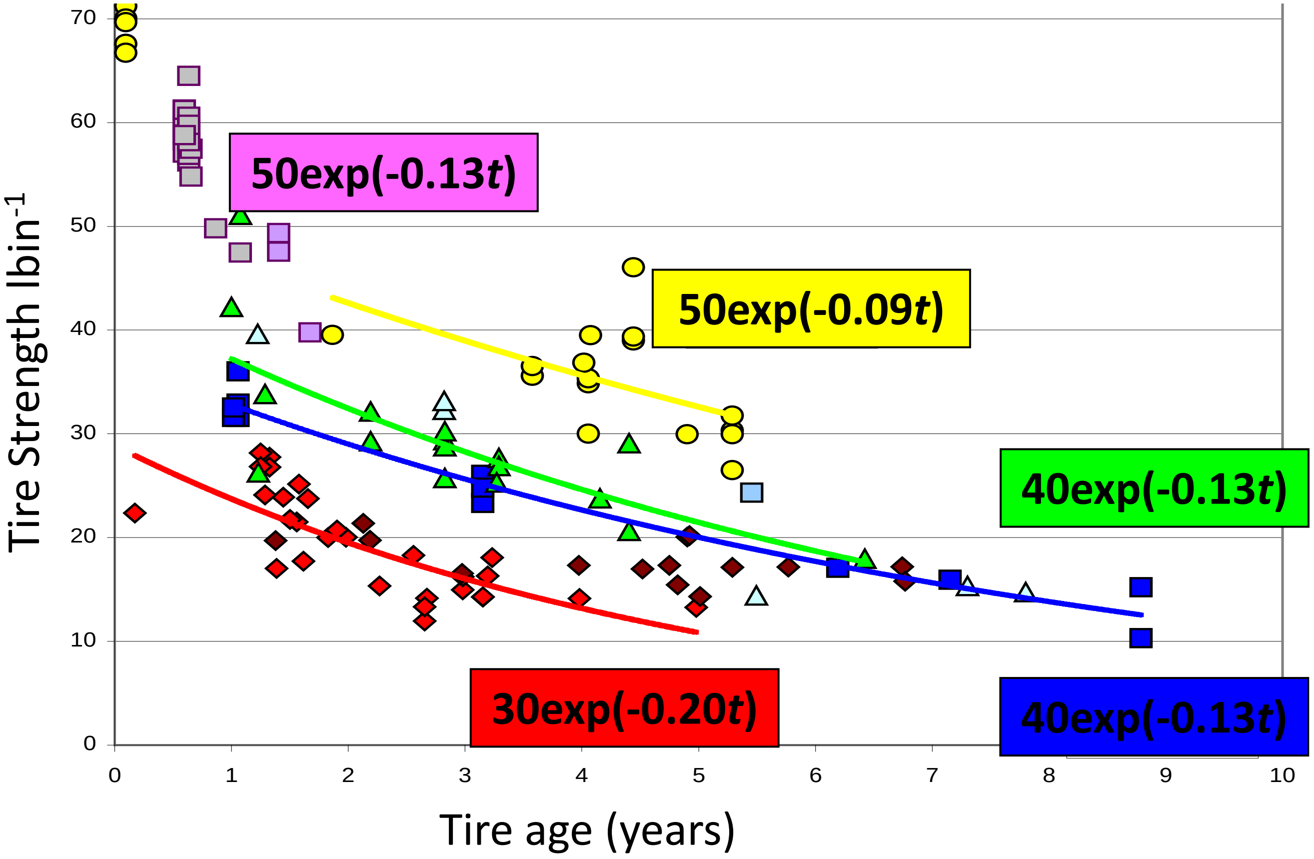




Some statistical methods employed to converge to a solution

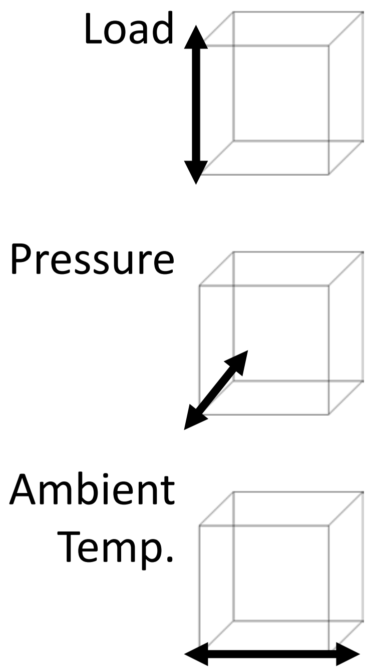
Regression analysis to model tire bond strength decay as a function of tire age

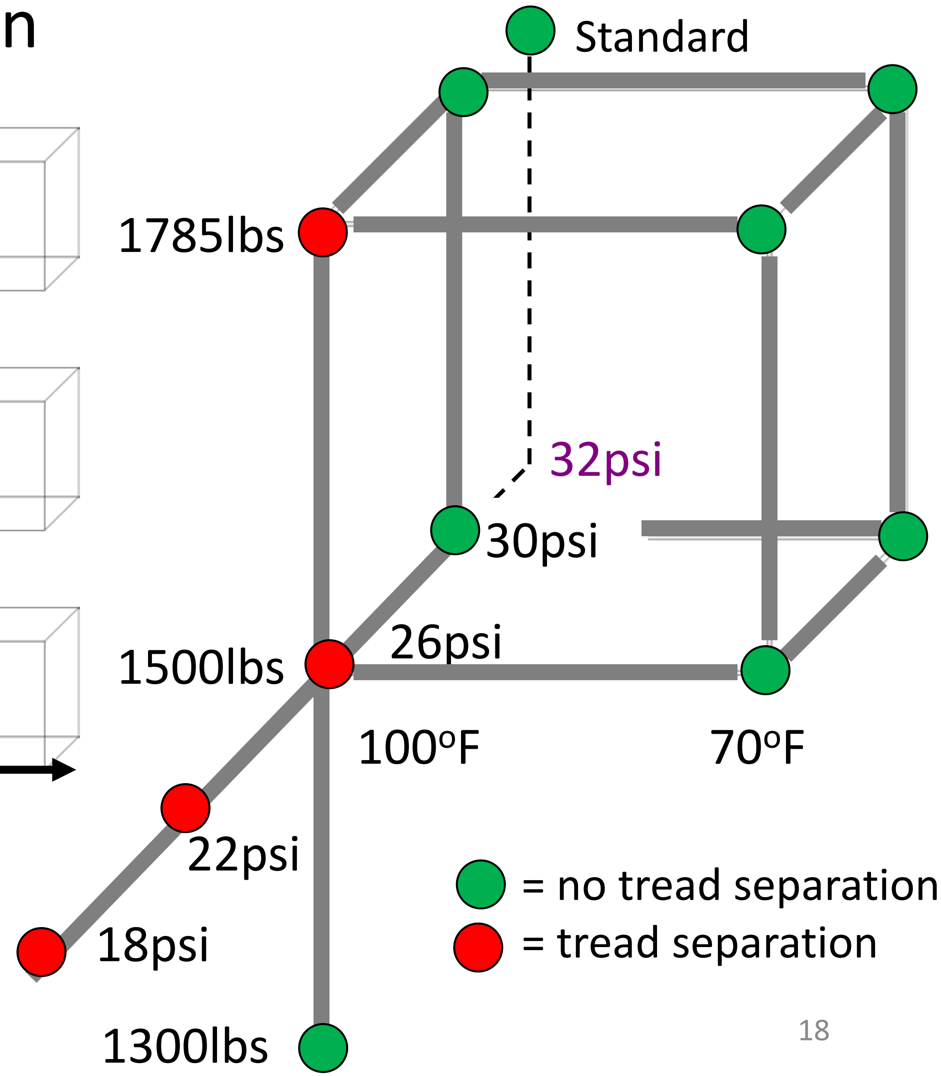


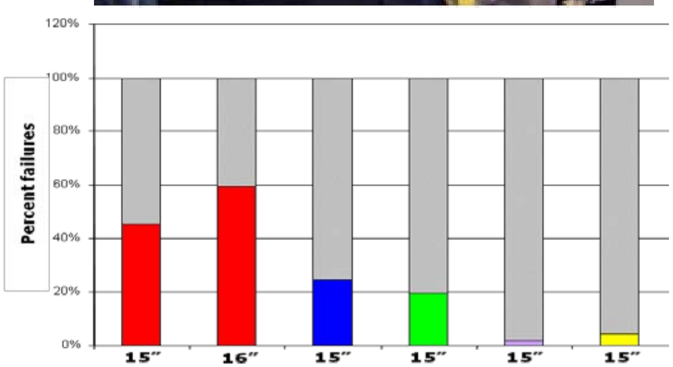


Some statistical methods employed to converge to a solution

Factorial design – to develop a lab test to replicate the failure, and the failure pattern

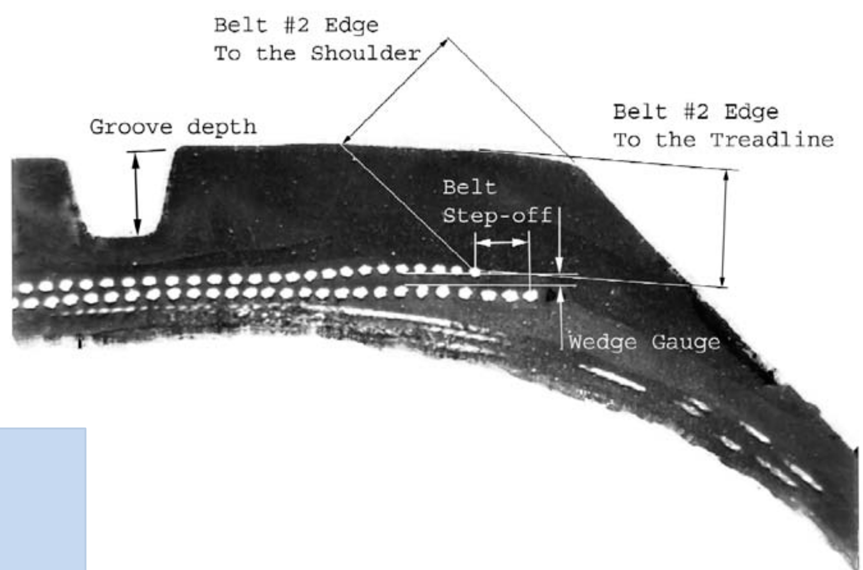




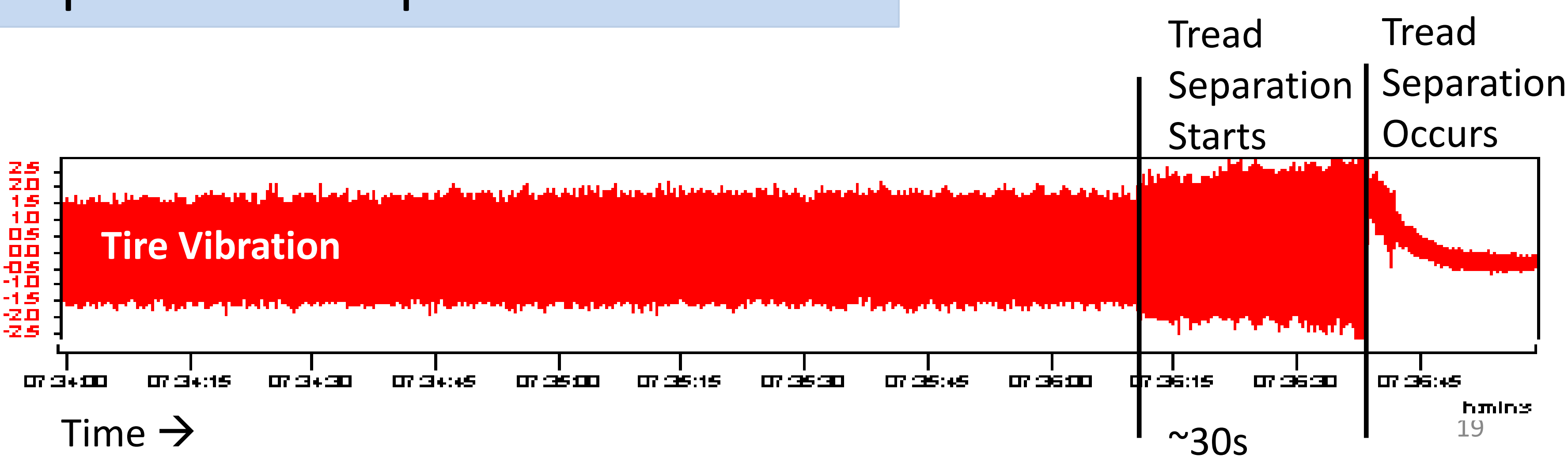


Some statistical methods employed to converge to a solution

Competing Risk Proportional Hazard regression – to test the root cause theory (can we turn the problem "on" and "off"?).

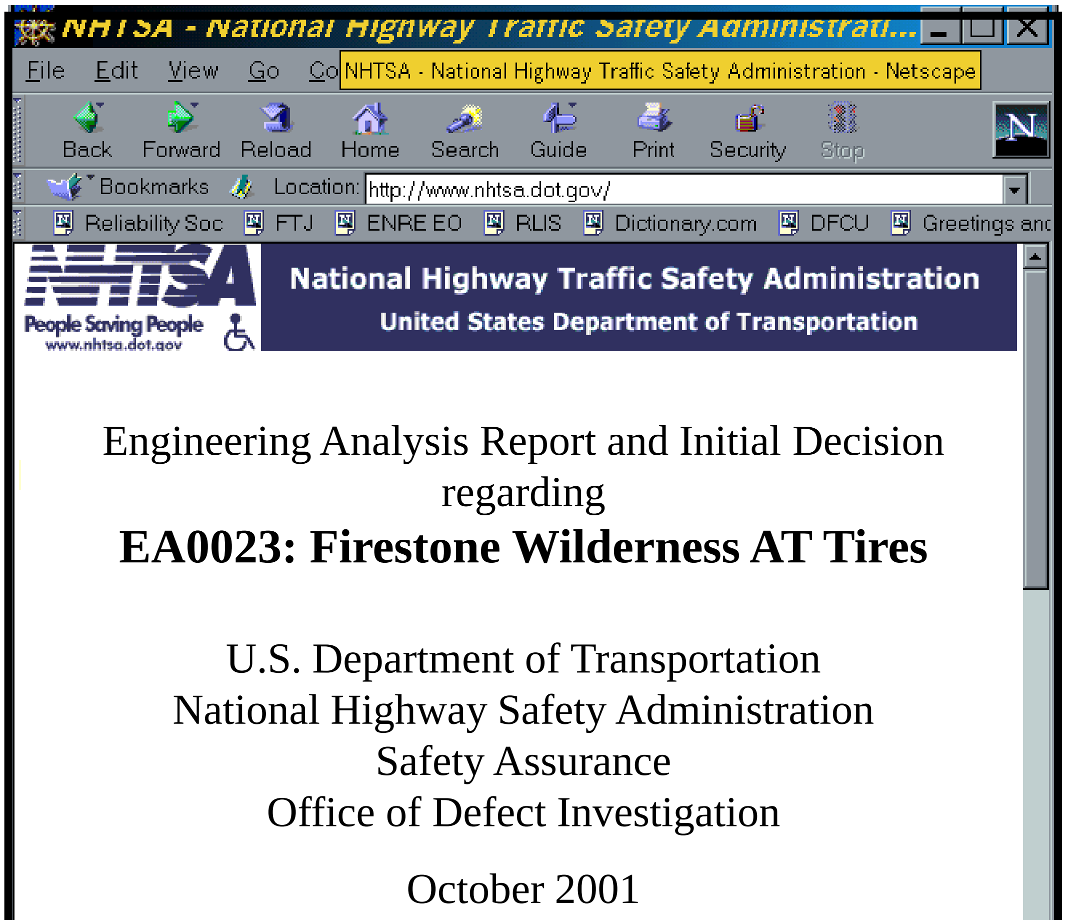


Problem: How can we measure tire parameter's prior to failure?



Some statistical methods employed to converge to a solution

The recall decision was made to replace 20 million tires ($3Bn) *before* the authorities asked us to do it



"… the set of **cumulative hazard function** curves for the recalled tires… demonstrate that if they are not removed from service, the focus tires from these plants – … will experience a similar increase in tread separation failures over the next few years.…"

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

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 usually interpreted 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, avoid mistakes, engineer and evaluate counter measures against a range of conditions**
  - This is recognised as an analytical problem
  - Key tool is Failure Mode & Effect Analysis (FMEA - barely referenced in reliability textbooks)

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

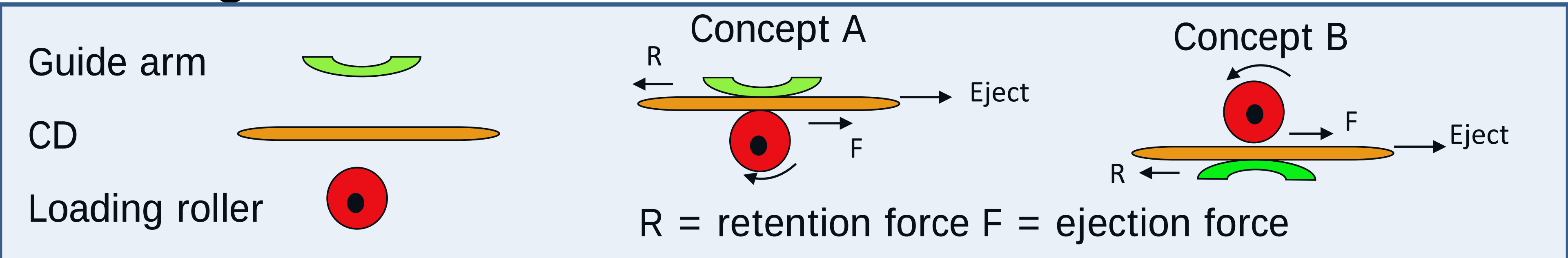
  "What is the cause of management's fantastic faith in the machinery?")

Reliability as Failure Mode Avoidance

- 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.

Mistake avoidance example

CD changer in a car



- 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**.
- The job of an engineer is to choose the design that will fail the least, ***not*** to predict how often the chosen design will fail.
- Reliability effort is best placed ensuring Concept B is chosen, rather than trying to predict how often Concept A will fail.

Robustness

- Robustness = product & process performance that is insensitive to disturbances.
- Concept popularized by G Taguchi, but the idea extends back at least as far as Jim Morrison (1957).
- 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 design configurations?

Measuring robustness

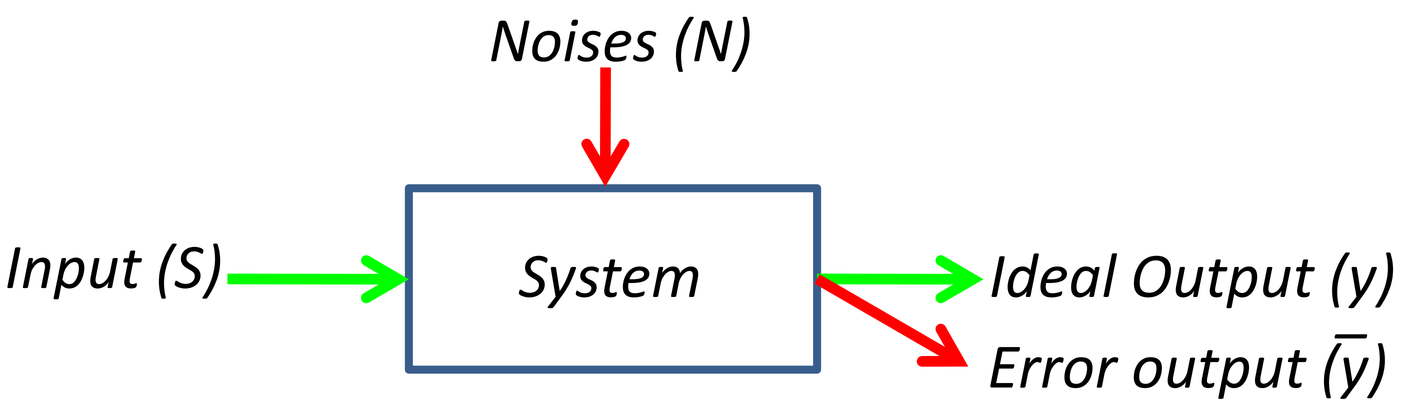
- To answer Q1, Taguchi used a signal to noise ratio:-
  S/N = $\log(\mu/\sigma)$

  μ=average product performance;

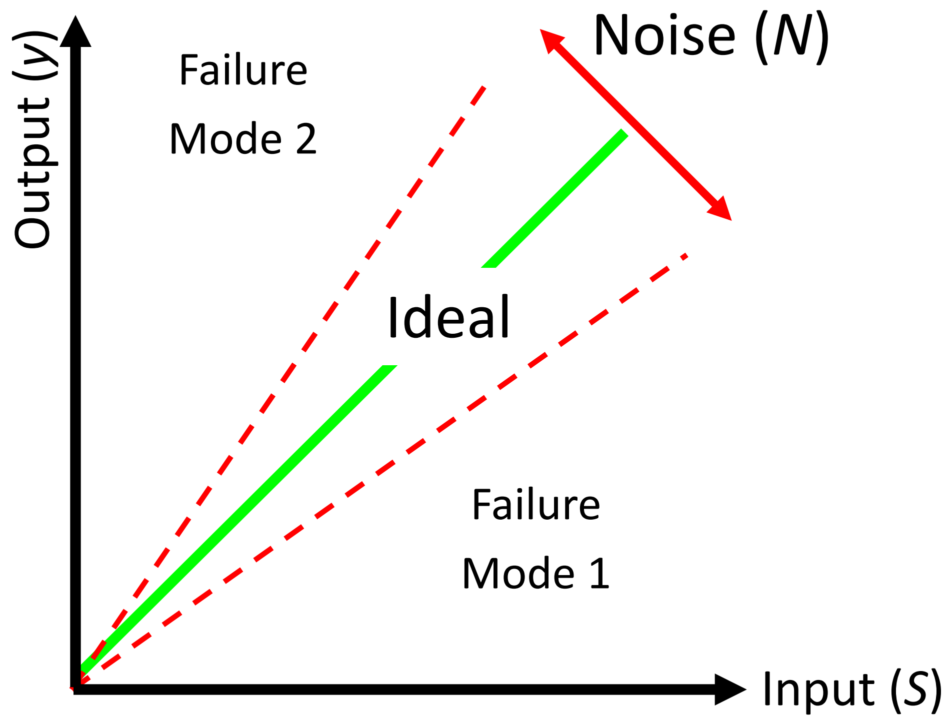
  σ = variation in performance induced by noises.
- Much controversy ensued in the statistical literature, in conferences, and 1-1 conversations ...

Measuring robustness

- Engineering function can be formulated as transforming or transporting ***Energy***, ***Material***, and ***Information*** from an input state (signal) to an output state (Pahl & Beitz, 1996)
- Since these are conserved quantities, the basic transfer function between the input & output will be linear
- But the system will be attacked by noises, which will disturb the function and create an error output which could be considered a failure mode (i.e. a counter measure will be required)



Measuring robustness



"Ideal" Function:

$y = \alpha_0 S.$

$\longrightarrow$

"Noise Disturbed" Function:

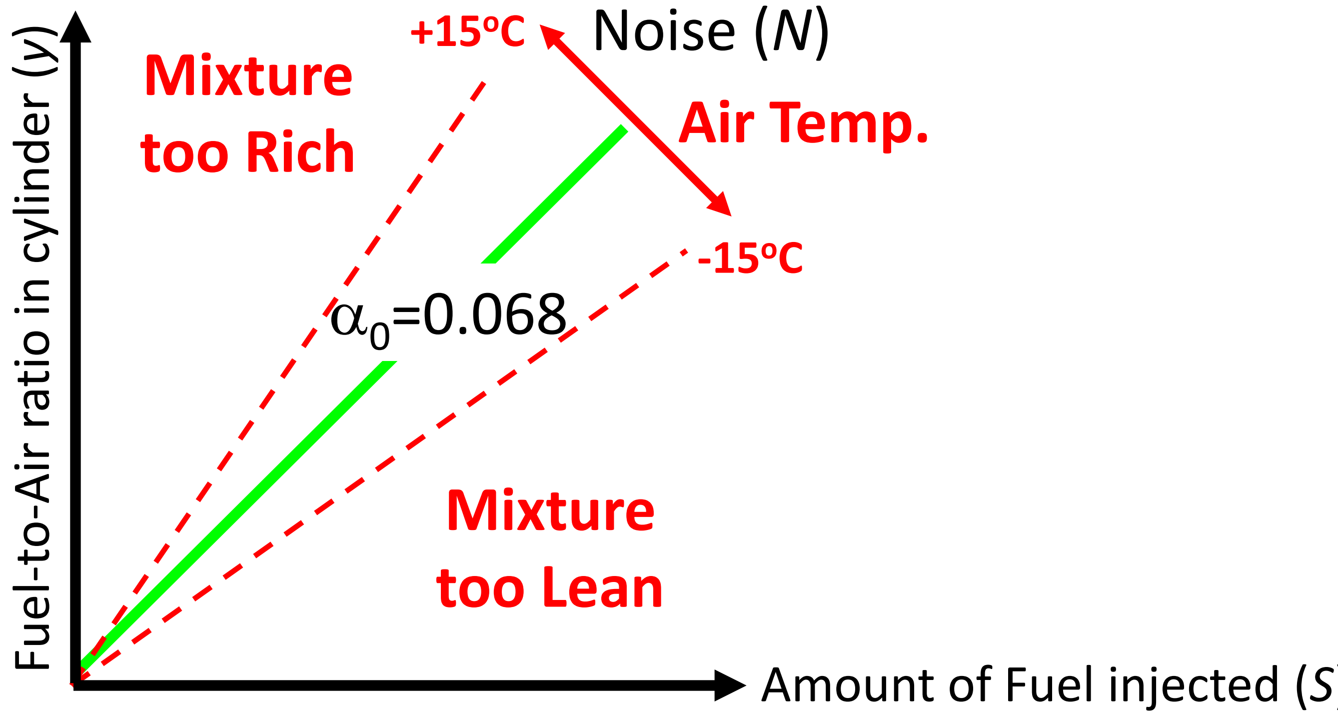
$y = \alpha_0(1+\alpha_1 N)S.$

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

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

Example – engine starting

<u>Objective:</u> achieve fast start times in spite of ambient air temperature

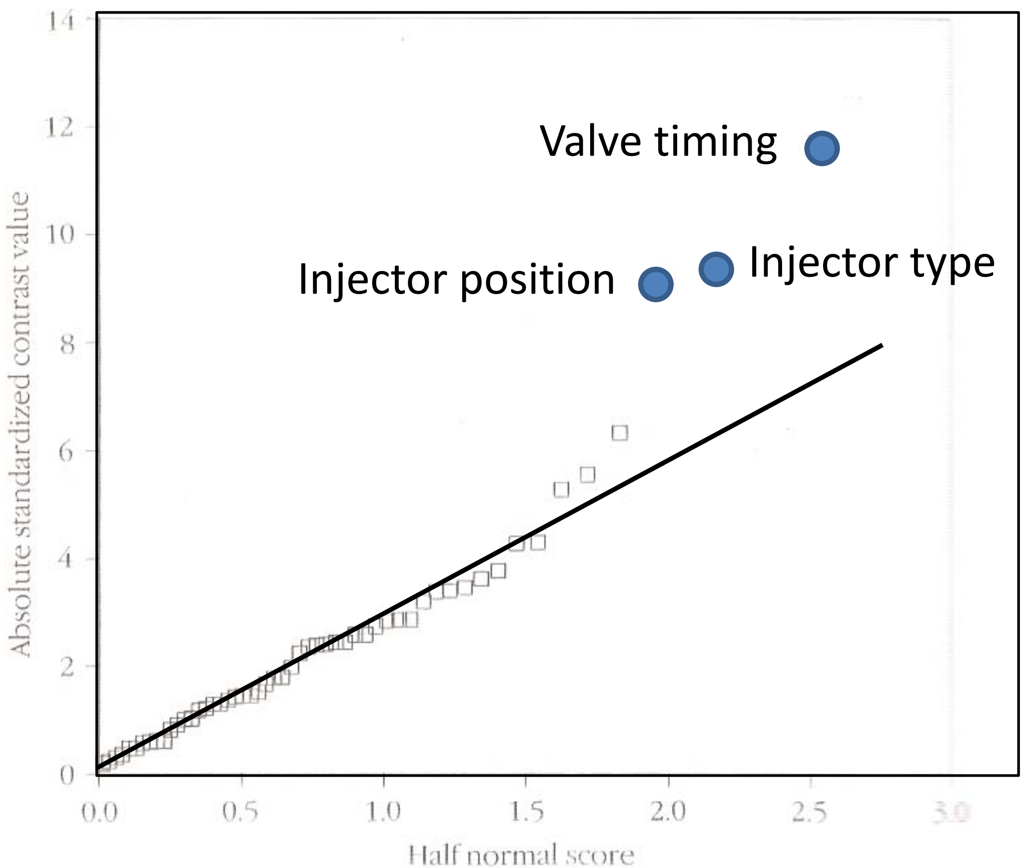


"Ideal" Function: $y=\alpha_0 S$. → "Noise Disturbed" Function: $y=\alpha_0(1+\alpha_1 N)S$.

Robustness can be achieved by experimenting with engine design parameters to minimize α_1 in the presence of the noise.

Example – engine starting

Factorial experiment in 7 design parameters related to engine configuration, run against a background of noise (air temperature).



- ½ Normal plot of factorial effects on the robustness parameter α_1
- 3 of the design parameters can be used to achieve robustness (i.e. low start times in spite of ambient air temperature)
- Details in Grove & Davis (1992) and Davis (2004)

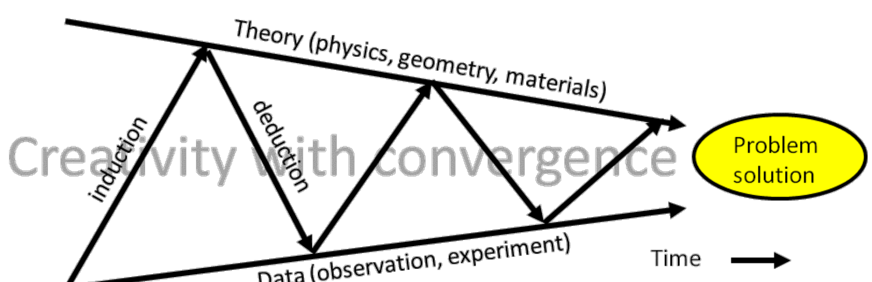
Engineering experiments

- Engineering experiments are often quick to run
- Graphical methods favoured approach to analysis (e.g. ½ Normal plots preferred to ANOVA tables)
- Because of the sequential nature of experiments, randomization is often not possible
- Saturated (and sometimes super-saturated) designs are quite common – confounding issues can be followed up with secondary experiments
- Inclusion of Control (*C*) and Noise (*N*) factors common – robustness is achieved by discovering interactions between *C* and *N*

$y=\alpha_0(1+\alpha_1 N)S$; if $\alpha_1=f(C)$, then $y=\alpha_0(1+f(C).N)S$

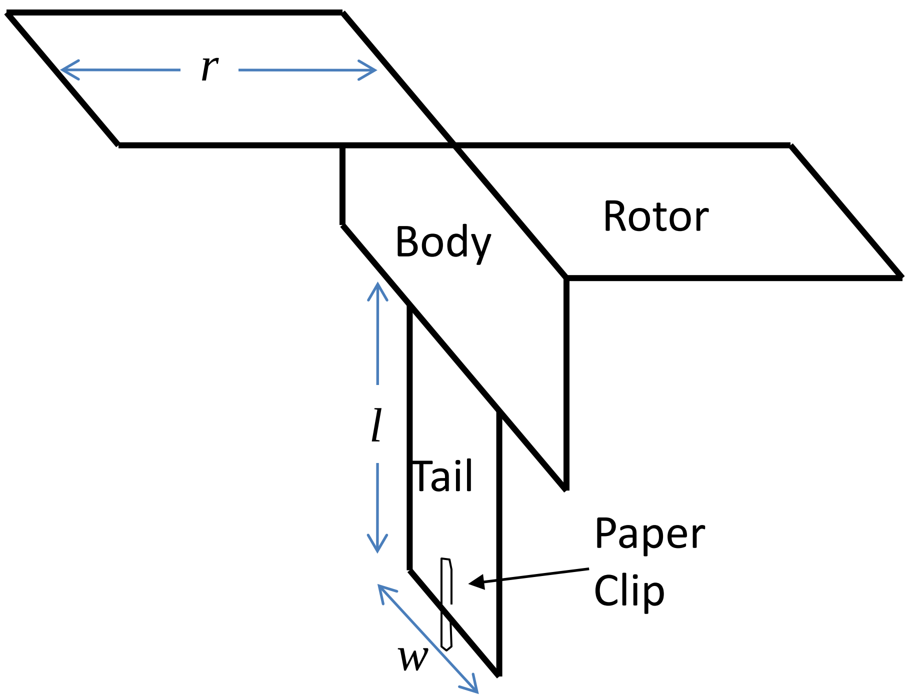
Dimensional Analysis

- Buckingham's Pi theorem (1914): A functional relationship in *n* variables and *m* fundamental units can be rewritten in terms of $N \geq n\text{-}m$ dimensionless variables.
- This extremely useful theorem can be used 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

Predict the flight time, T, of the helicopter



Typical factors that might be used in a response surface experiment (assume all other variables are fixed):

- Rotor radius (r)
- Tail length (l)
- Tail width (w)

$T=f(r, l, w)$

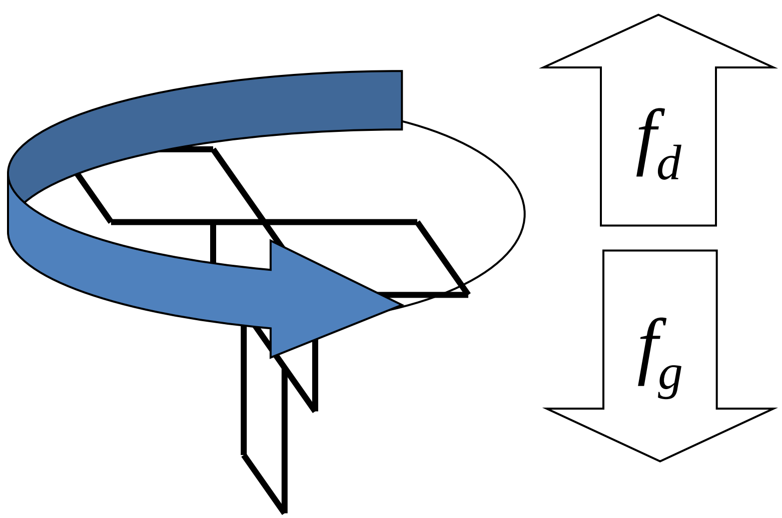
We could approximate this with a 2nd order response surface which would need ~15 runs to estimate.

$$T = -0.03 - 0.01l - 0.01w + 0.41r - 0.02r^2 - 0.002l.w + 0.001l.r + 0.001w.r$$

Dimensionally consistent only by virtue of linearity

Paper helicopter physics

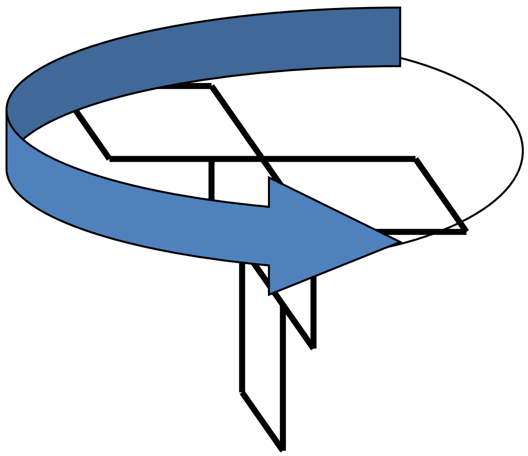
- Time of flight (T) is determined by v and the launch height (h)
- The helicopter very quickly comes to a steady state velocity (v)
- v is determined by the balance between the force of gravity f_g and drag f_d
- f_g is determined by the mass of the helicopter ($m=f(r,l,w)$) and acceleration due to gravity, g.
- f_d is determined by the area swept out by the rotor radius (r) and the density of air (ρ).



Without knowing the exact form of the relationship we can at least say that $T=F_1(m,g,r,\rho,h)$

Paper helicopter physics

- $T = F_1(m, g, r, \rho, h)$
- We know exactly how T depends on h → $T = h/v$
- So we are looking for an expression of the form $v = F_2(m, g, \rho, r)$



Term	Units
$v = h/T$	ms^{-1}
m	kg
g	ms^{-2}
ρ	kgm^{-3}
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 $\Phi_v \equiv v r^a \rho^b g^c; \;\; \Psi_m \equiv m r^d \rho^e g^f$

Analyze the dimensions of the core variables

$$\left[\Phi_v\right] \equiv ms^{-1} m^a kg^b m^{-3b} m^c s^{-2c} = m^{1+a-3b+c} kg^b s^{-1-2c}$$

$$\left[\Phi_m\right] \equiv kg m^d kg^e m^{-3e} m^f s^{-2f} = m^{d-3e+f} kg^{1+e} s^{-2f}$$

Enforce non-dimensionallity → a=- ½; b=0; c=- ½; d=-3; e=-1; f=0

$$\Phi_v = \frac{v}{\sqrt{gr}} = \frac{h}{T\sqrt{gr}}; \Psi_m = \frac{m}{\rho\, r^3}$$

Paper helicopter experiment

- We can now fit a (dimensionless) equation of the form $\Phi_v = F_3(\Psi_m)$
- Choose a small number of experimental runs to fit this equation.
- Change r, l, w, calculate m, measure T, calculate Φ_v and Ψ_m.

Tail Length (l)	Tail Width (w)	Rotor Radius (r)	Φ_v	Ψ_m
5	3.2	12	1.069	1.975
5	3.438	8.744	1.405	3.410
7	5.1	7.62	1.675	4.845

$$\Phi_v = 0.664 + 0.211\Psi_m$$ (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)

$$T = \frac{h}{\sqrt{gr}\left(0.664 + 0.211\dfrac{m}{\rho r^3}\right)}$$

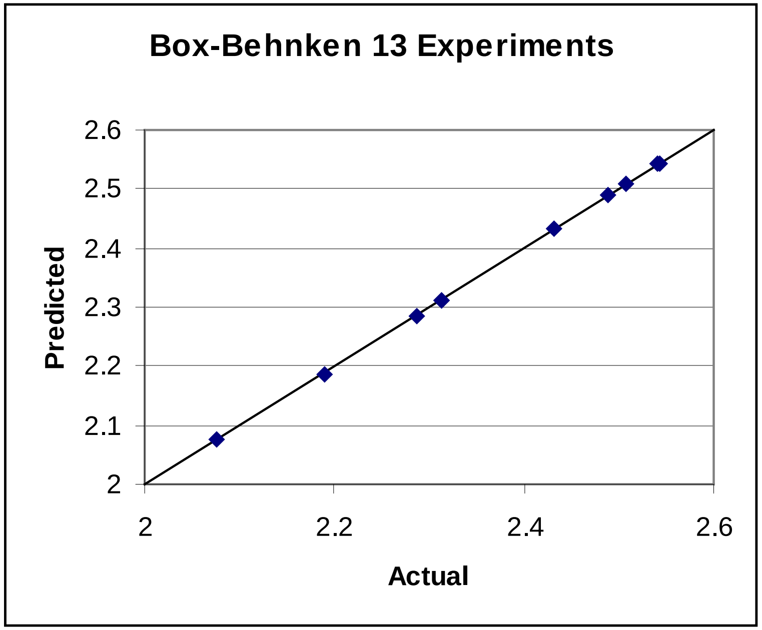
(non-linear and dimensionally consistent)

Paper helicopter-Validation

- Perform some validation "experiments"

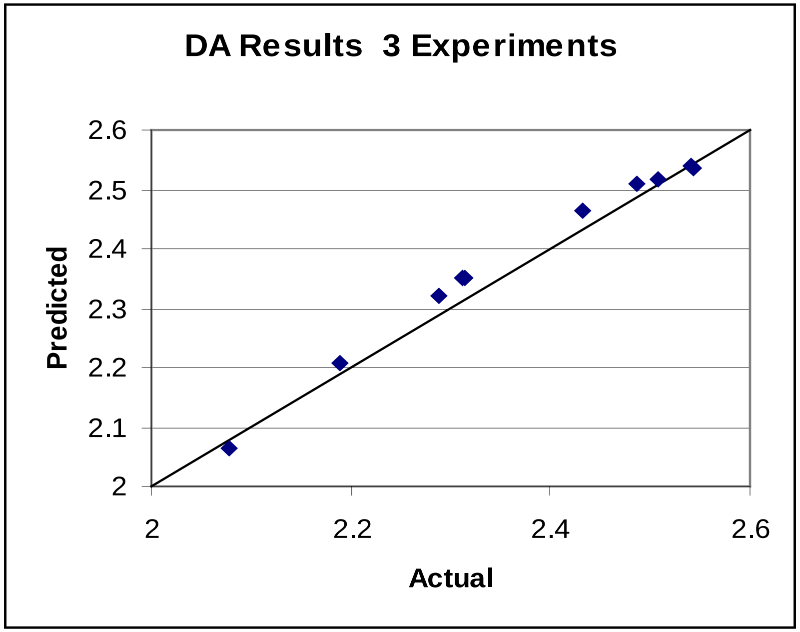
Box-Behnken Design

- 13 experimental runs
- 6 fitted parameters



Simple experiment Dim. Analysis

- 3 experimental runs
- 2 fitted parameters



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

Concluding remarks

- Statistics is the science of making inferences through inductive logic and reasoning in the face of uncertainty.
- In engineering at least, the interplay between deduction & induction is vital in finding solutions to problems
- As statistical collaborators, one of our main contributions is to ensure convergence of investigations.
- Many available training programs haven't helped - too much emphasis on probability theory and a "black box" mathematical approach to analysis; instead of using statistical methods with graphics and use of the "grey cells".
- Engineering and manufacturing industry provides a wealth of interesting challenges for statistics – please get involved!

Statistical Engineering

Tim Davis

Royal Statistical Society International Conference
Brighton Centre, UK; Sept. 15, 2010.

References & Bibliography

- TP Davis (2004). “Measuring robustness as a parameter in a transfer function”. In *Reliability and Robust Design in Automotive Engineering*, ISBN Number: 0-7680-1380-1, Society of Automotive Engineers, Detroit, MI, USA
- TP Davis (2006). “Science, engineering, and statistics”. *Applied Stochastic Models in Business and Industry,* Vol. 22, Issue 5-6, pp401-430.
- VV Krivtsov, DE Tananko, and TP Davis. (2002) “A regression approach to tire reliability analysis”, *Reliability Engineering & System Safety*, volume 78, number 3, pp 267-273.
- DM Grove and TP Davis. (1992) *Engineering, Quality, and Experimental Design*, Longman Scientific and Technical, Harlow, UK.
- CR Mischke, (1980) Mathematical model building – an introduction to engineering. ISU Press, ISBN: 0 8138 1005 1
- GEP Box, (1976) "Science and statistics". Journal of the American Statistical Association, Vol. 71, No 356, pages 791-799
- Feynman, RP. (1986) "Personal observations on the reliability of the shuttle". *Appendix F, Report of the presidential commission on the space shuttle Challenger accident*. William P. Rogers (Chair), US Government, Washington DC
- Morrison, SJ (1957). "The study of variability in engineering design". *Applied Statistics*, Vol. 6, #2, pages 133-138.
- E Buckingham (1914). “On physically similar systems; illustrations of the use of dimensional equations”. Physical Review, Vol IV, #4, pp345-376.
- WE Deming (1950) *Some theory of sampling*. Dover, New York. ISBN 0-486-64684-X
- G Pahl & W Beitz (1996). Engineering Design – a systematic approach (2nd edition). Springer, London. ISBN 3-540-19917-9.
- C Gearhart (2004). Using dimensional analysis to build a better transfer function. *SAE Technical Paper* 2004-01-1129
- TP Davis. (1998) "The fallacy of reliability prediction in automotive engineering". *Automotive Excellence*, Spring 1998, pages 19-21, The American Society for Quality.