STATISTICS AND ANIMAL EXPERIMENTATION: THE GOOD, THE BAD AND THE UGLY

Dr. Manuel Berdoy   Oxford University Veterinary Services.

OSLO June 09

The Essence of Statistical Thinking, Common Mistakes and Solutions:

1. Why statistics and P Values
2. From P values to Power Calculations
3. Experimental Design & Analyses: controlling variation
4. Recapitulation, Refinement, Reduction

1. Why STATISTICS and P VALUES ?

"Lies, Damned lies and Statistics".   This is not "Statistics", but politics.

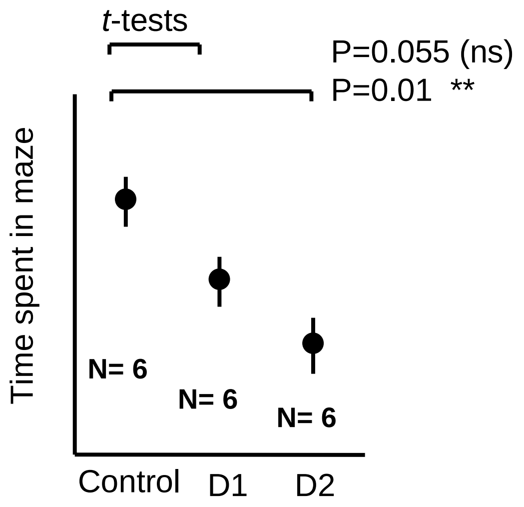
"All statistics has shown us is that most of us have more than the average number of legs"
Factually correct, but "statistics" is not about stating the obvious either.

Statistics is about finding the <u>truth</u> in an <u>imperfect</u> world, and saving time and effort… and (sometimes) lives.

Searching for Truth = Experiment

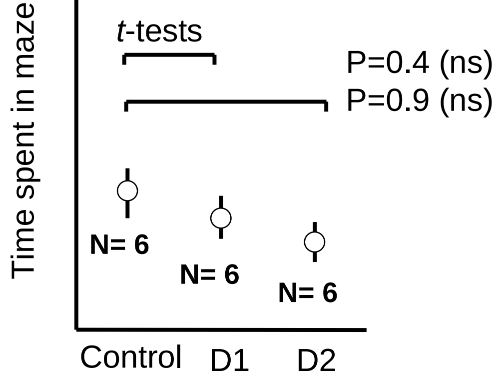
Test of drug doses on cognitive abilities (tested in maze).
<u>**Exp 1**</u>: 3 treatments administered to N=18 male rats, group-housed in 3 cages of 6 rats). **Control** administered to cage 1 (6 rats), **D1** to cage 2,etc.

Conclusion (kind of encouraging)
D1 no effect.
D2 significant effect

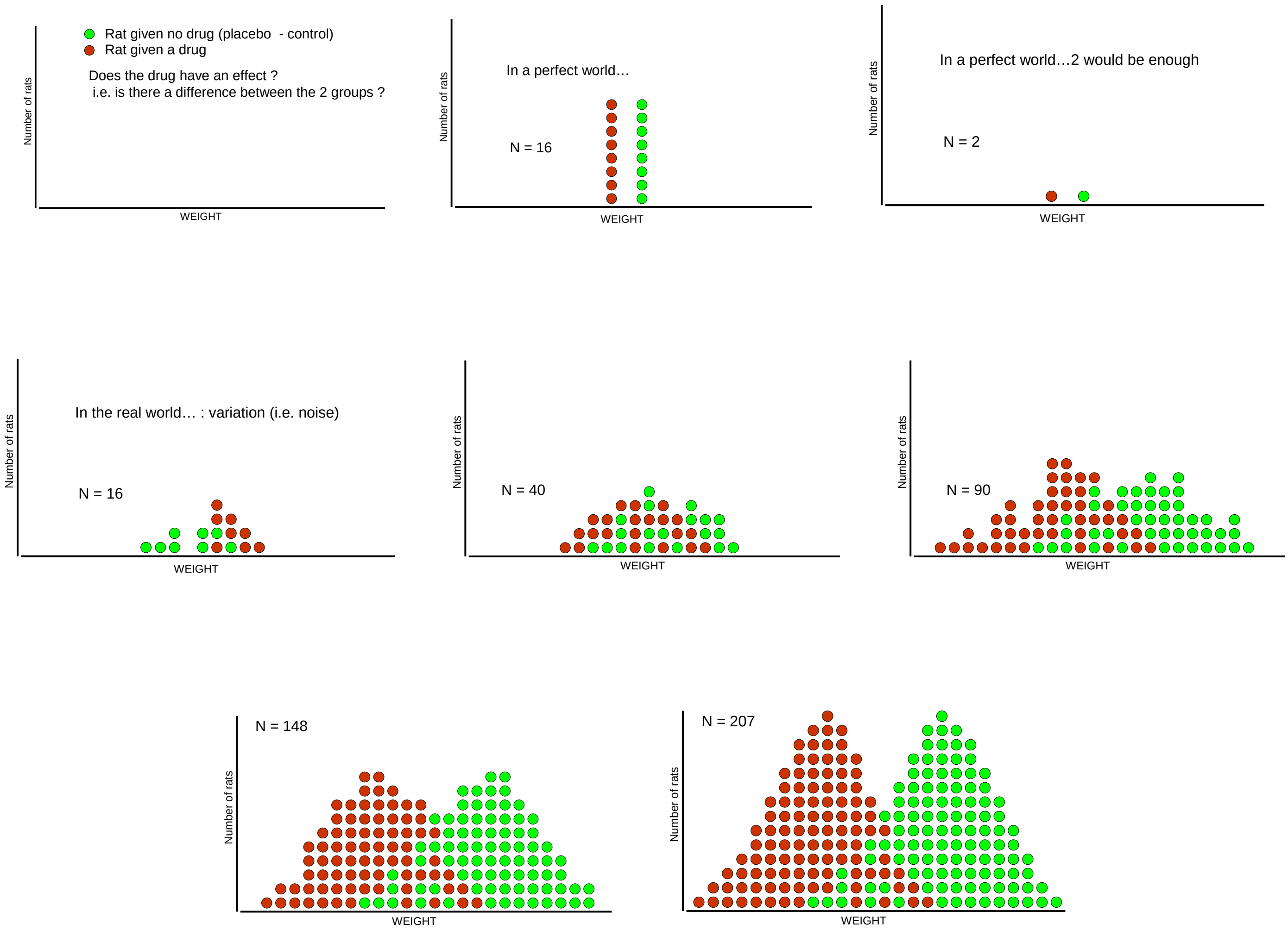


<u>**Exp.2**</u> Repeated on females

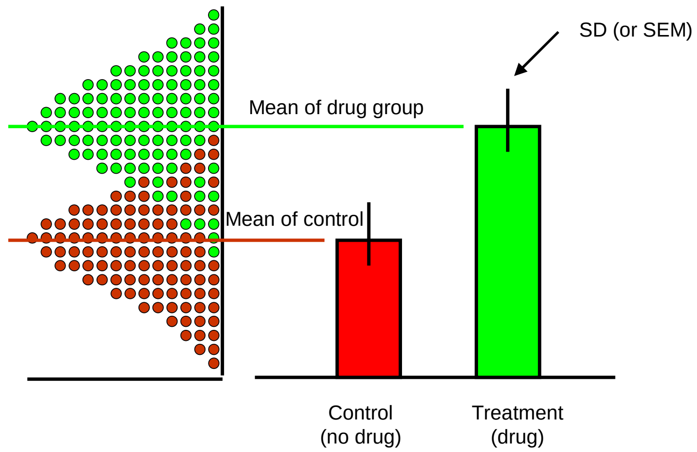
Conclusion:
D1 no effect.
D2 no effect



Searching for "Truth": reminder



Presentation reminder



A question: how many is enough ?...

2. POWER CALCULATION

• Not <u>always</u> realistic

• but we do it to some extent anyway
  • Power calculation is a way of formalising the process
  • Ethical grounds & scientific grounds for it.
  (Animal Welfare regulation, Grant Awarding bodies)

Ethical + Scientific Issues

Ethical, using a medical example:
If test of new drug will have adequate power with a sample of 100 patients, then inappropriate to use 200.
(equivalent for animals under the Act, since by definition painful).

Scientific + ethical:
Conversely,
If test of new drug requires 200 patients to yield adequate power, then inappropriate to use 100. Patients accept to be part of the study on the assumption that it will yield useful results.
(animal equivalent is that no result due to lack of power is a waste of animals).

Some Reminders

HO= Nul hypothesis = Nil hypothesis = no effect

"State of Nature"	Reject HO (**Find effect**)	Accept HO (**Find no effect**)
No effect (HO is true)	Type I error Alpha p value	CORRECT
Effect (HO is false)		

The prob. that exp. will give a false positive result (e.g.due to random fluctuations)

IT DOES HAPPEN !

Some Reminders

HO= Nul hypothesis = Nil hypothesis = no effect

"State of Nature"	Reject HO (**Find effect**)	Accept HO (**Find no effect**)
No effect (HO is true)	Type I error Alpha p value	CORRECT
Effect (HO is false)	CORRECT	Type II error Beta (1- power)

Prob. of detecting a specified effect at specified significance level

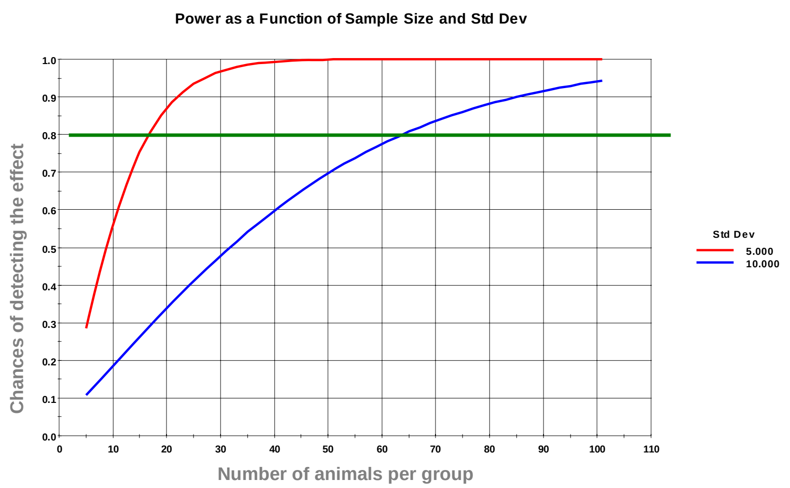
The 6 variables "determining" the chance of statistical significance

• Significance level **= Chance of False <u>Positive</u>**
  [arbitrary, set at P= 0.05 min]
• Desired Power of experiment **= Chance of False <u>Negative</u>**
  [arbitrary, set at 0.80 - 0.90]
• Alternative Hypothesis (1 vs 2 tailed)

• Size of the effect of biological interest **(= Signal)**
• Variation (i.e. Standard Deviation) **(= Noise)**
• Sample size (N)

Note: this is a closed system, i.e. fix any five and the sixth can be derived

Some examples: using the Software "Power and Precision" (www.PowerAnalysis.com)
(see also www.vet.ox.ac.uk/training/mod5links.html)



Example of designs:

- •Factorial
- •Randomised Blocks
- •Latin Square
- •Cross-Over
- •Repeated Measure
- •Covariance

Two common mistakes in biomedical-literature:

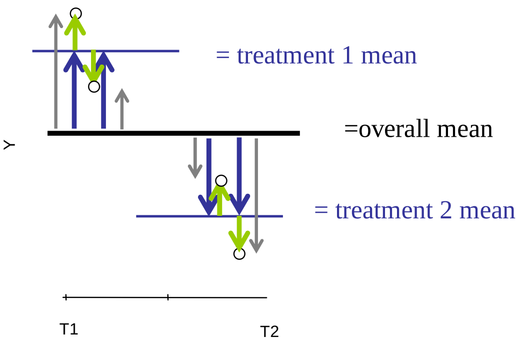
- multiple t-tests
- no blocking

N= 27 rats, assigned to 3 levels of treatment

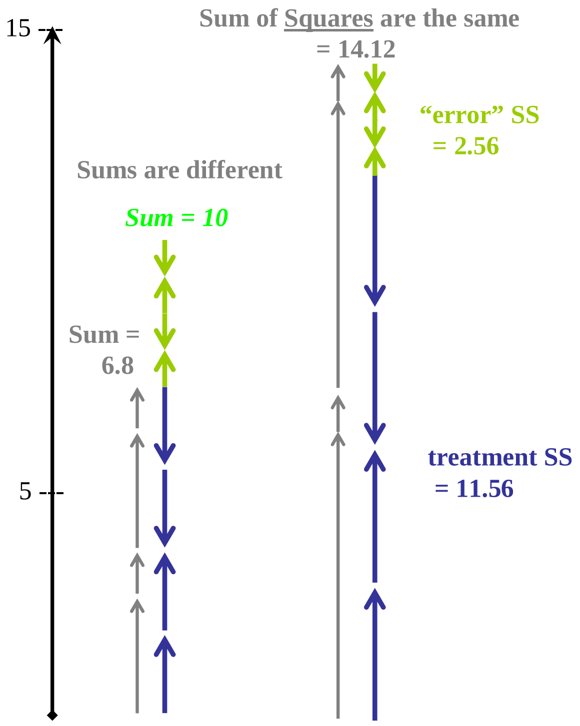
ANalysis Of Variance (ANOVA) table

Source	DF	SS	MS	F	P
Treatment	2	2861	1430.5	13.13	0.00014
Error	24	2614	108.9		
Total	26	5475			

Why Squares? : A reminder (using a simpler case):



Source	DF	SS
Treatment	1	11.56
Error	2	2.56
Total	3	14.12



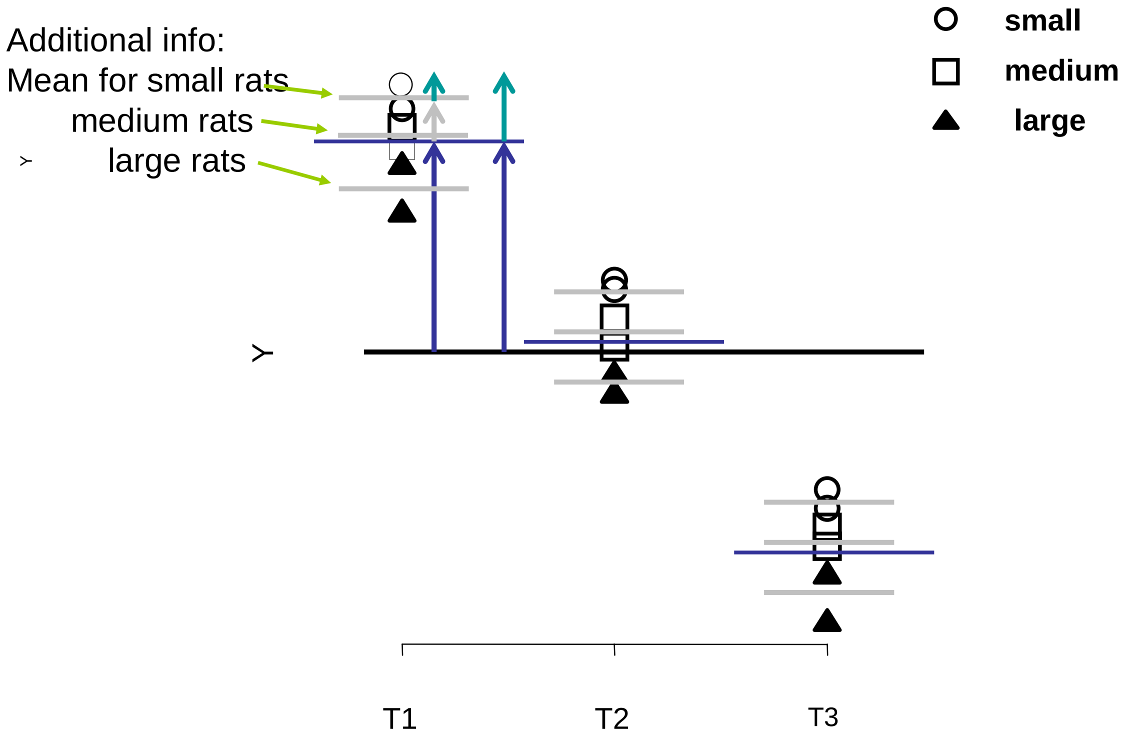
Blocking (i.e. controlling) for size

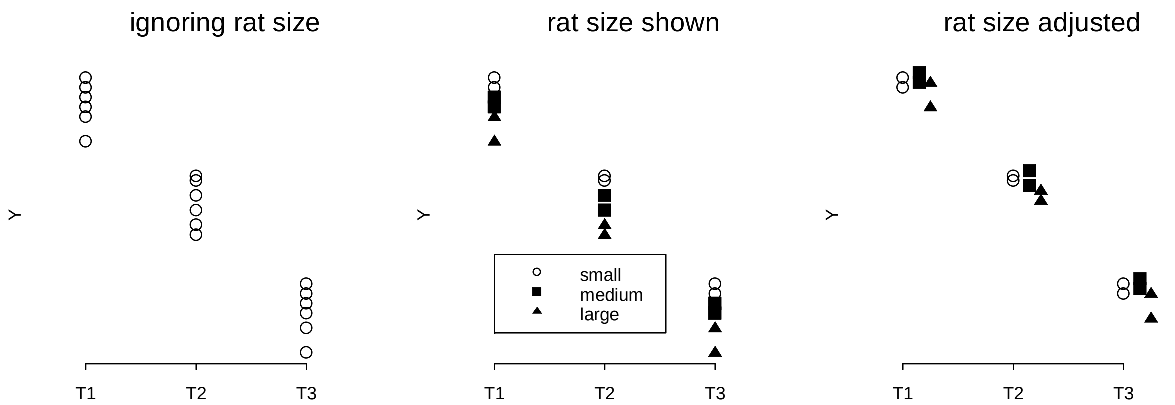
	Big rats "block"	Medium rats	Small rats
N= 9	Rat Rat Rat	Rat Rat Rat	Rat Rat Rat

Treatment 1 (↑ first rat), Treatment 2 (↑ second rat), Treatment 3 (↑ third rat)

N= 18	Rat Rat Rat	Rat Rat Rat	Rat Rat Rat
N= 27	Rat Rat Rat	Rat Rat Rat	Rat Rat Rat

Etc…



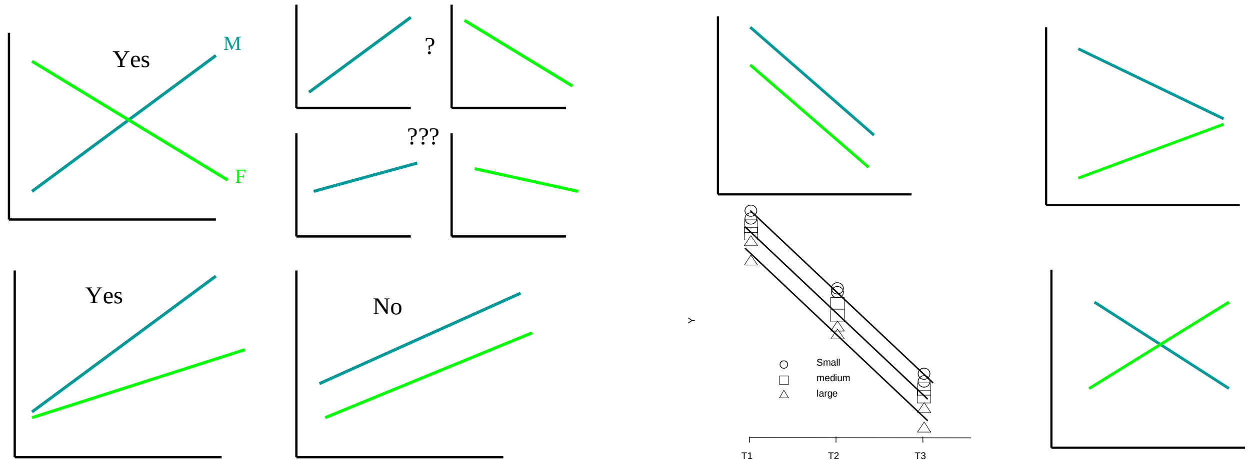


Source	DF	SS	MS	F	P
Treatment	2	2861	1430.5	13.13	0.00014
Error	24	2614	108.9		
Total	26	5475			

WITH BLOCKING for size

Source	DF	SS	MS	F	P
Block	2	2279	1139.7	74.94	0.0000
Treatment	2	2861	1430.5	94.06	0.0000
Error	22	334	15.2		
Total	26	5475			

"interactions": information for (almost) free



Sources of "noise"

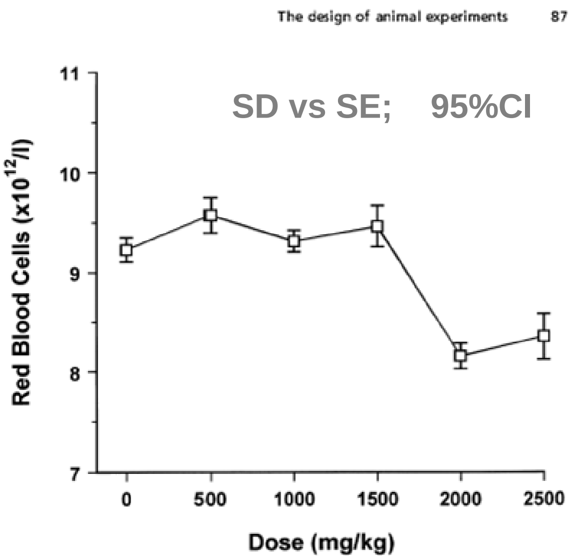
The experimental unit ?

Age,
Sex,
Weight
Stress
Subclinical disease

Temperature etc
Animal House/Barn/Cage
Position in rack

Time (hours/months)
Experimenter/carer (competence, unintentional bias)
…

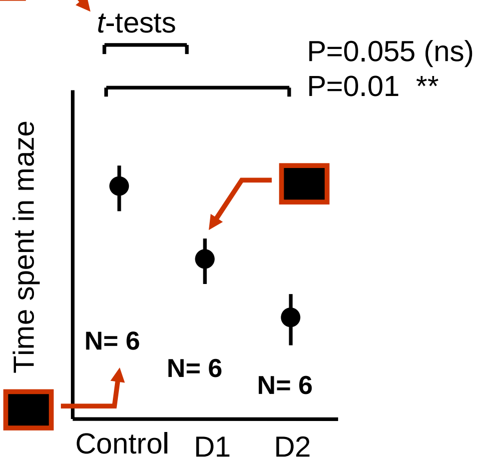
Presentation



Searching for Truth = Experiment Revisited

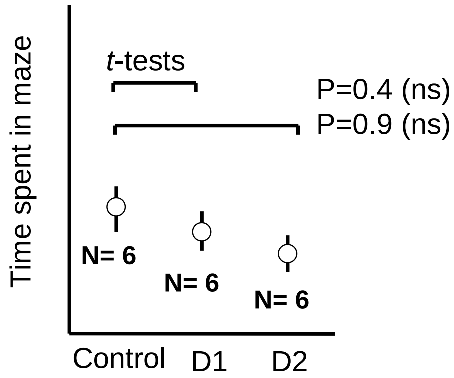
Test of drug doses on cognitive abilities (tested in maze).
<u>Exp 1</u>: 3 treatments administered to N=18 male rats, group-housed in 3 cages of 6 rats). **Control** administered to cage 1 (6 rats), **D1** to cage 2,etc.

Conclusion (kind of encouraging)
D1 no effect.
D2 significant effect



<u>Exp.2</u> Repeated on females

Conclusion:
D1 no effect.
D2 no effect



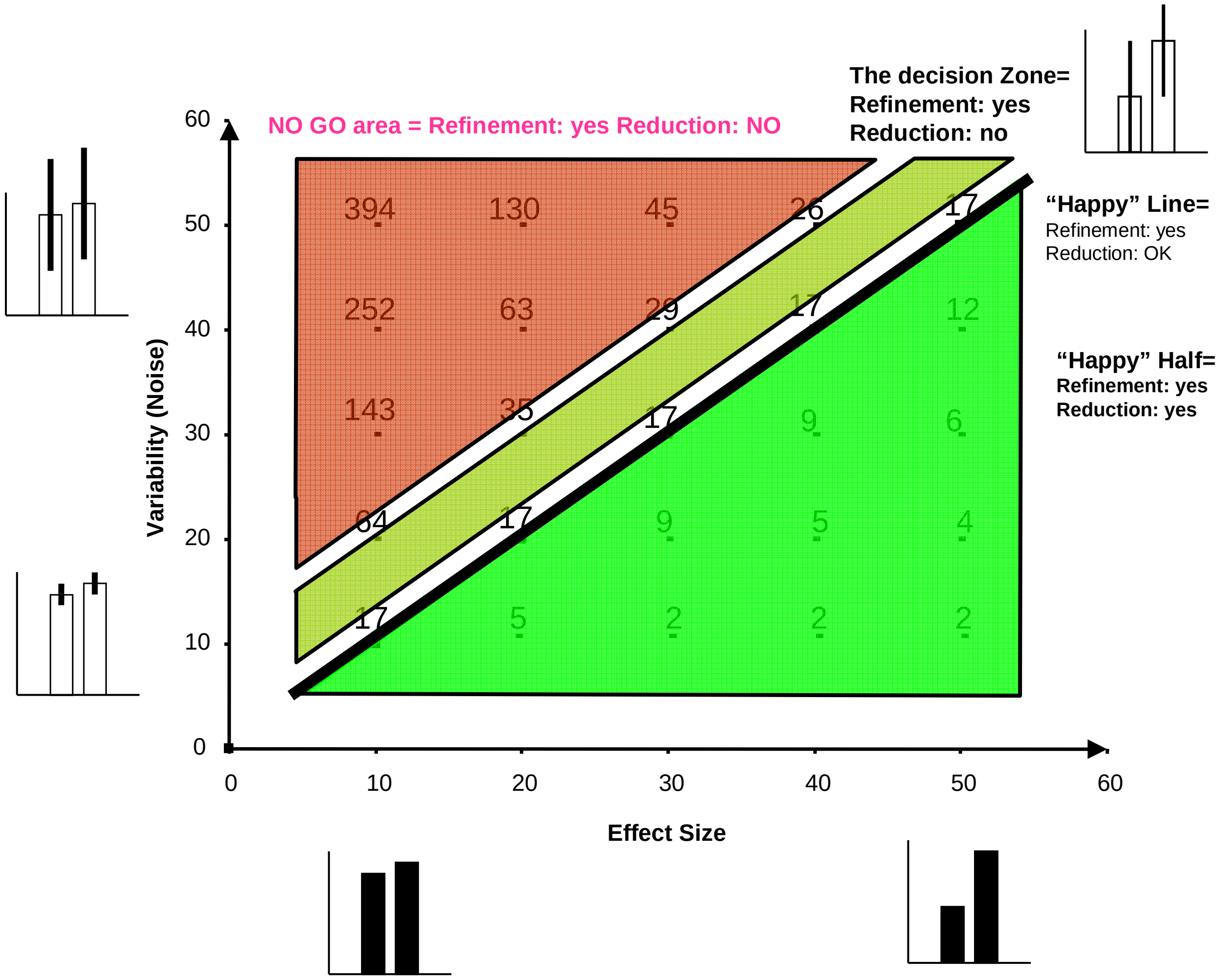
The 6 variables "determining" the chance of statistical significance

- Significance level **= Chance of False <u>Positive</u>**
  [arbitrary, set at P= 0.05 min]
- Desired Power of experiment **= Chance of False <u>Negative</u>**
  [arbitrary, set at 0.80 - 0.90]
- Alternative Hypothesis (1 vs 2 tailed)

- Size of the effect of biological interest (= Signal)
- Variation (i.e. Standard Deviation) (= Noise)
- Sample size (N) *

Note: this is a closed system, i.e. fix any five and the sixth can be derived

Standardisation & Reduction vs Refinement (my thoughts)



In Short: Take home messages

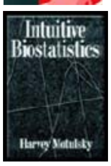
Power
- Remember the law of diminishing returns (power curve)
- Remember the (squared) effect of variation on numbers
- Use biggish experiments (factorial) rather many small ones (+ remember: additional bonus of "interactions")
- Remember that power is affected by variation and effect size (Refinement vs Reduction)

Precision
- Identify and reduce sources of unwanted variation
- Include them in your experimental design (rather than just worry about it afterwards)

Design
- Know about experimental design (ignorance is no defence)
- Talk to someone
- Do it before you start

Reading (examples of 3 kinds of books)

Festing, M., Overend, P., Gaine Das, R., Cortina Borja, M. & Berdoy, M. (2002). The design of Animal Experiments: Reducing the Use of Animals in Research Through Better Experimental Design.
Royal Society of Medicine Press, London.
(Intro to experimental design, see chapter 4 for summary of designs)



Motulsky , H. (1995) Intuitive Biostatistics; OUP;
See also chapter 37 for choosing tests
(and on web: at www.graphpad.com/www/Book/Choose.htm)

Grafen, A & Hails, R. 2002
Modern Statistics for the Life Sciences, OUP
On Model formulae, part of General Linear Model conceptual framework (GLM) which is a catch all for most tests:
e.g. mazespeed = treatment ratweight