Welcome to the course on

Experimental Design and Statistics: The Good, the Bad and the Ugly

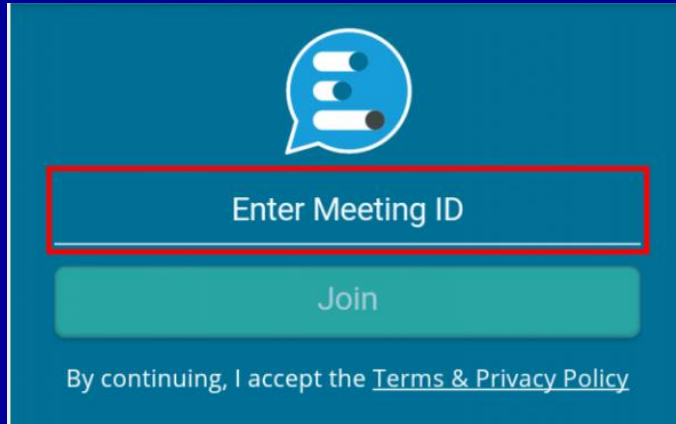
Manuel Berdoy

Make sure you have downloaded the app on your smart phone (IOS or Android) from:

get.meetoo.io

Whilst we wait to start you can enter the meeting

Our Meeting ID is: 107-854-181



Experimental Design and Statistics.

- Aim: understand the Essence of Experimental Design and Statistics

- Ethos for today (important):
  we are all interacting friends.

- Material for distribution throughout the day
- Group exercises

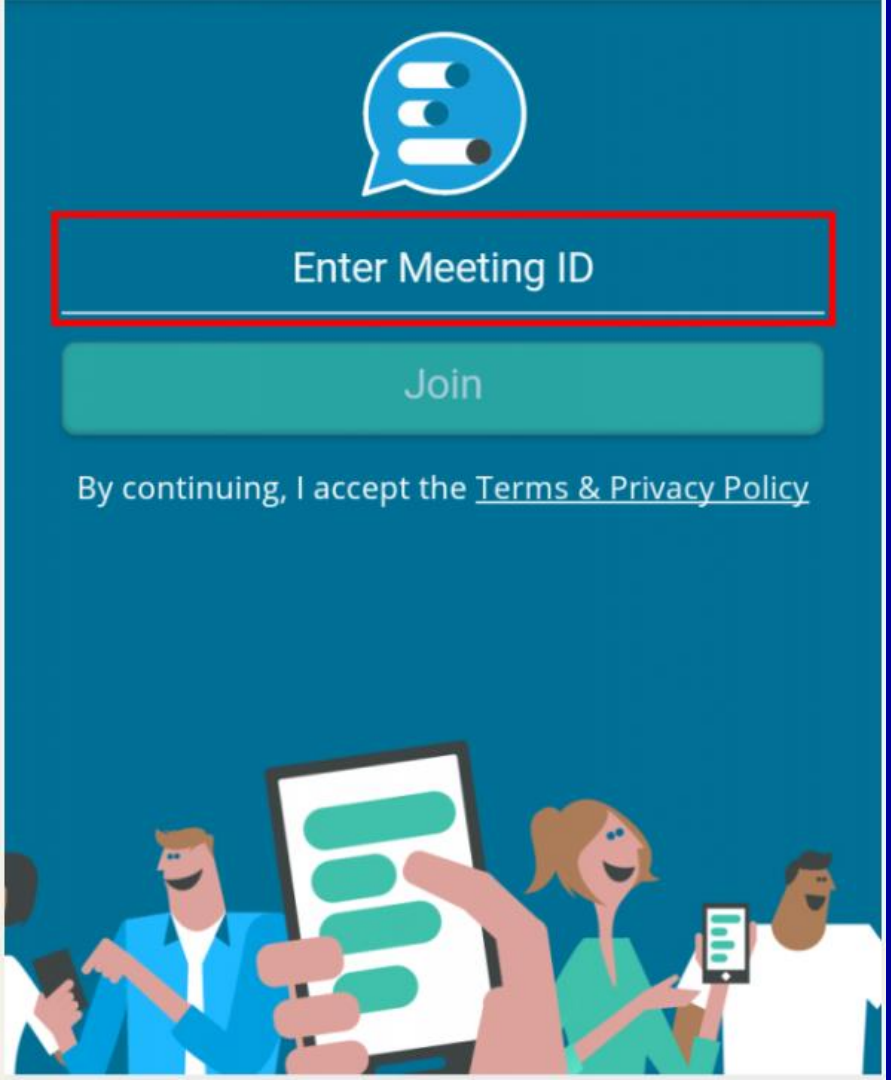
Experimental Design and Statistics.

Your Handouts:

1. **From P values to Power Calculation (LECTURE NOTES)**
2. **Analysis of Variance and the Control of Variation. (LECTURE NOTES)**
3. Workshop exercises

3b, 3c papers, as part of workshop exercises

4. Assumption of Parametric tests, and data transformation
5. + 5b Overview of Several Experimental designs and analyses
6. Some relevant websites (mentioned during the lectures)
7. Adjusted vs Sequential Sum of Squares (additional material)



Manuel 's Experimental Design

Meeting ID:

107-854-181

Document 1: From P values to Power calculation

Why Statistics ?

“Lies, Damned lies and Statistics”.

This is not “Statistics”, but politics.

“All statistics shows is that most of us have more than the average number of legs”

Factually correct, but that is also not what

“statistics” is about.

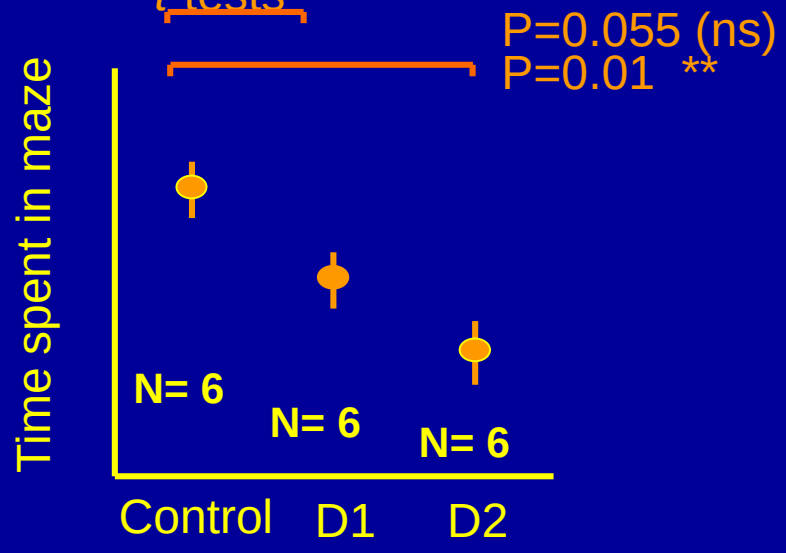
Statistics is about finding the truth in an imperfect world, and saving time and effort… and (sometimes) lives.

Searching for Truth = Experiment

Test of drug doses on cognitive abilities (tested in maze).

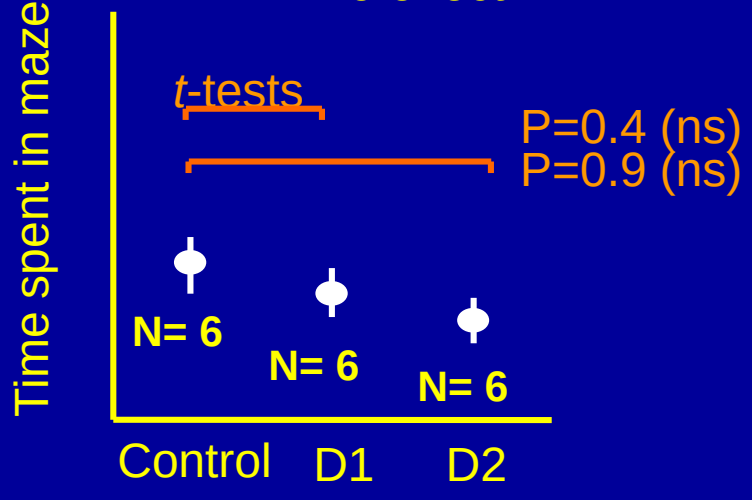
Exp 1: 3 treatments administered to N=18 male rats, group-housed in 3 cages of 6 rats). **Control** administered to cage 1 (6 rats), **D1** to cage 2,etc.

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



Exp.2 Repeated on females

Conclusion:
D1 no effect.
D2 no effect



Searching for Truth = Experiment ?

EDISON'S BULBS FAIL TO LIGHT UP AUCTION
First all-science collection sells modestly at Christie's.
www.nature.com/news

CHRISTIE'S

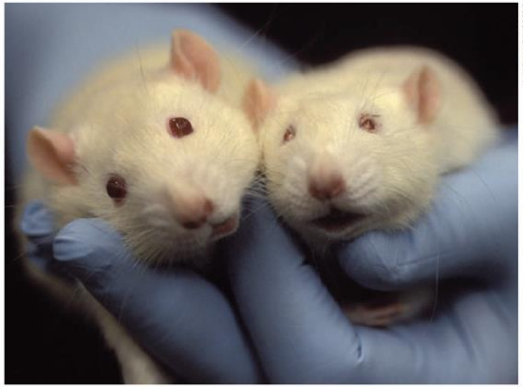
Animal experiments under fire for poor design

In the contentious world of animal research, one question surfaces time and again: how useful are animal experiments as a way to prepare for trials of medical treatments in humans? The issue is crucial, as public opinion is behind animal research only if it helps develop better drugs. Consequently, scientists defending animal experiments insist they are essential for safe clinical trials, whereas animal-rights activists vehemently maintain that they are useless.

Now a British team has made the first attempt to answer the question in a scientific way, and the result suggests that animal researchers need to raise their game. The team claims that animal experiments are often poorly designed, and so fail to lay the ground properly for subsequent human studies.

The study looked at six treatments that have been evaluated in detail in human trials. The researchers assessed whether animal studies had accurately predicted the outcome of the human work, a task that involved reviewing more than 200 papers. In three of the six cases, the answer was no (P. Perel *et al. Br. Med. J.* doi:10.1136/bmj.39048.407928.BE; 2006).

Ian Roberts, an epidemiologist at the London



K. KASMAUSKI/CORBIS

Are animals being wasted in badly thought through experiments?

e.g. lack of reporting of:

Randomisation,
Double blinding,

External Validity

+ other problems…

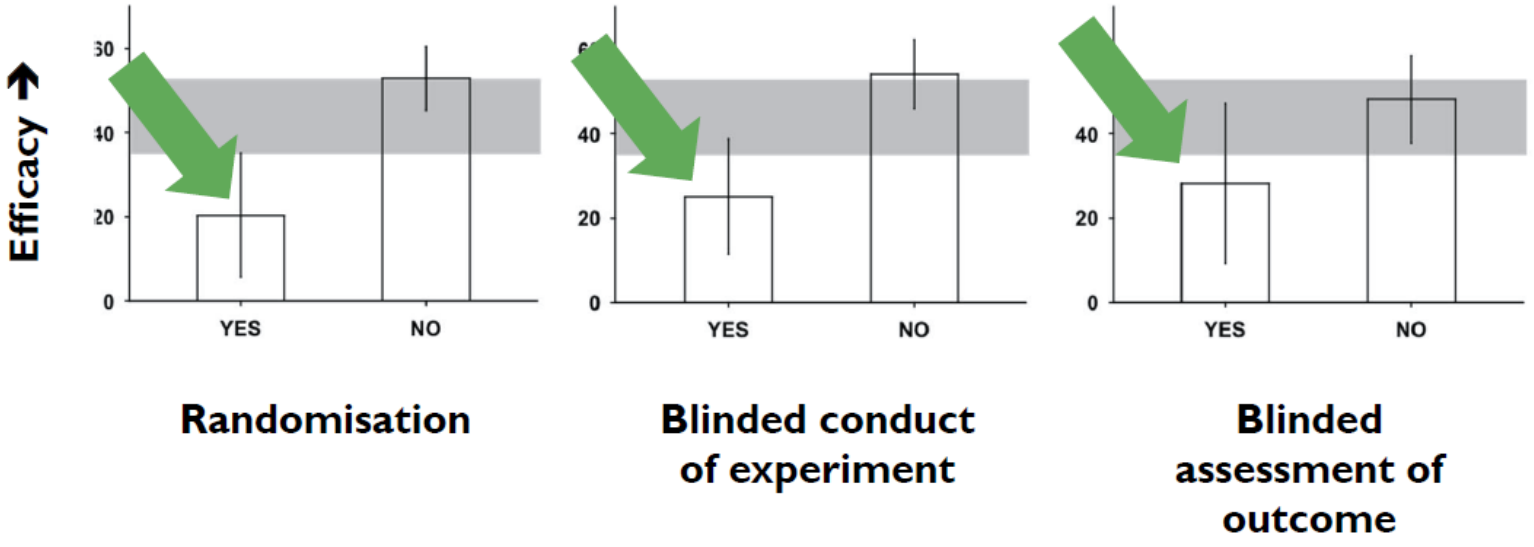
The problem: Basic errors are widespread

- experimental unit wrongly identified in 48% of 99 papers surveyed (Hurlbert 1984)
- 54% of 141 articles in *Infection* and *Immunity* had errors of analysis, reporting or both (Olsen 2003)
- 79 of 157 neuroscience papers used incorrect comparisons of results (Neuwenhuis et al 2009 -)
- "Random allocation of animals to experimental groups was reported in only 13% of all the studies in the sample" (Kilkenny et al 2009)

  Only 14% of all papers [susceptible to observer bias] also reported that they used blinding." (Blinding is an effective way of reducing bias)

We know that these errors have an effect:

- 11 publications, 29 experiments, 408 animals
- Improved outcome by 44% (35-53%)



Macleod et al, 2008

We know that these errors have an effect:

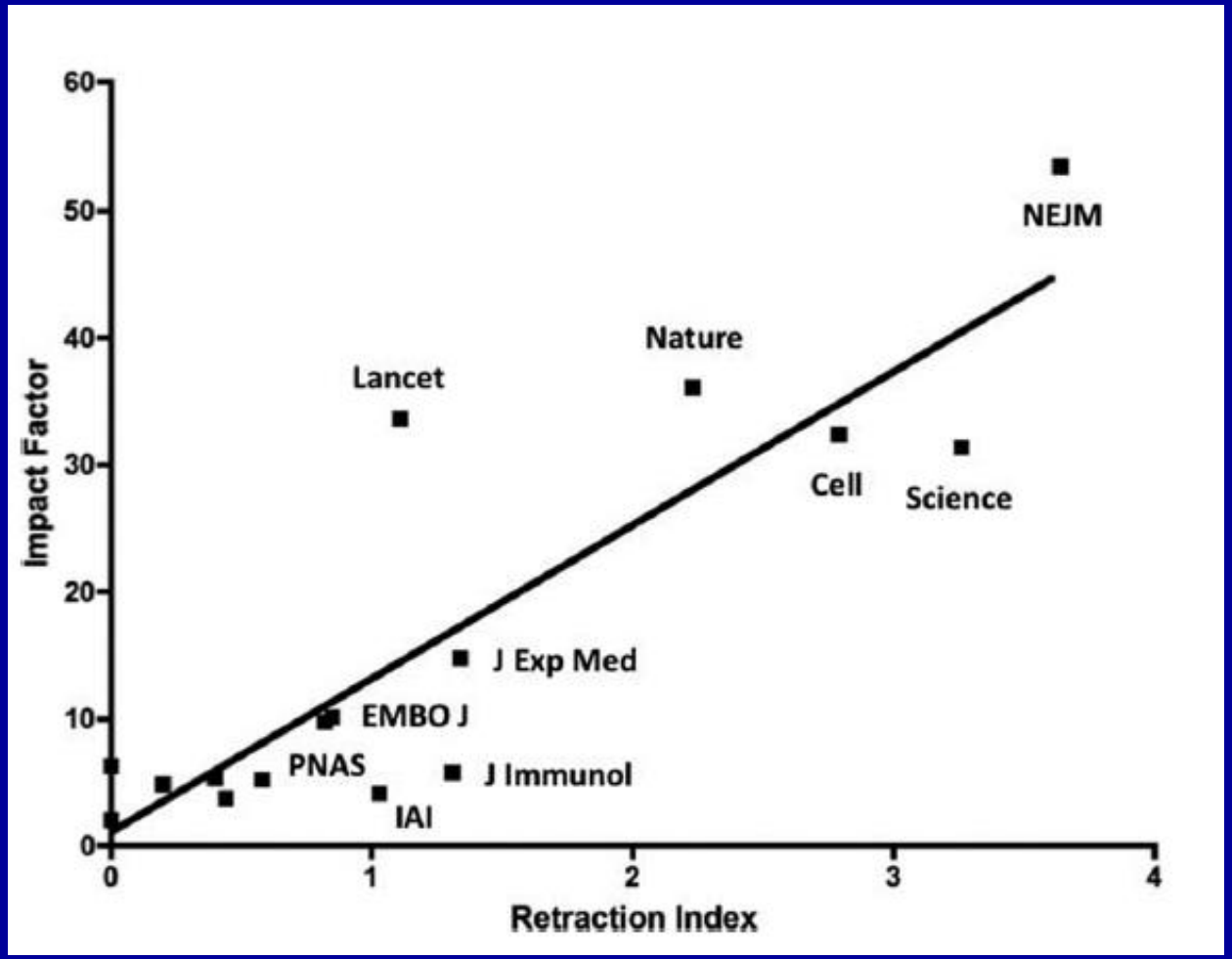
BIAS

- Of 290 studies surveyed those which did not report that they randomised were more likely to report positive findings than those that did.
- Those that reported neither randomising nor blinding were even more likely to report positive findings.

(Bebarta et al 2003)

WASTE of ANIMALS:

Many researchers do not use efficient “blocked” and “factorial” designs



Fang & Casadevall 2011

Retraction Watch: Tracking retractions as a window into the scientific process https://retractionwatch.com/

ANNOUNCEMENT

Reducing our irreproducibility

Over the past year, *Nature* has published a string of articles that highlight failures in the reliability and reproducibility of published research (collected and freely available at go.nature.com/huhbyr). The problems arise in laboratories, but journals such as this one compound them when they fail to exert sufficient scrutiny over the results that they publish, and when they do not publish enough information for other researchers to assess results properly.

From next month, *Nature* and the Nature research journals will introduce editorial measures to address the problem by improving the consistency and quality of reporting in life-sciences articles. To ease the interpretation and improve the reliability of published results we will more systematically ensure that key methodological details are reported, and we will give more space to methods sections. We will examine statistics more closely and encourage authors to be transparent, for example by including their raw data.

Central to this initiative is a checklist intended to prompt authors to disclose technical and statistical information in their submissions, and to encourage referees to consider aspects important for research reproducibility (go.nature.com/oloeip). It was developed after discussions with researchers on the problems that lead to irreproducibility, including workshops organized last year by US National Institutes of Health (NIH) institutes. It also draws on pub-

We know that these errors have an effect:

LACK OF EXTERNAL VALIDITY **(no agreement in 50% of cases)**

The cost of of low reproducibility in preclinical research (i.e. not just animal research) in the US alone, is estimated at:…

$ 28 000 000 000

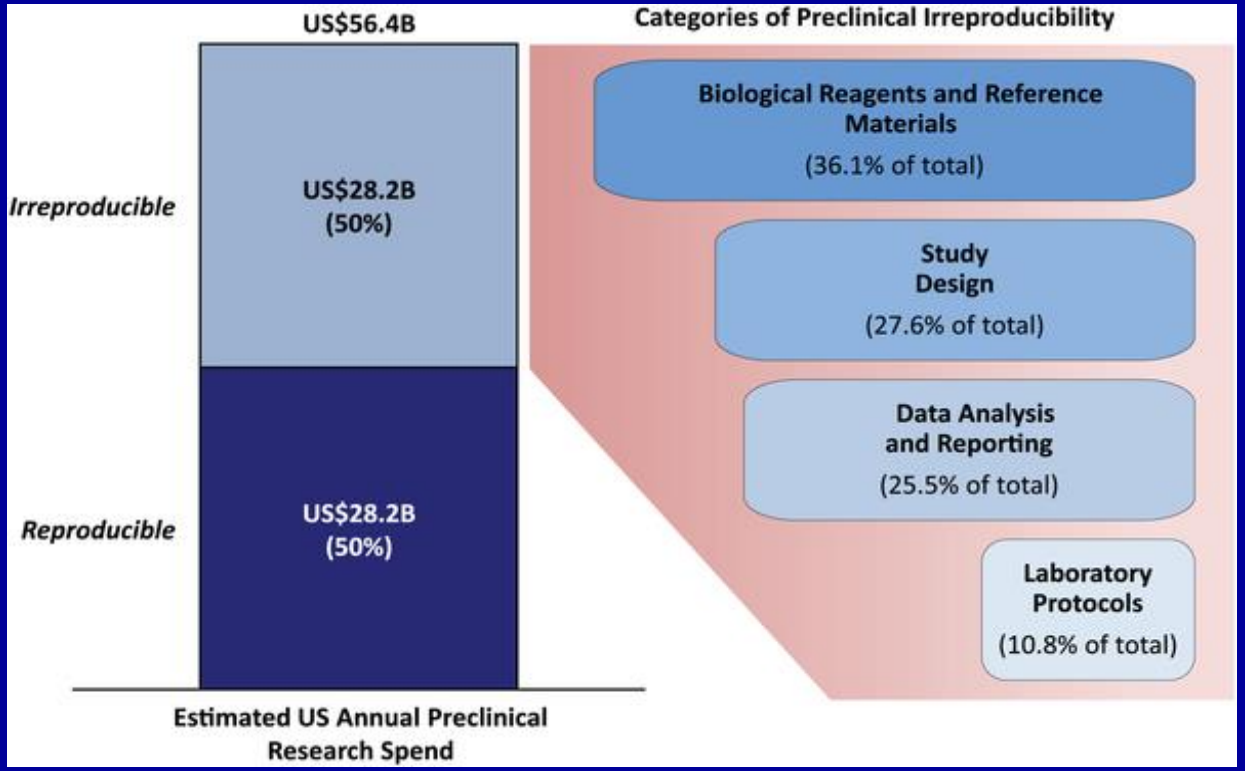
(Freedman, Cockburn & Simcoe PLOS Biology | DOI:10.1371/, 2015)

The Economics of Reproducibility in Preclinical Research

Leonard P. Freedman, Iain M. Cockburn, Timothy S. Simcoe

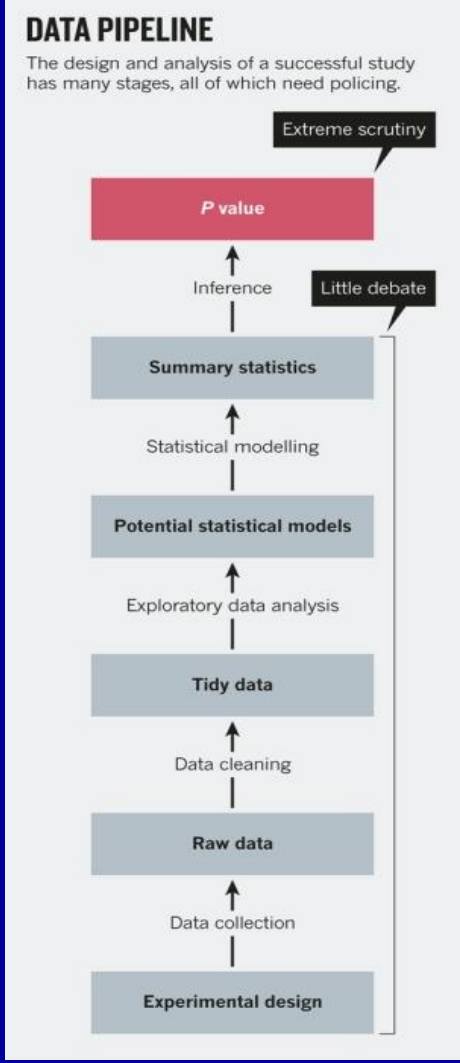
Published: June 9, 2015 • DOI: 10.1371/journal.pbio.1002165

Fig 2. Estimated US preclinical research spend and categories of errors that contribute to irreproducibility.



Freedman LP, Cockburn IM, Simcoe TS (2015) The Economics of Reproducibility in Preclinical Research. PLoS Biol 13(6): e1002165. doi:10.1371/journal.pbio.1002165

Jeffrey T. Leek& Roger D. Peng Nature 2015

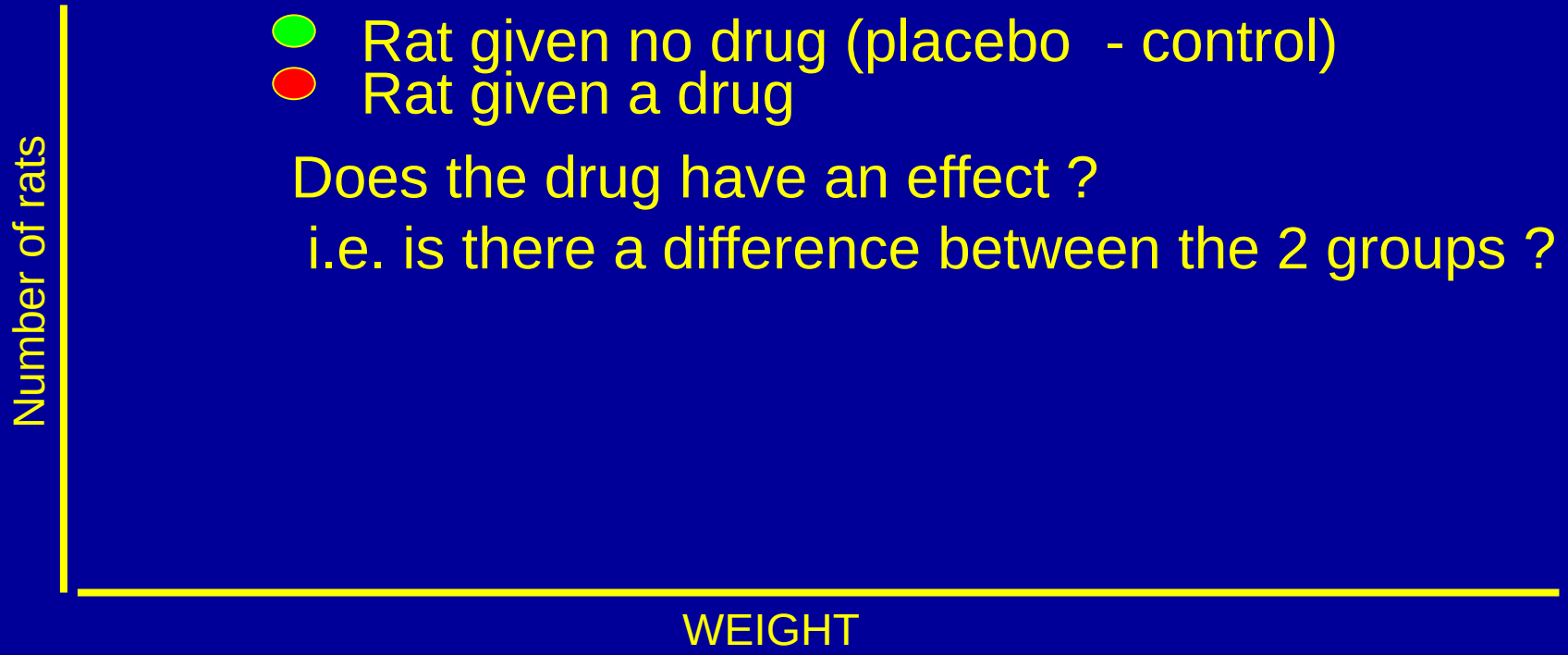


Searching for “Truth”: reminder

- Rat given no drug (placebo - control)
- Rat given a drug

Does the drug have an effect ?

i.e. is there a difference between the 2 groups ?

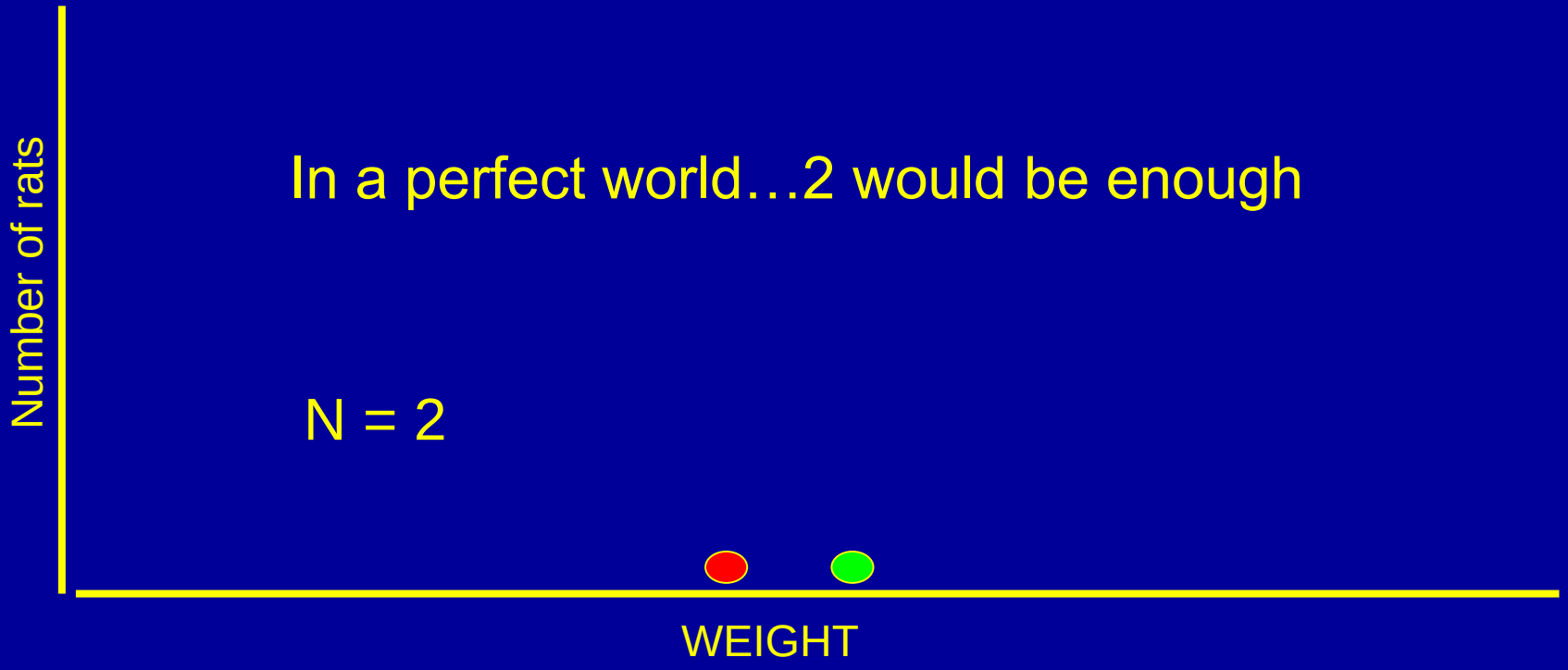


Searching for “Truth”: reminder

Searching for “Truth”: reminder

In a perfect world…2 would be enough

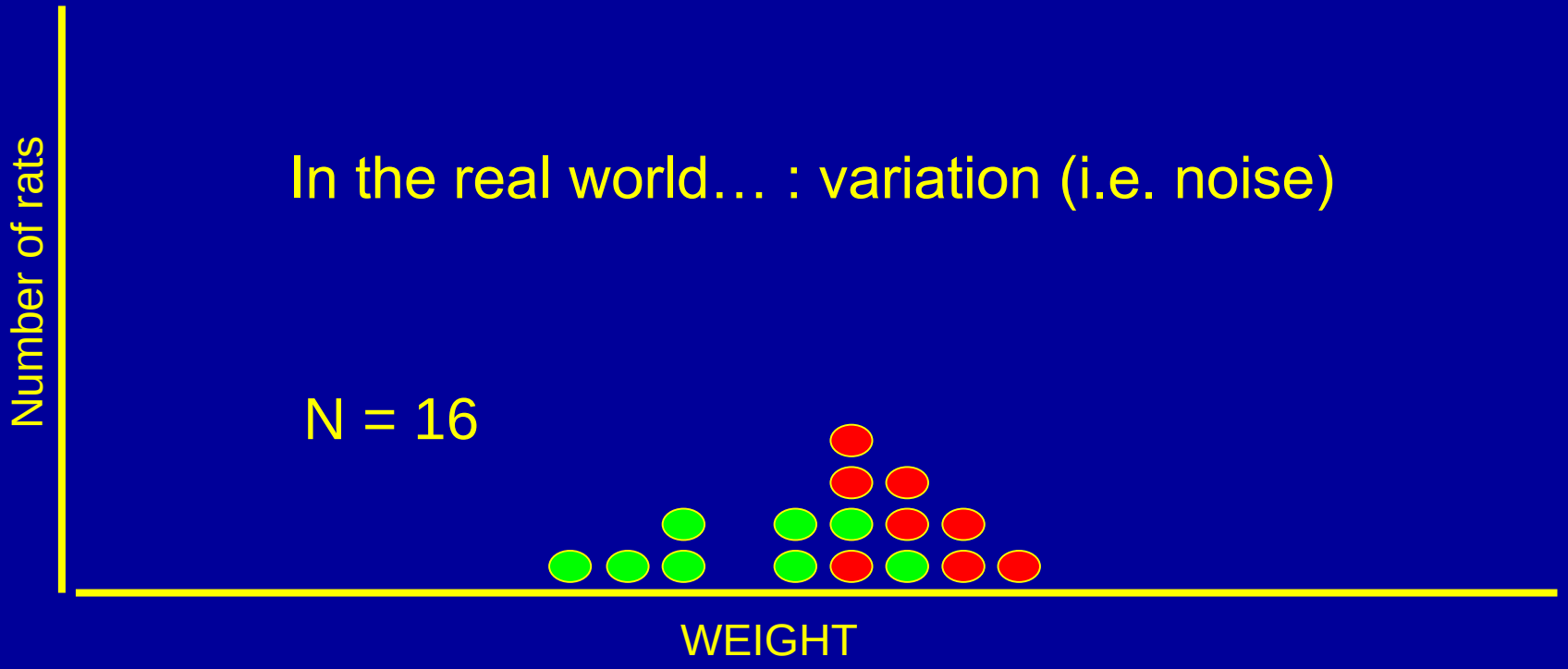
N = 2

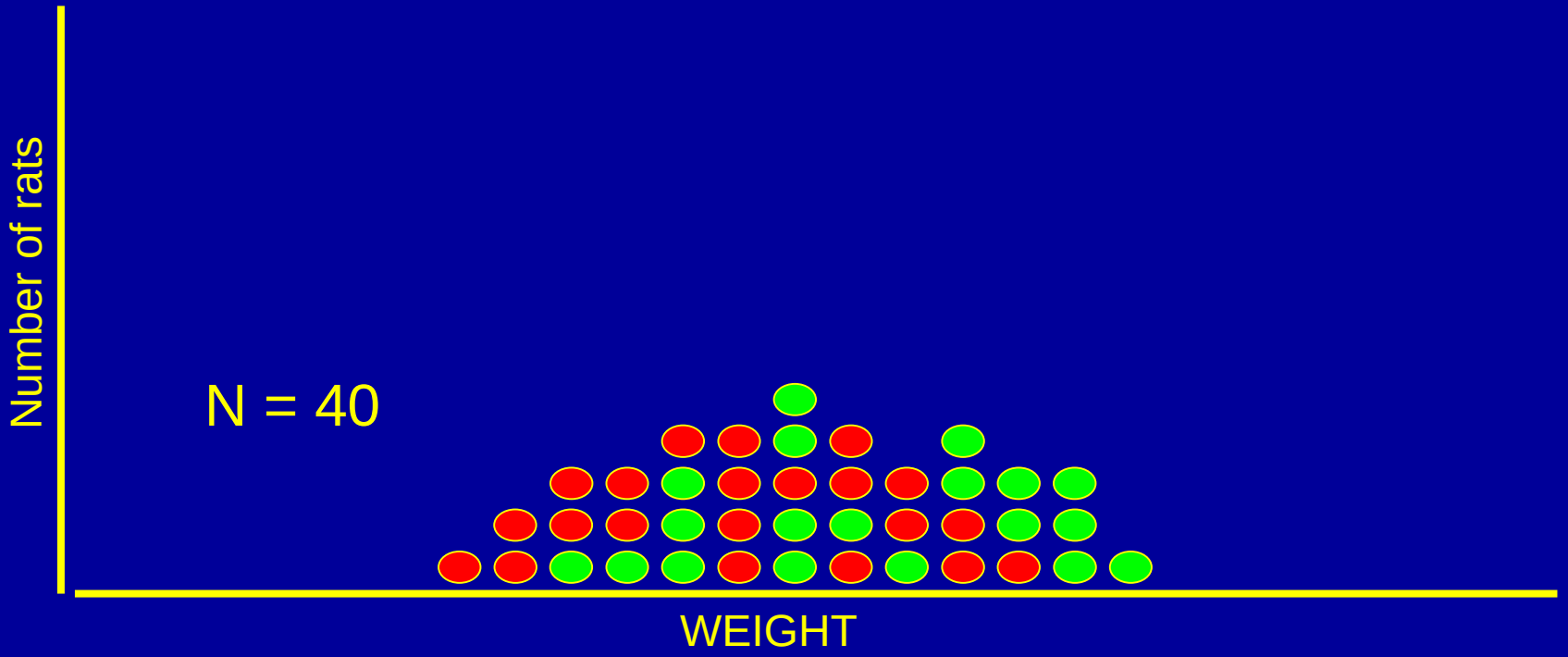


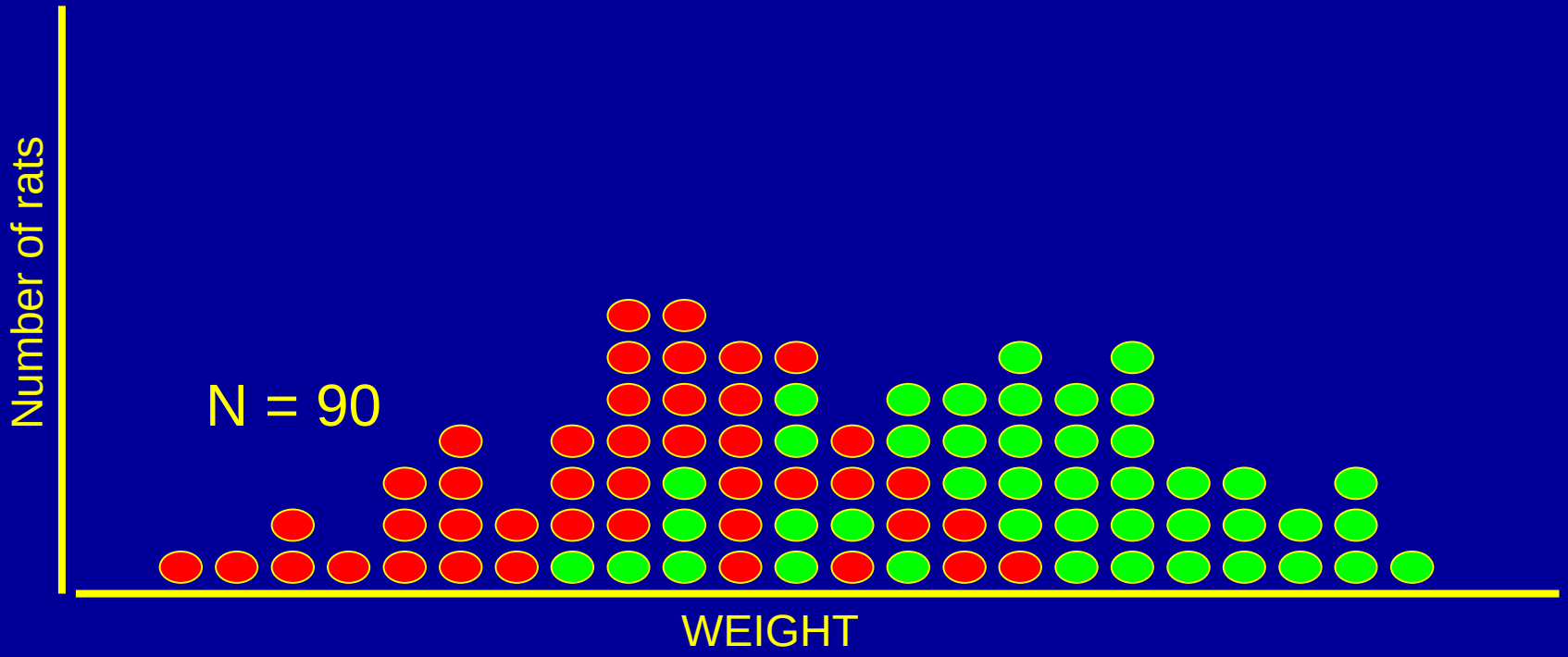
Searching for “Truth”: reminder

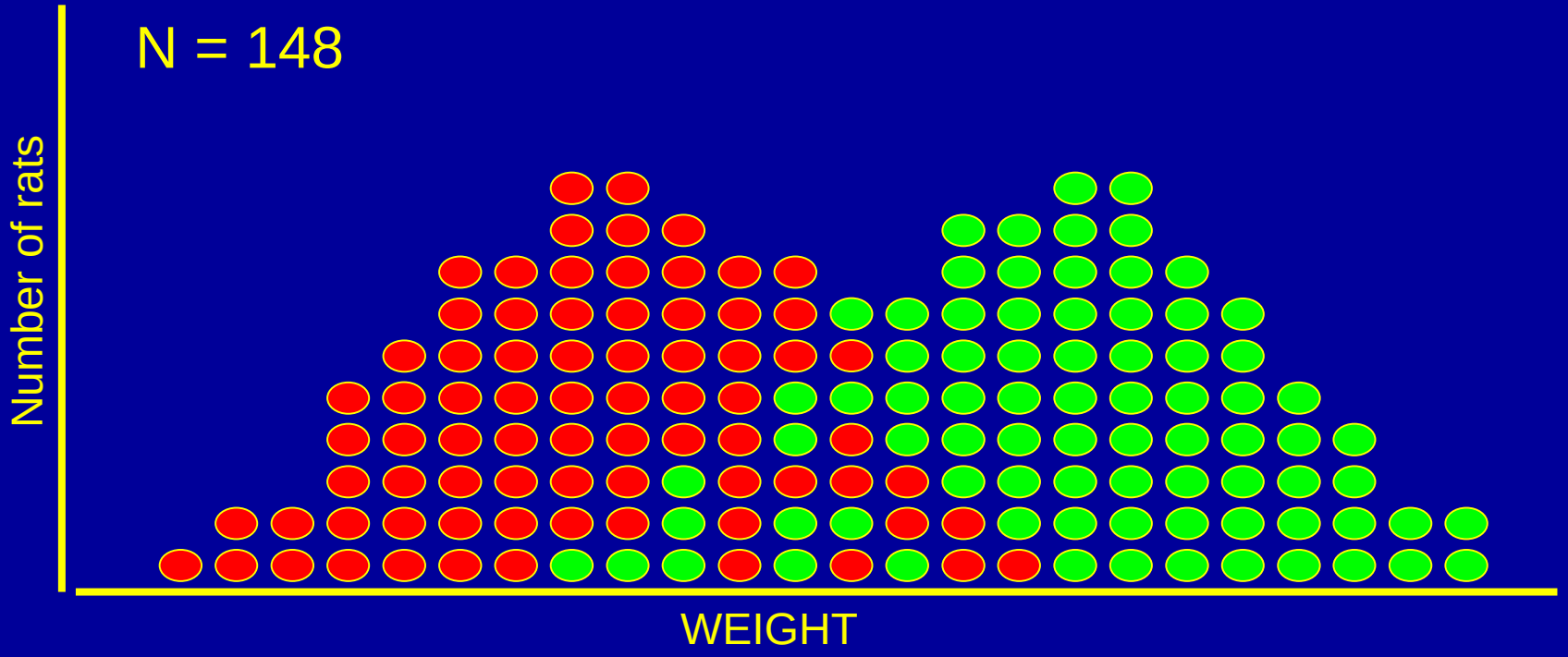
In the real world… : variation (i.e. noise)

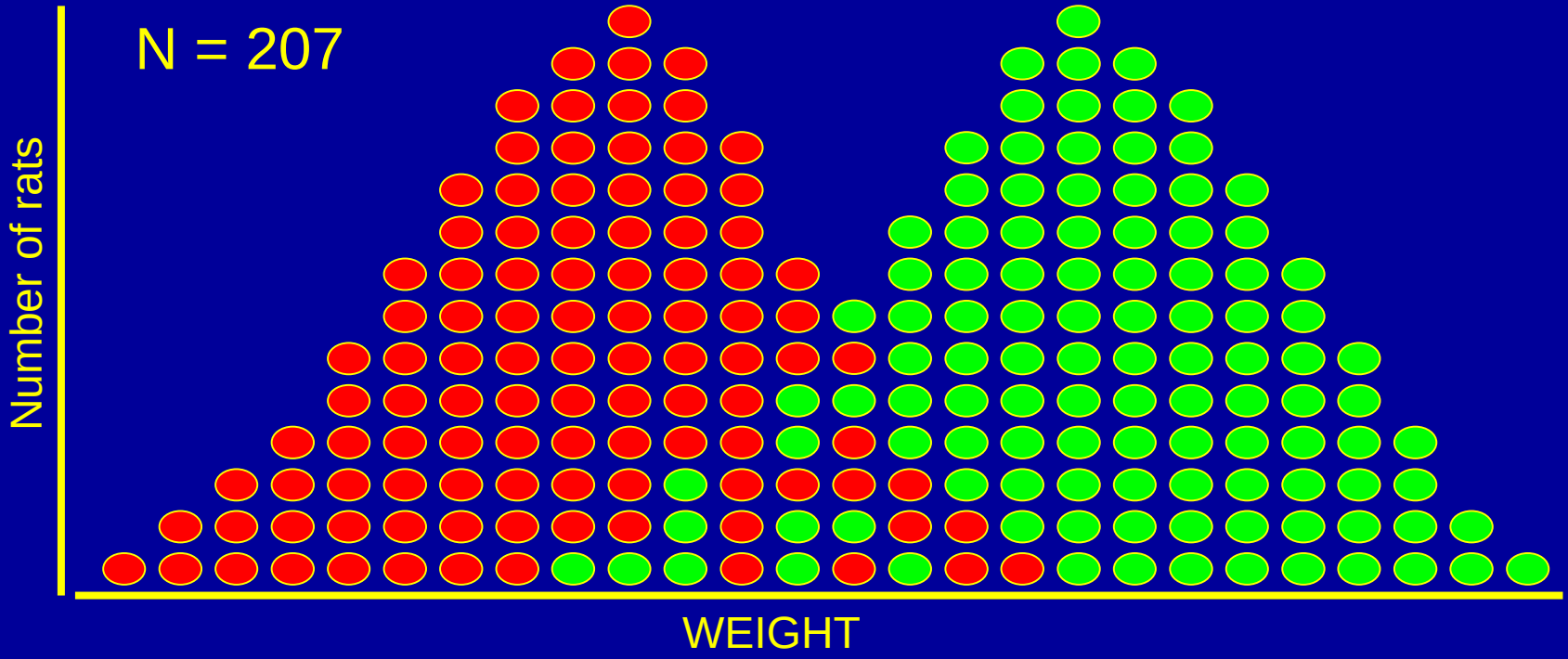
N = 16




Number of rats
N = 40
WEIGHT


Number of rats
N = 90
WEIGHT


N = 148
Number of rats
WEIGHT


N = 207
Number of rats
WEIGHT

How confident are we that there is a difference? i.e. How much would you be prepared to bet ?:

POLL OPEN

1. Nothing

46.15%

2. My Bicycle

28.21%

3. My Car

10.26%

4. My Neighbour’s life

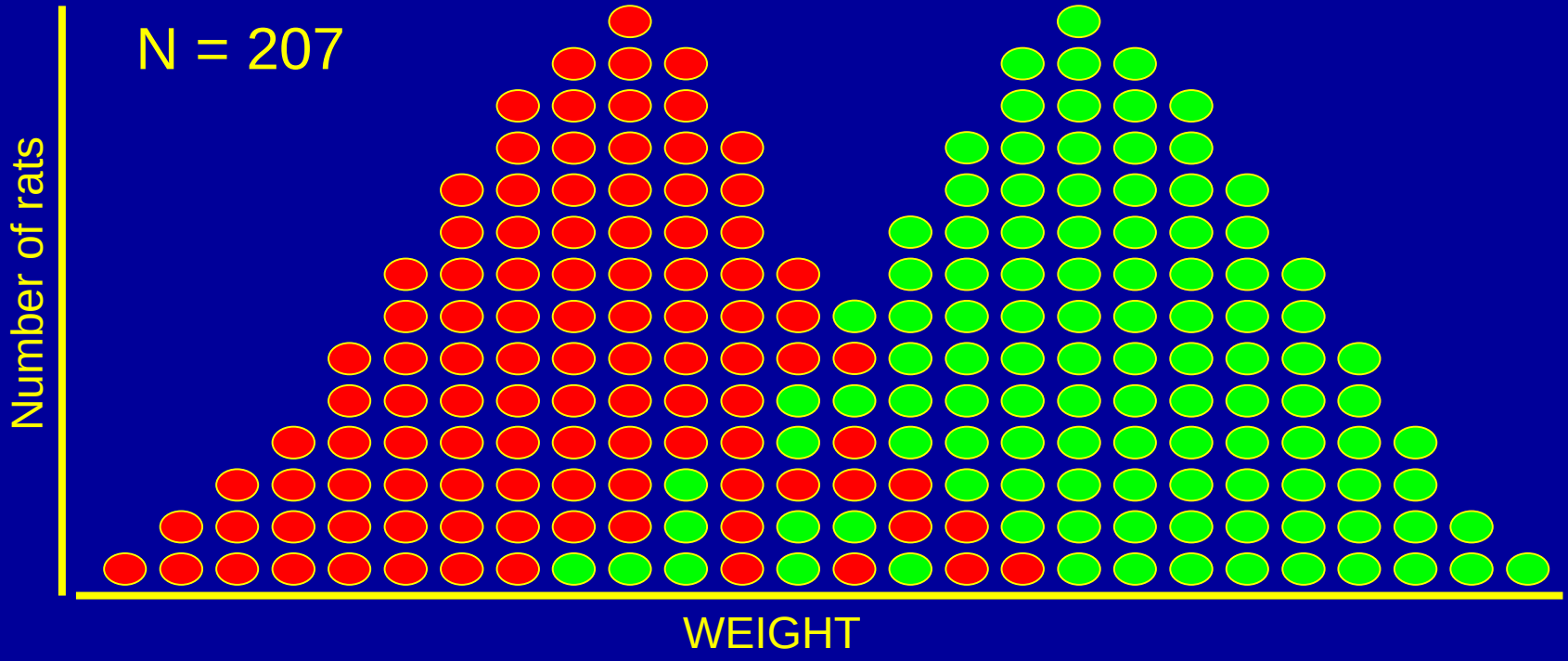
15.38%

5. My Life

0%

$P = 4 \times 10^{-51}$ = 4 chances in a sexdecillion-th

P = 0.000000000000000000000000000000000000000000004 *

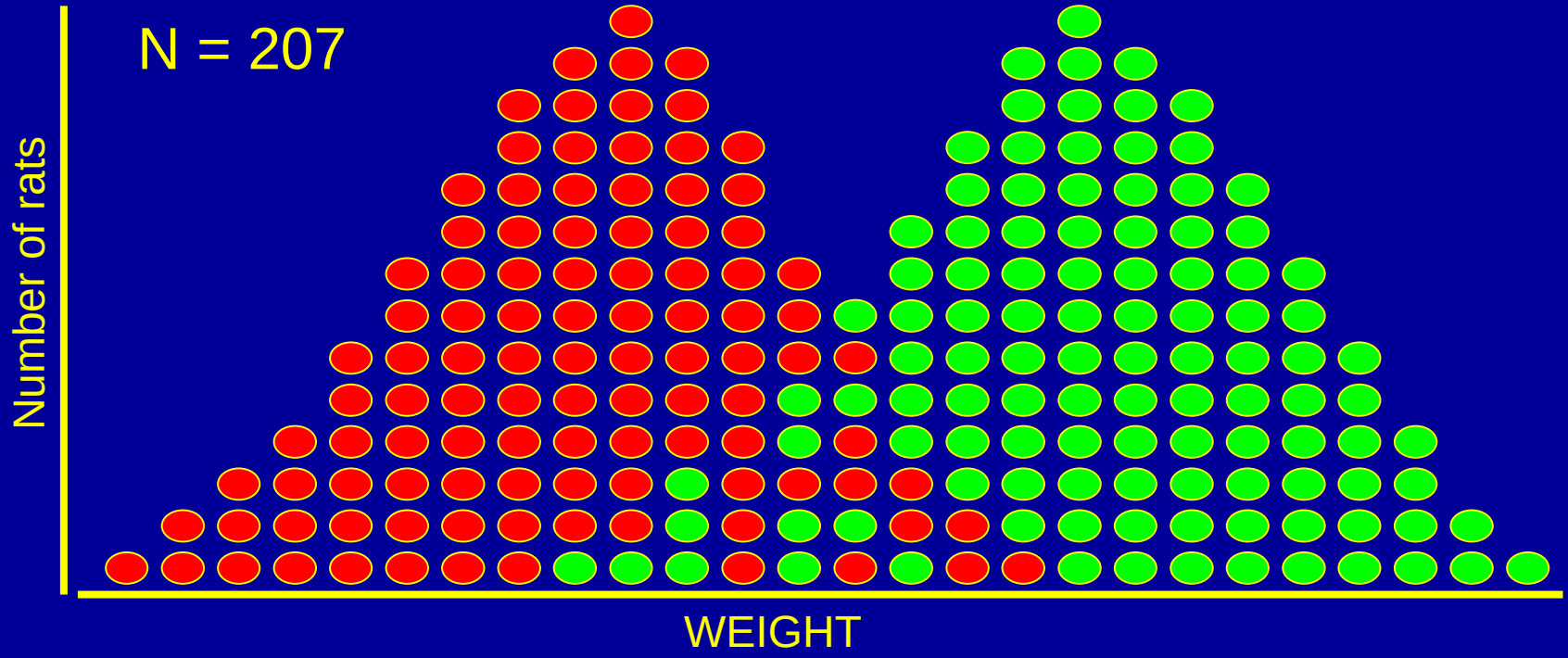


$P = 4 \times 10^{-51}$ = 4 chances in a sexdecillion-th

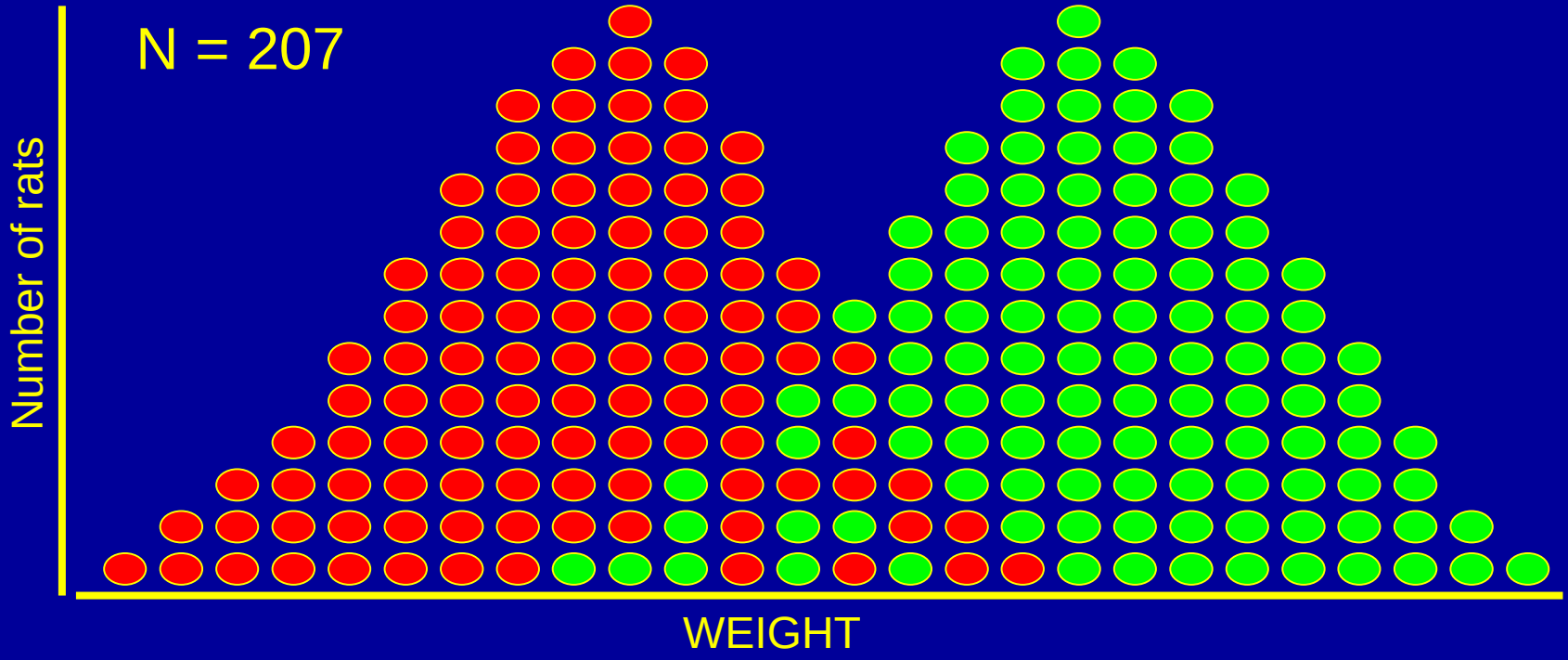
P = 0.000000000000000000000000000000000000000000004

AIR CRASH P = 0.000002 (1/400,000) per year.
CAR CRASH P = 0.0001 (1/6,500) per year.

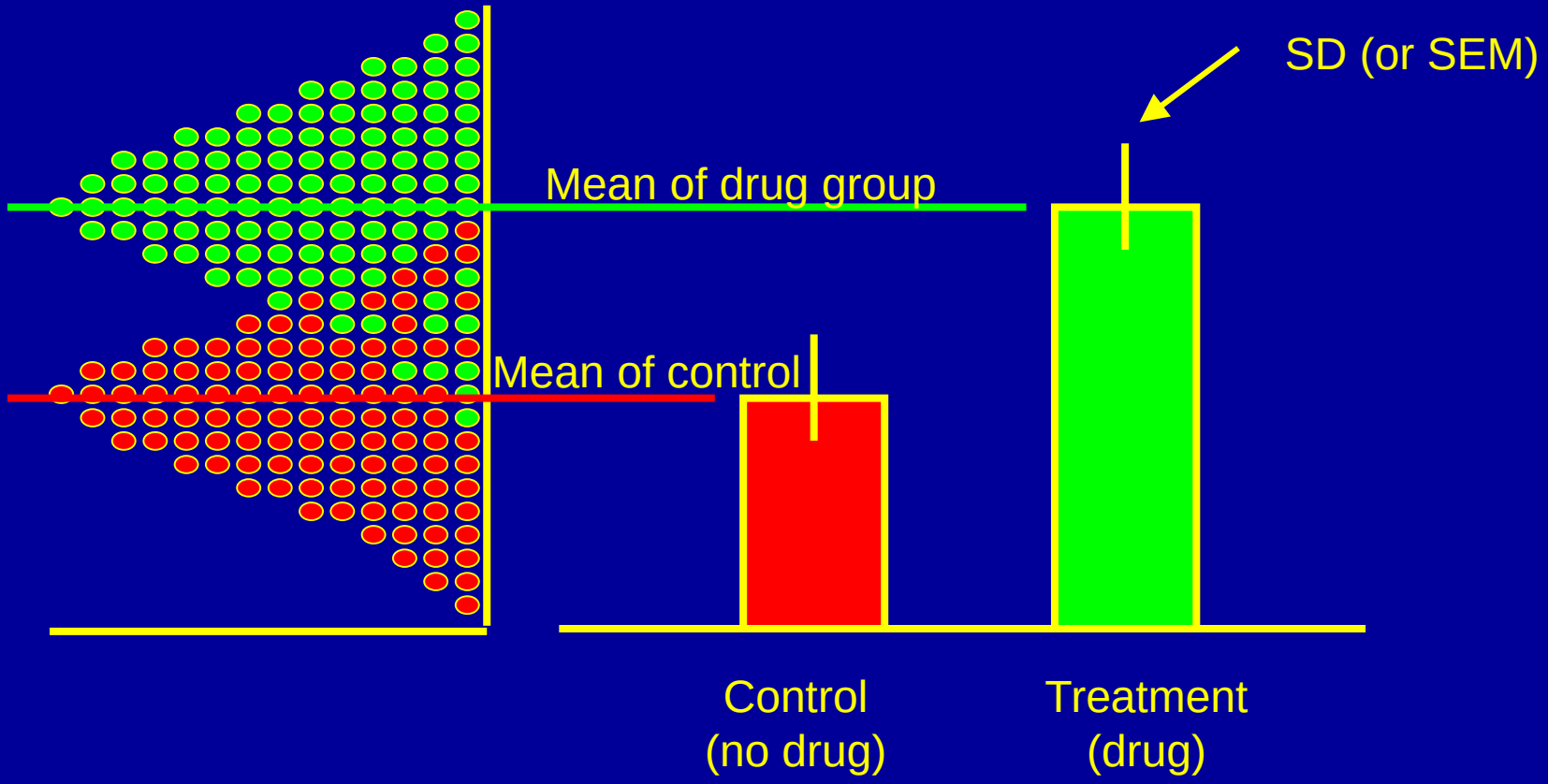
USA NSC 2003 data



Presentation reminder



Presentation reminder



Searching for Truth = Experiment

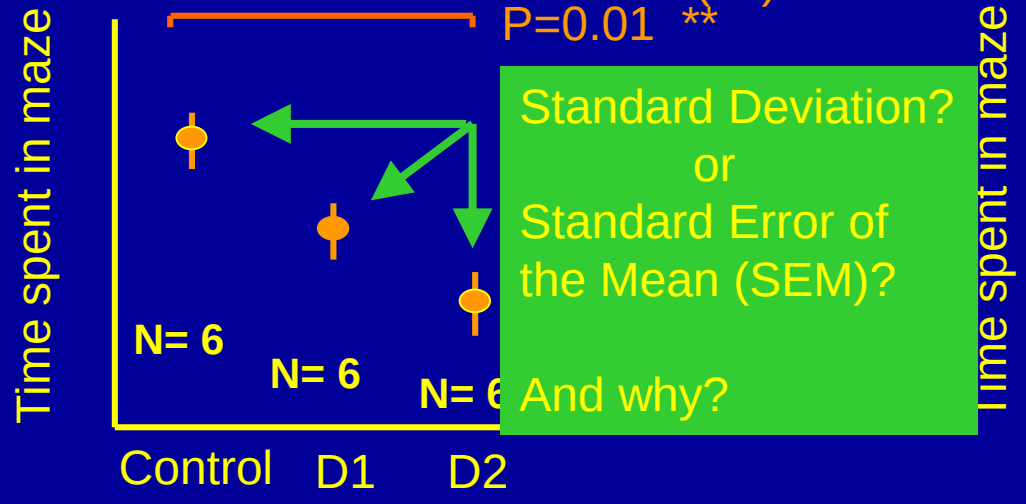
Test of drug doses on cognitive abilities (tested in maze).

Exp 1: 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

t-tests

P=0.055 (ns)
P=0.01 **



Standard Deviation?
or
Standard Error of the Mean (SEM)?

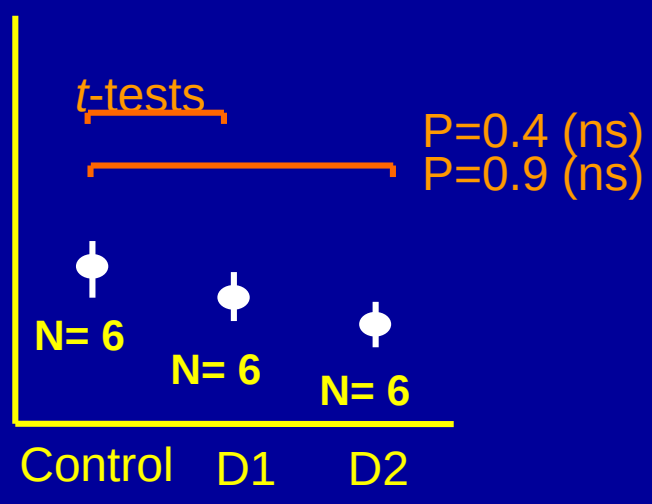
And why?

Exp.2 Repeated on females

Conclusion:
D1 no effect.
D2 no effect

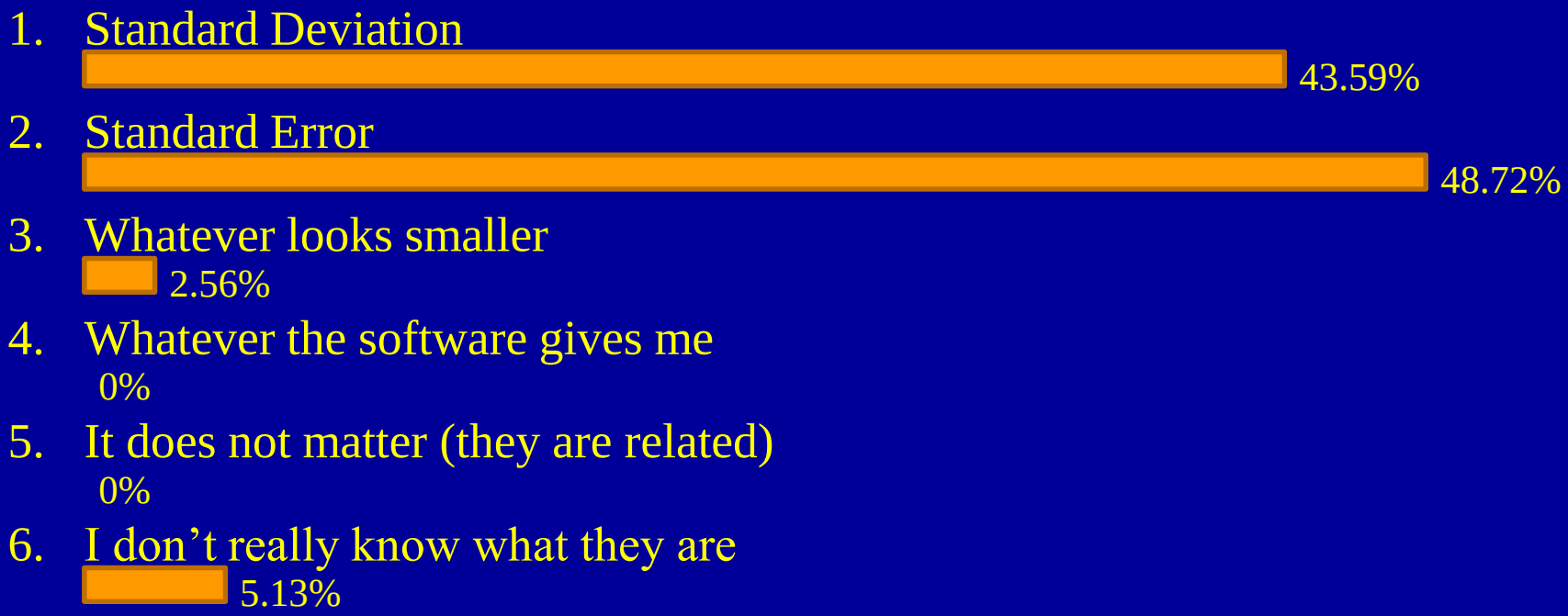
t-tests

P=0.4 (ns)
P=0.9 (ns)



POLL OPEN

When comparing means I prefer to use:



Relationship between SE and SD

$$SEM = \frac{SD}{\sqrt{N}}$$

You need to write this down

Power Calculation

- Not always realistic

- And we do it to some extent anyway
  - Power calculation is a way of formalising the process
  - Ethical grounds & scientific grounds for it.

  (Regulatory, 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.
(and with animals, non consenting and possibly 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

"Absence of evidence is no evidence of absence"

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 (1-Beta = Power)	Type II error Beta

Prob. of detecting a specified effect at specified significance level

The 6 variables “determining” the chance of statistical significance

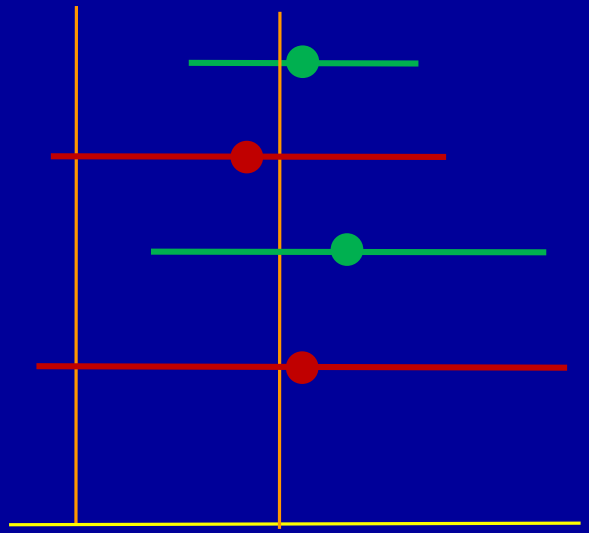
- Significance level **= about False Positive**
  [arbitrary, set at P= 0.05 min]
- Desired Power of experiment
  [arbitrary, set at 0.80 - 0.90]
- Alternative Hypothesis (1 vs 2 tailed)

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

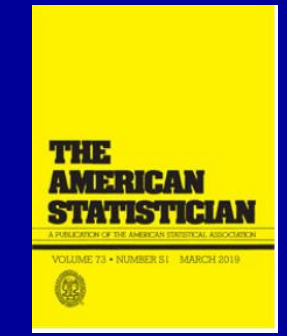
Statistical Inference in the 21st Century: A World Beyond p < 0.05 The American Statistician ; Volume 73 2019 - Issue sup1:

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

“Out of the four studies, half reported no significant difference therefore…….”



No effect Observed effect and confidence



Statistical Inference in the 21st Century: A World Beyond p < 0.05 The American Statistician ; Volume 73 2019 - Issue sup1:

- P-Hacking; HARKing
- “0.05 cliff”

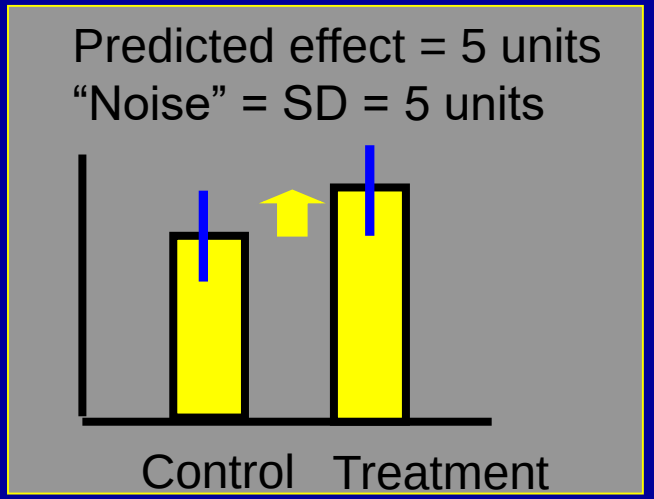
The 6 variables "determining" the chance of statistical significance

- Significance level — **= about False Positive**
  [arbitrary, set at P= 0.05 min]
- **Desired Power of experiment** — **= about False Negative**
  [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

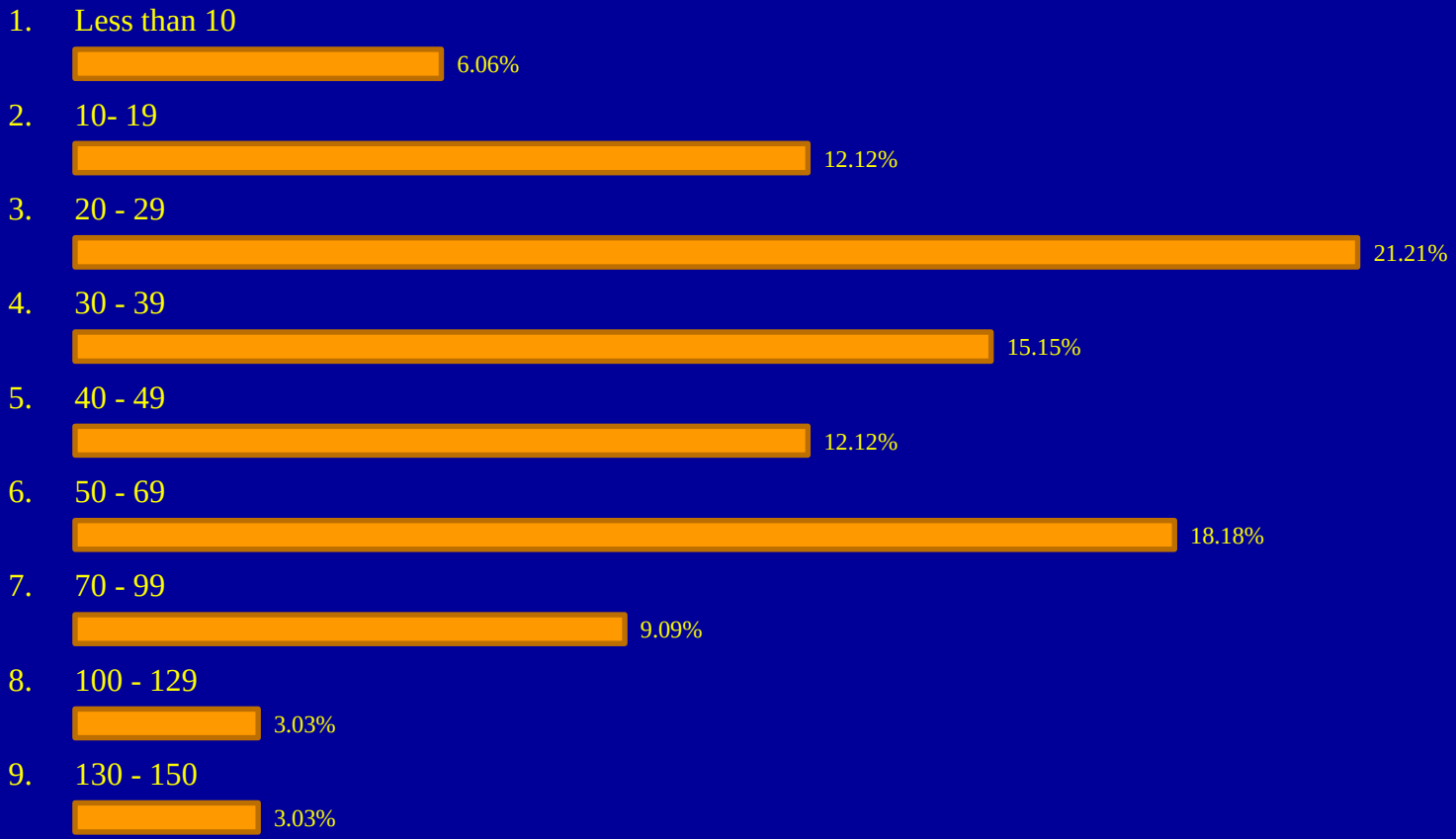
Power



How many animals do we need ?
(to have enough power)

How many animals are needed?

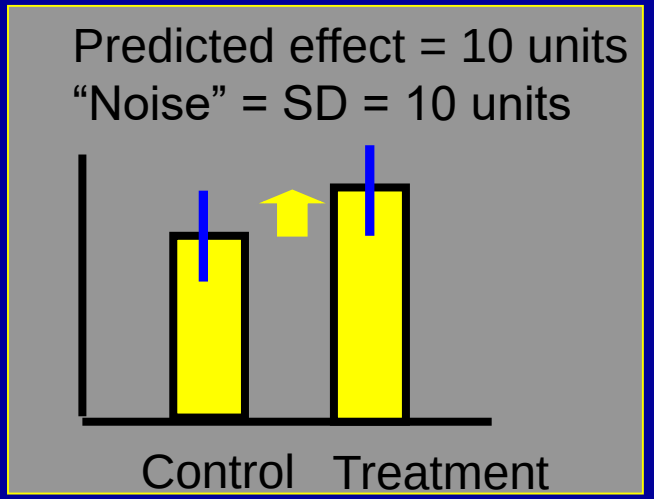
POLL OPEN



Power

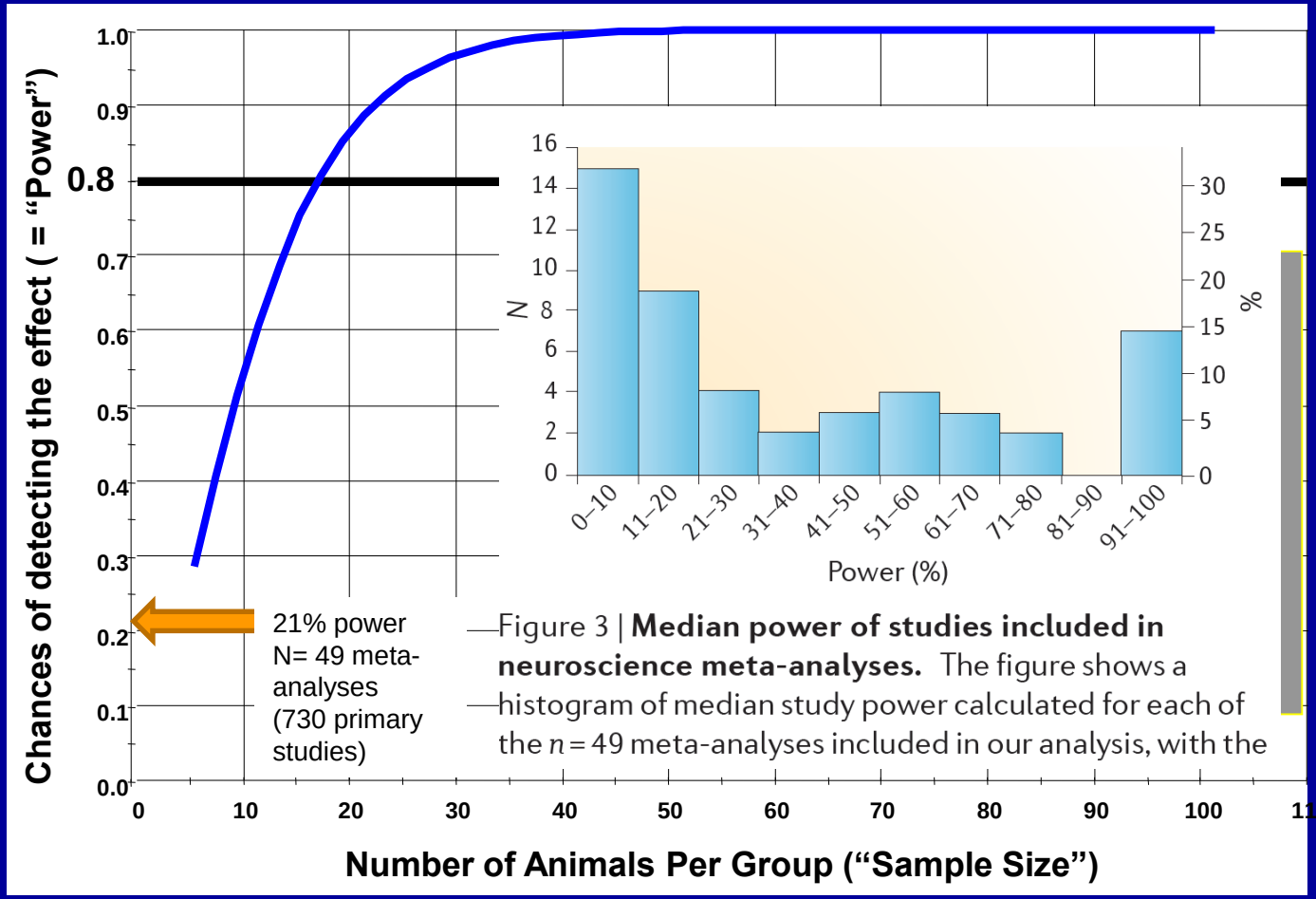
Power calculation,
using Power and Precision software
(they are many others)

+ GPOWER



How many animals do we need ?
(to have enough power)

Power


Chances of detecting the effect (= "Power")
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
10
20
30
40
50
60
70
80
90
100
110
Number of Animals Per Group ("Sample Size")
21% power
N= 49 meta-analyses
(730 primary studies)
N
%
0–10
11–20
21–30
31–40
41–50
51–60
61–70
71–80
81–90
91–100
Power (%)
Figure 3 | Median power of studies included in neuroscience meta-analyses. The figure shows a histogram of median study power calculated for each of the n = 49 meta-analyses included in our analysis, with the

POLL OPEN

1000 hypotheses tested. 100 are true. How many significant results expected with p< 0.05, 80% power.

1. 75

13.33%
2. 85

16.67%
3. 95

3.33%
4. **100**
0%
5. 105
0%
6. 115

6.67%
7. 125

43.33%
8. Don't know

16.67%

What we find

The Truth

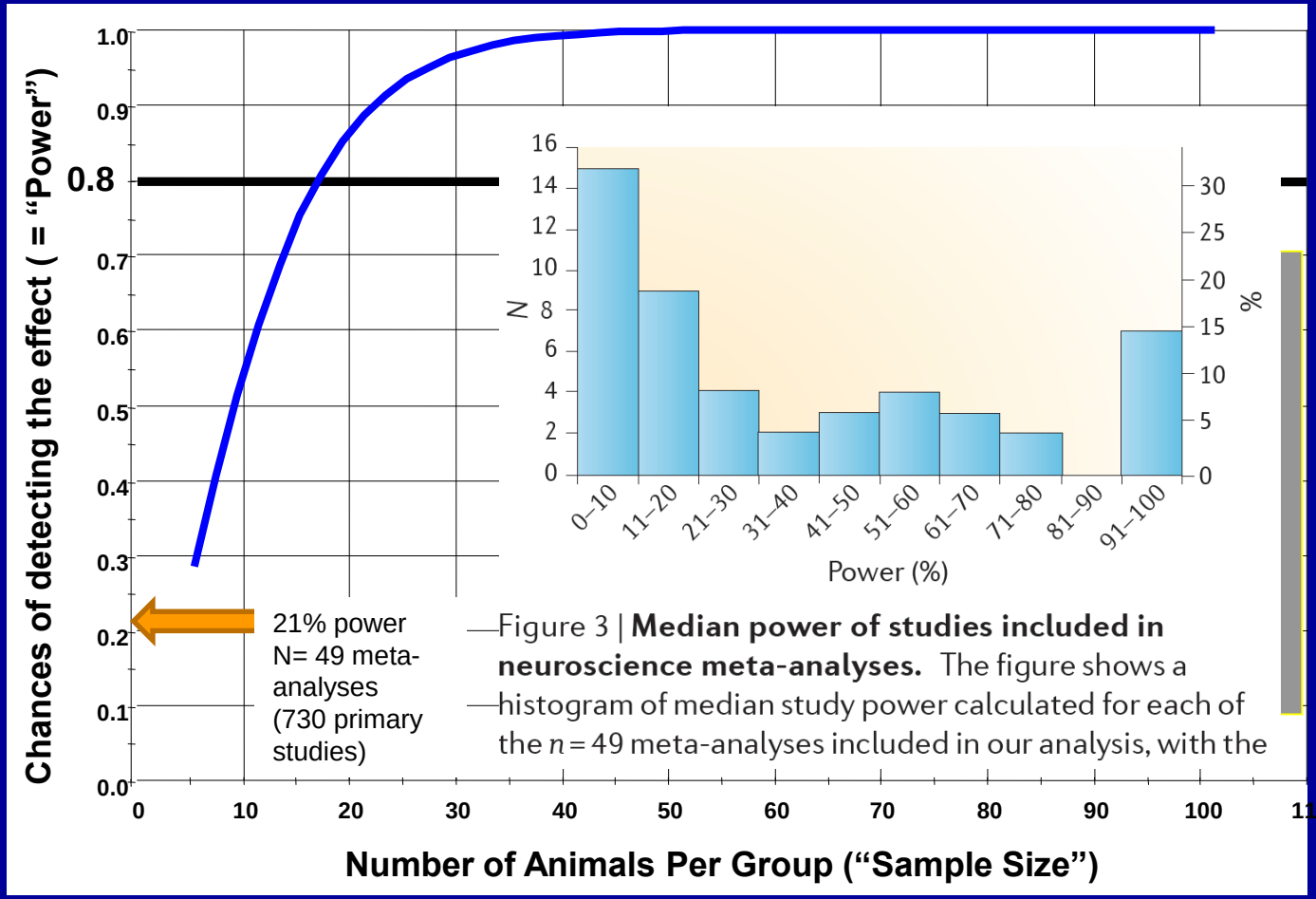
	Effect	**No effect**
No Effect	Type 1 error ($\alpha = 0.05$)	
Effect	$(1-\beta = 0.80)$	Type 2 error (β)

1000 hypotheses tested. 100 are true. How many significant results expected with p< 0.05, 80% power.

What we find

The Truth	Effect	No effect
900 No Effect	Type 1 error ($\alpha = 0.05$) 900 x 0.05 = 45	($1-\alpha = 0.95$) 900 x 0.95 = 855
100 Effect	$1-\beta = 0.80$ 100 x 0.8 = 80	Type 2 error ($\beta = 0.20$) 100 x 0.2 = 20
1000	45 + 80 = 125 (25% overestimate)	855 + 20 = 875 (3% underestimate)

Power


Chances of detecting the effect (= "Power")
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
10
20
30
40
50
60
70
80
90
100
110
Number of Animals Per Group ("Sample Size")
21% power
N= 49 meta-analyses
(730 primary studies)
N
%
0–10
11–20
21–30
31–40
41–50
51–60
61–70
71–80
81–90
91–100
Power (%)
Figure 3 | Median power of studies included in neuroscience meta-analyses. The figure shows a histogram of median study power calculated for each of the n = 49 meta-analyses included in our analysis, with the

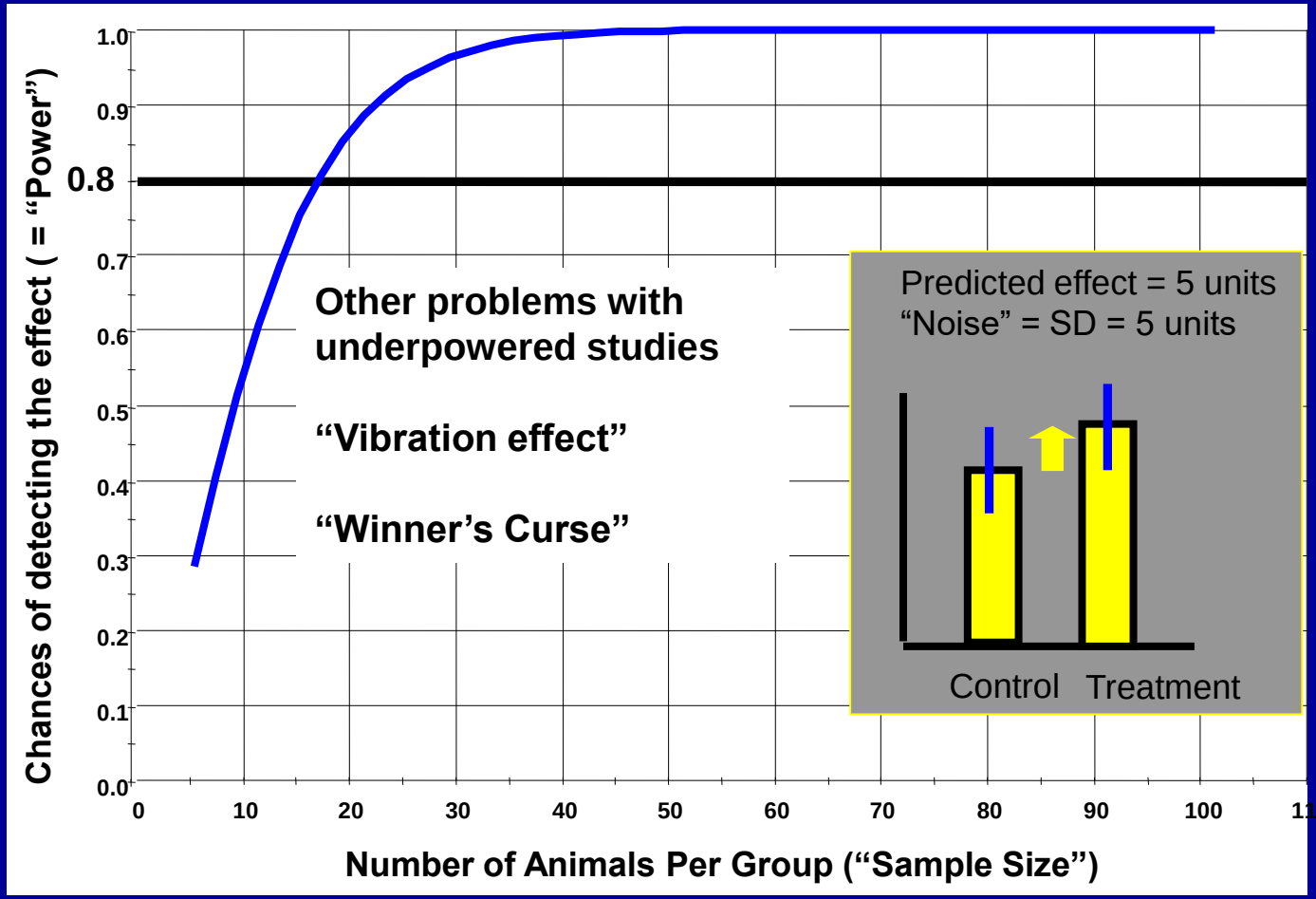
1000 hypotheses tested. 100 are true. How many significant results expected with p< 0.05, **20% power.**

What we find

The Truth	Effect	No effect
900 No Effect	Type 1 error (α = 0.05) 900 x 0.05 = 45	(1-α = 0.95) 900 x 0.95 = 855
100 Effect	1-β = ~~0.80~~ **0.20** 100 x 0.8 = 20	Type 2 error (β = 0.80) 100 x 0.2 = 80
1000	45 + 20 = 65 (35% underestimate)	855 + 20 = 935 (4% overestimate)

BUT most (69%) of significant results are wrong (45 out of 65).

Power


Chances of detecting the effect (= "Power")
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
10
20
30
40
50
60
70
80
90
100
110
Number of Animals Per Group ("Sample Size")
Other problems with underpowered studies
"Vibration effect"
"Winner's Curse"
Predicted effect = 5 units
"Noise" = SD = 5 units
Control
Treatment

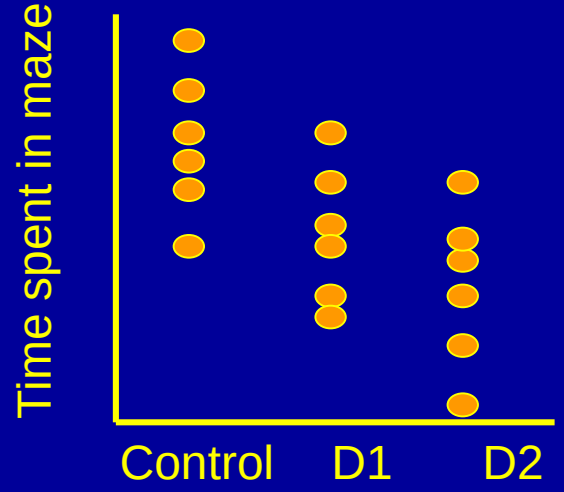
Sources of Noise

Age,
Sex,
Weight
Stress
Subclinical disease

Temperature etc
Animal House/Barn/Cage
Position in rack

The experimental unit ?

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



Look at the data !

This the topic of the next section…

Experimental Designs & Analyses

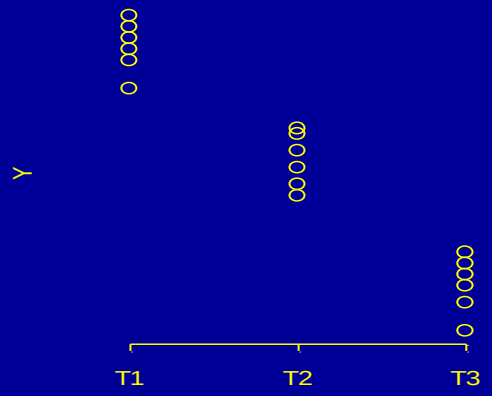
Example of designs:

- Factorial

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

Two common mistakes in biomedical-literature:

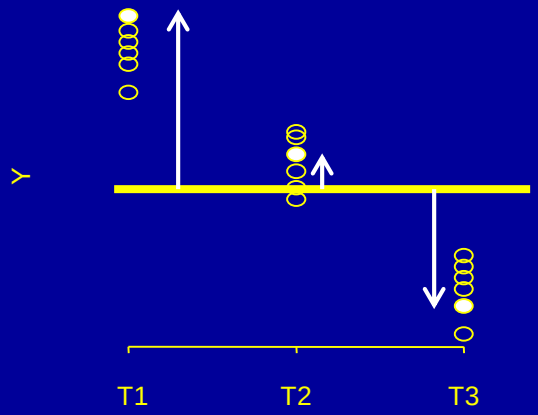
- multiple t-tests
- no blocking



N= 27 rats,
assigned to 3 levels of treatment

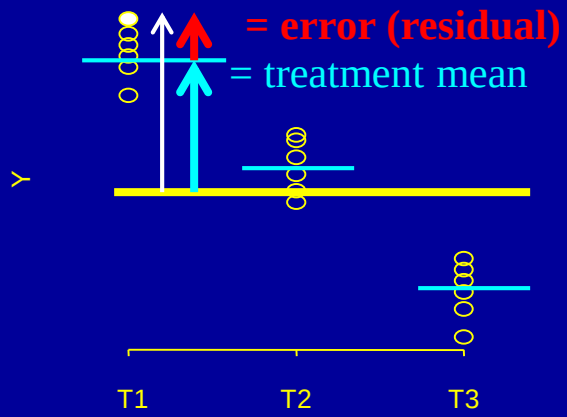
ANOVA table

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



=overall mean

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



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

POLL OPEN

Why Squares?

1. To remove negative signs (or they add up to zero)

62.96%

2. To give more weight to outliers

14.81%

3. Because that is best way to partition variation

11.11%

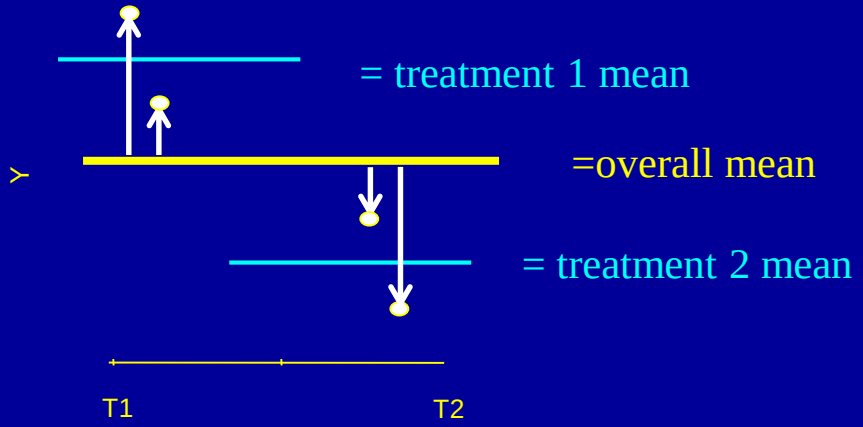
4. Because it works

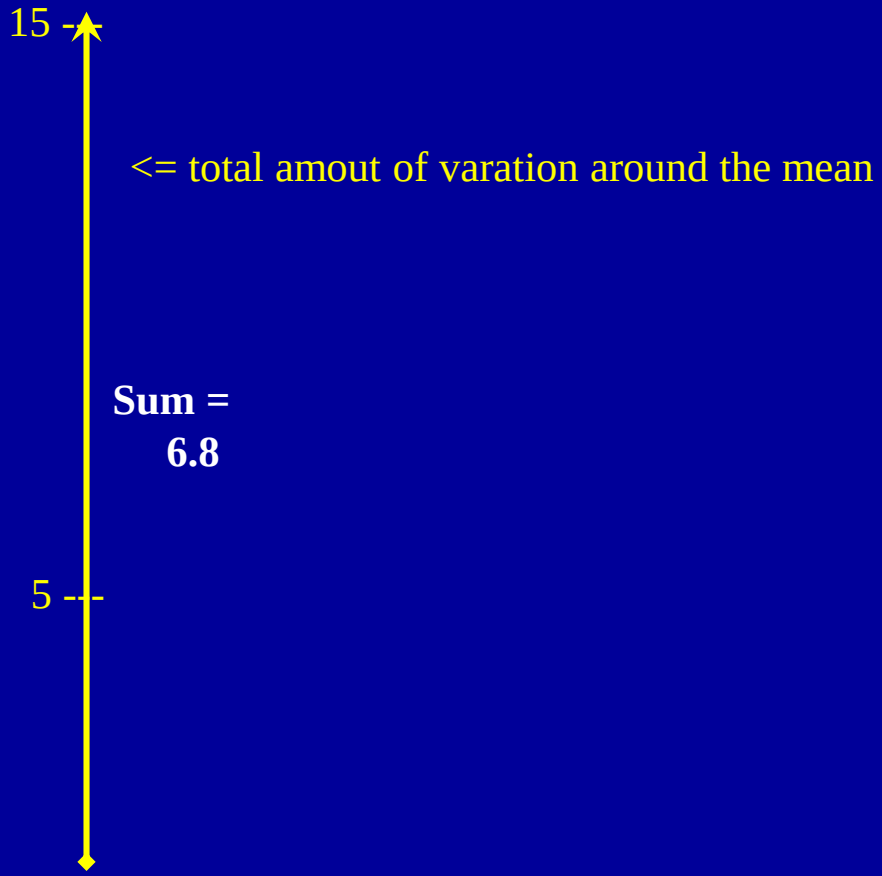
7.41%

5. Dont know

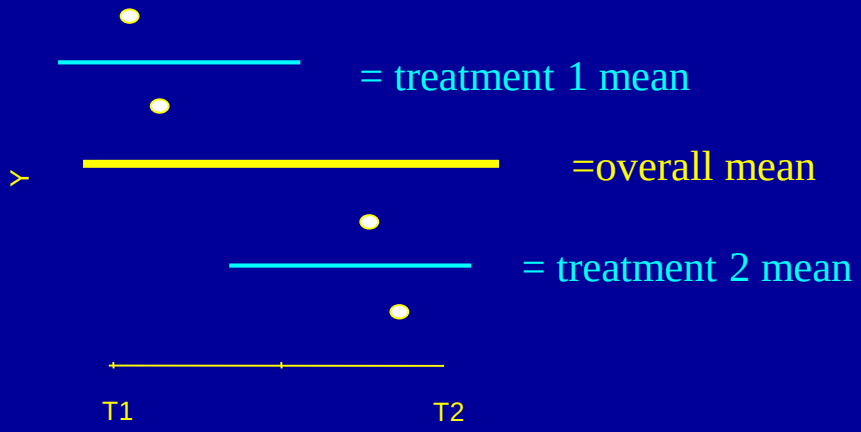
3.7%

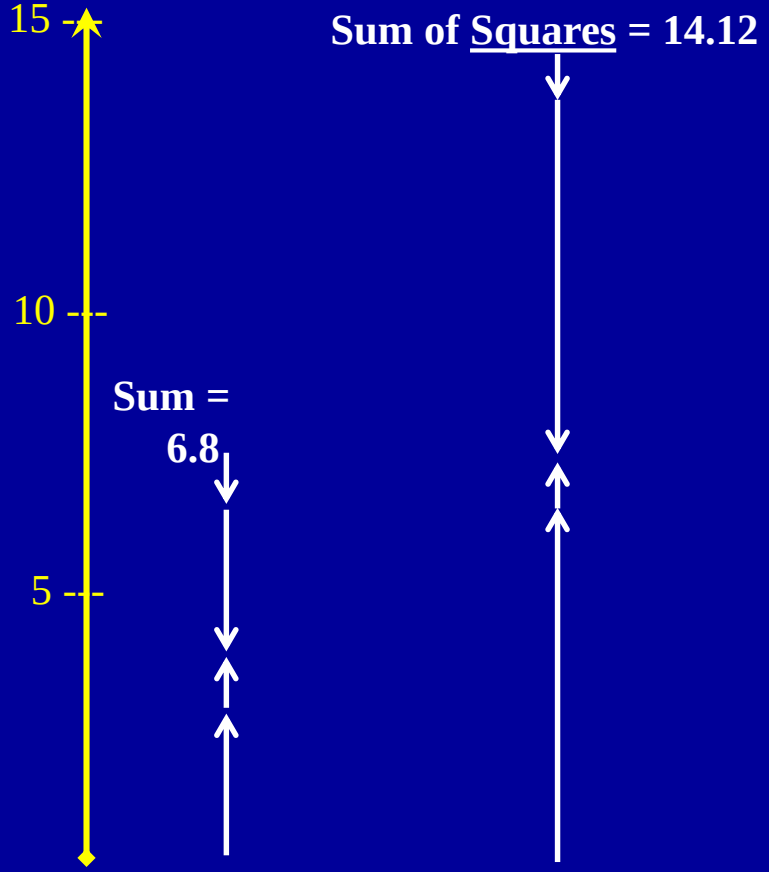
Why Squares?


= treatment 1 mean
=overall mean
= treatment 2 mean
T1
T2

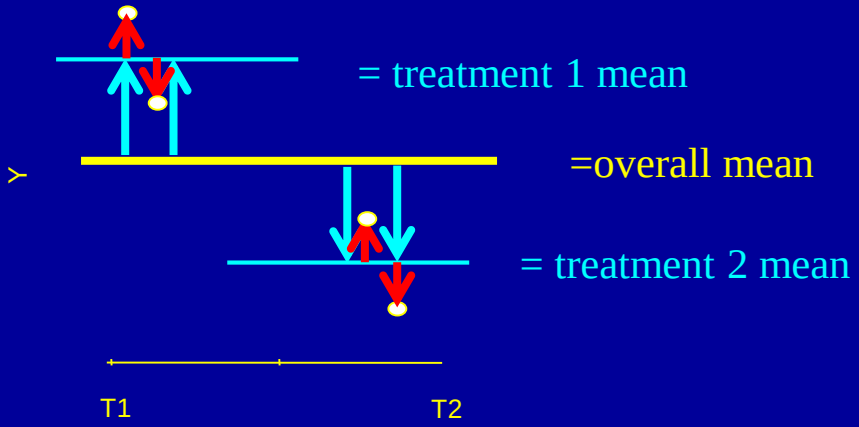

15
5
<= total amout of varation around the mean
Sum =
6.8

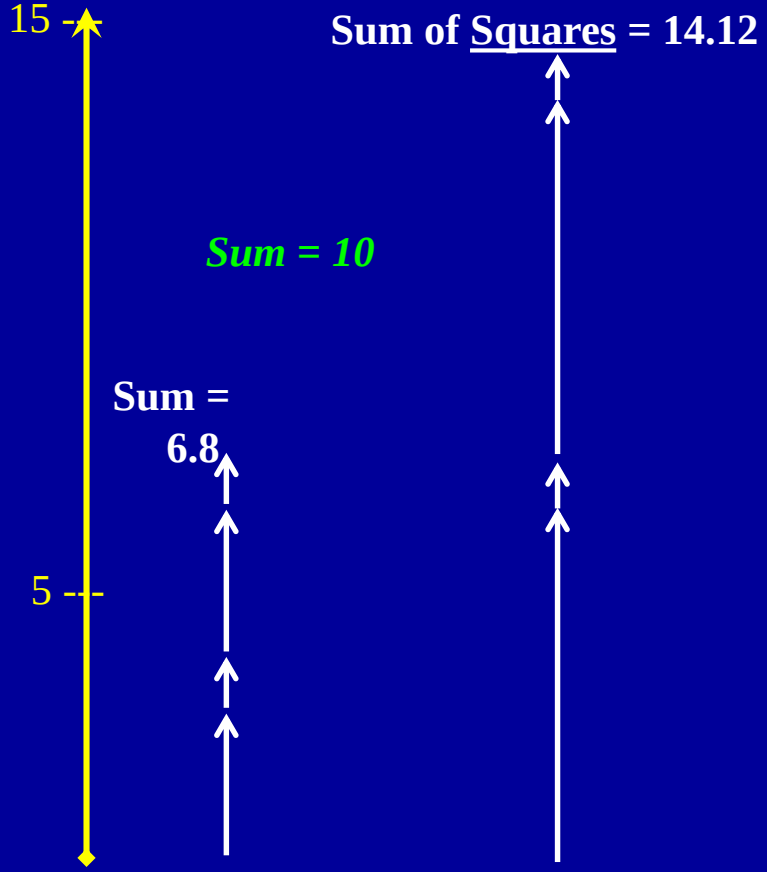
Why Squares?


= treatment 1 mean
=overall mean
= treatment 2 mean
T1
T2

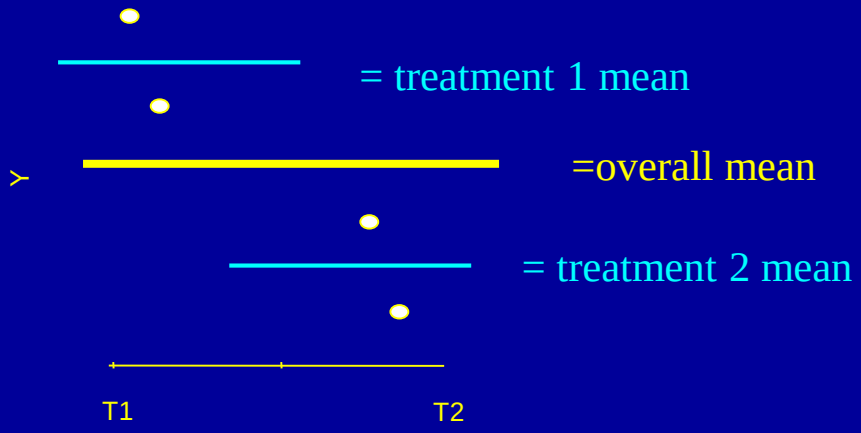

15
10
5
Sum = 6.8
Sum of Squares = 14.12

Why Squares?

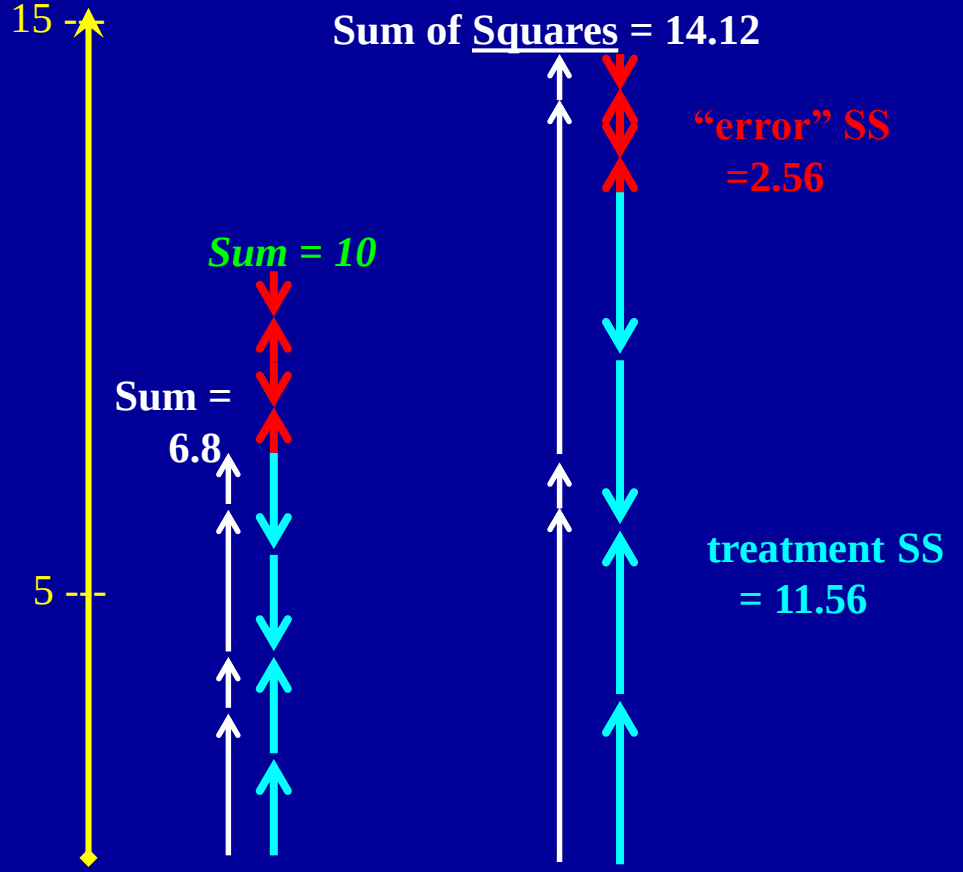

= treatment 1 mean
=overall mean
= treatment 2 mean
Y
T1
T2


15
5
Sum of Squares = 14.12
Sum = 10
Sum = 6.8

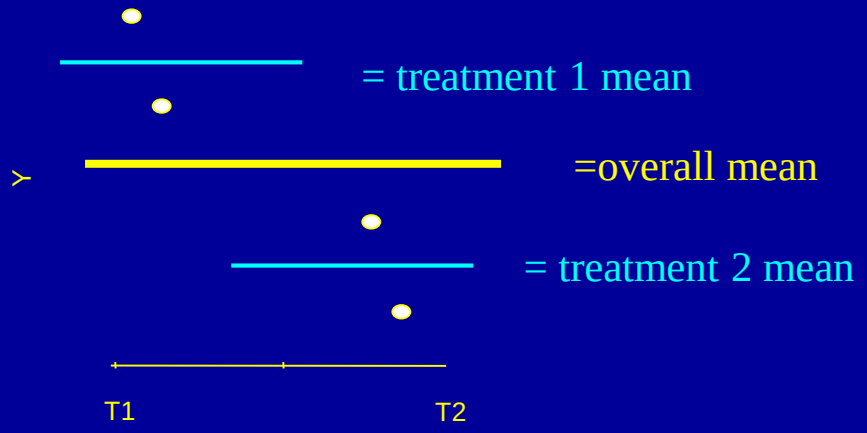
Why Squares?



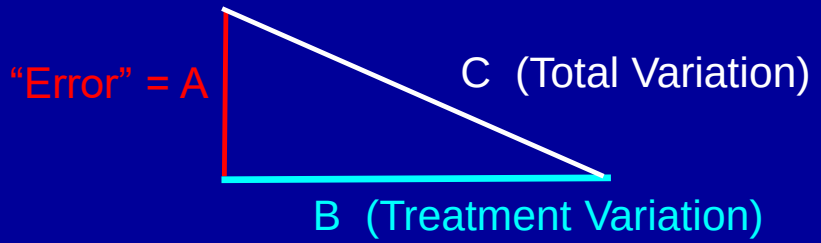
Source	DF	SS
Treatment	1	11.56
Error	2	2.56
Total	3	14.12

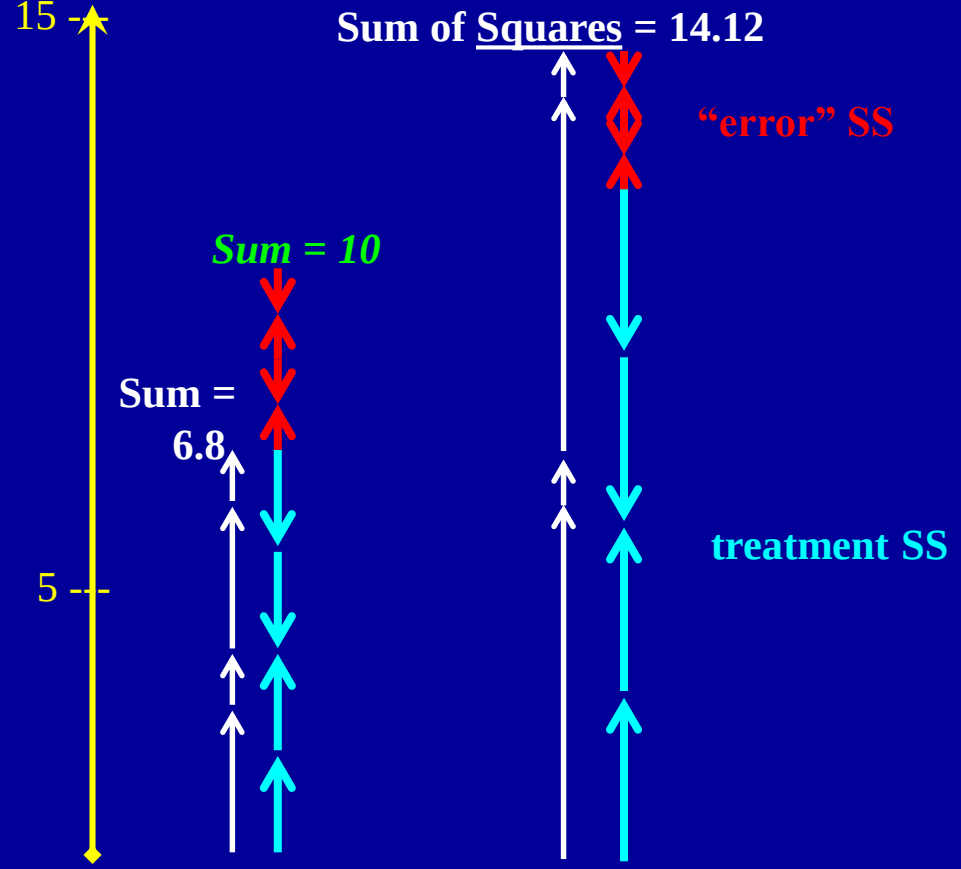


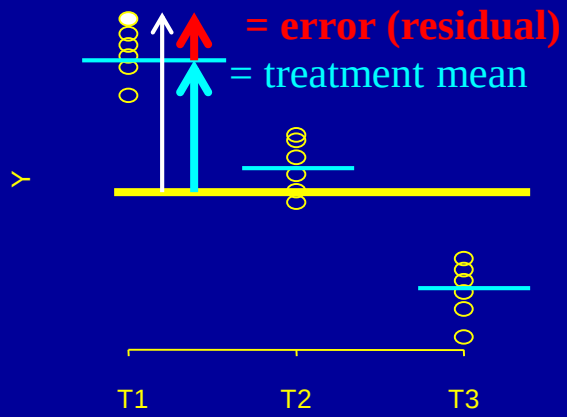
Why Squares?



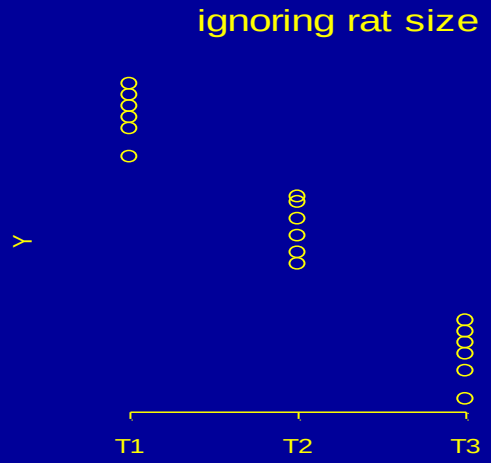
Source	DF	SS
Treatment	1	11.56
Error	2	2.56
Total	3	14.12

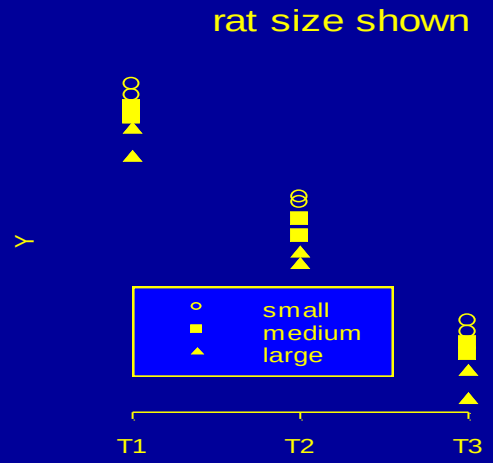






Source	DF	SS	MS	F	P
Treatment	2	2861	1430.5	13.13	0.00014
Error	24	2614	108.9		
Total	26	5475			





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

Blocking (i.e. controlling) for size

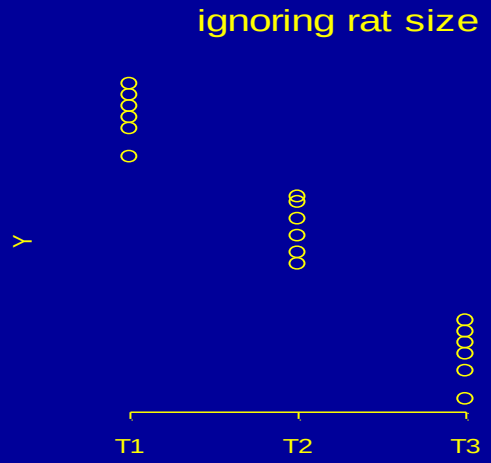
Big rats "block"	Medium rats	Small rats
Rat Rat Rat…	Rat Rat Rat…	Rat Rat Rat…

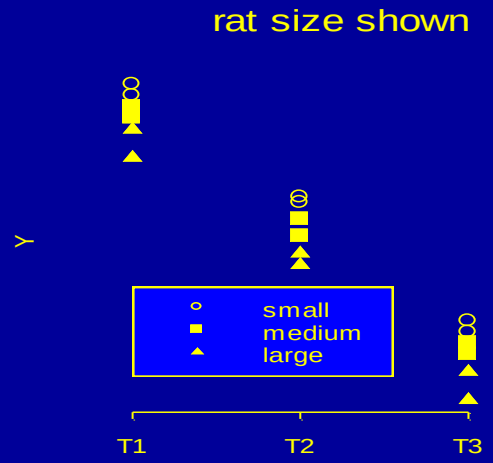
Blocking (i.e. controlling) for size

Blocking (i.e. controlling) for size

	Big rats "block"	Medium rats	Small rats
N= 9	**Rat Rat Rat**	Rat Rat Rat	Rat Rat Rat
N= 18	**Rat Rat Rat**	Rat Rat Rat	Rat Rat Rat
N= 27	**Rat Rat Rat**	Rat Rat Rat	Rat Rat Rat

Etc…



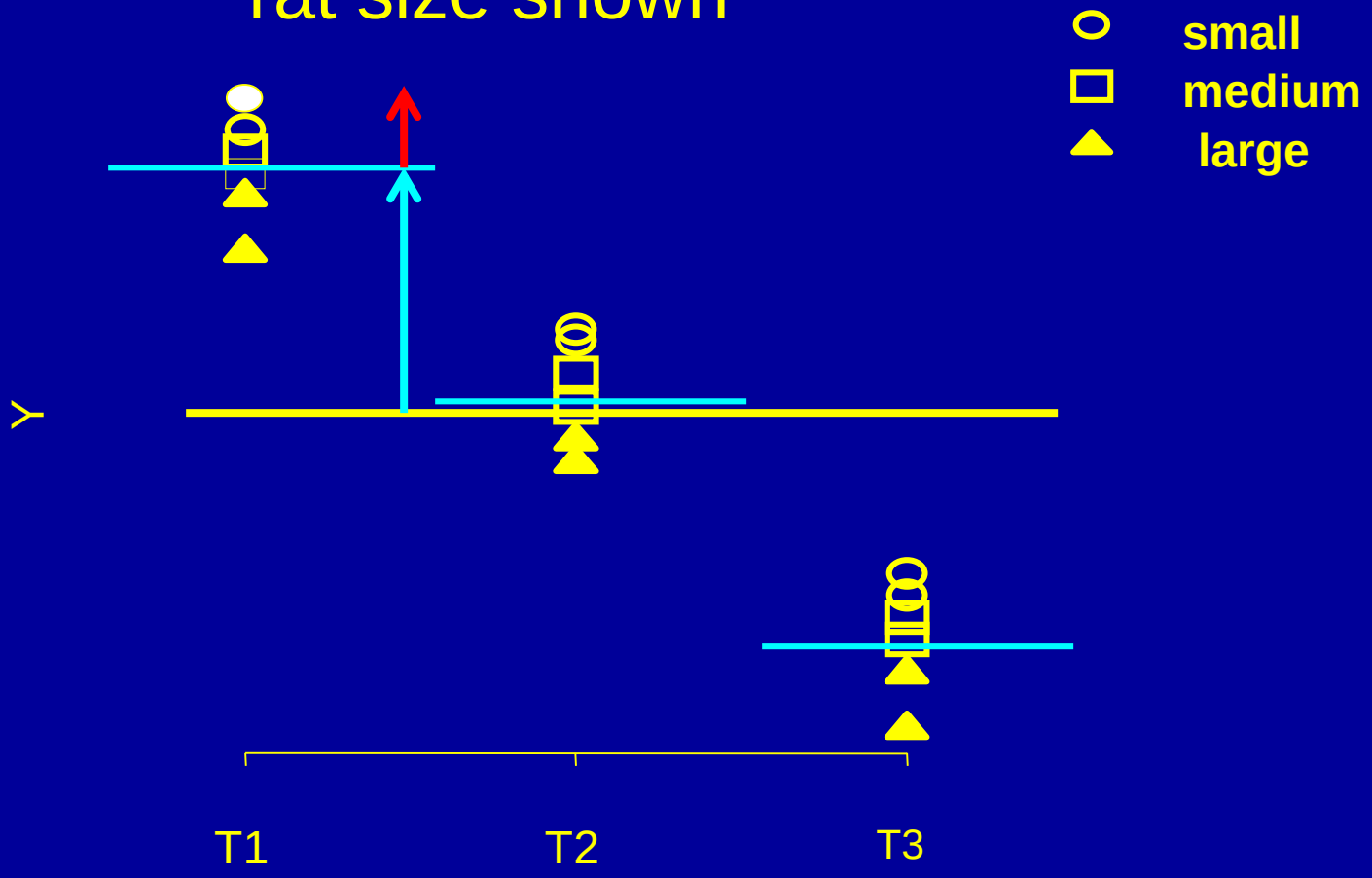


Source	DF	SS	MS	F	P
Treatment	2	2861	1430.5	13.13	0.00014
Error	24	2614	108.9		
Total	26	5475			

WITH BLOCKING

Source	DF	SS	MS	F	P
Block(=size)	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			

rat size shown



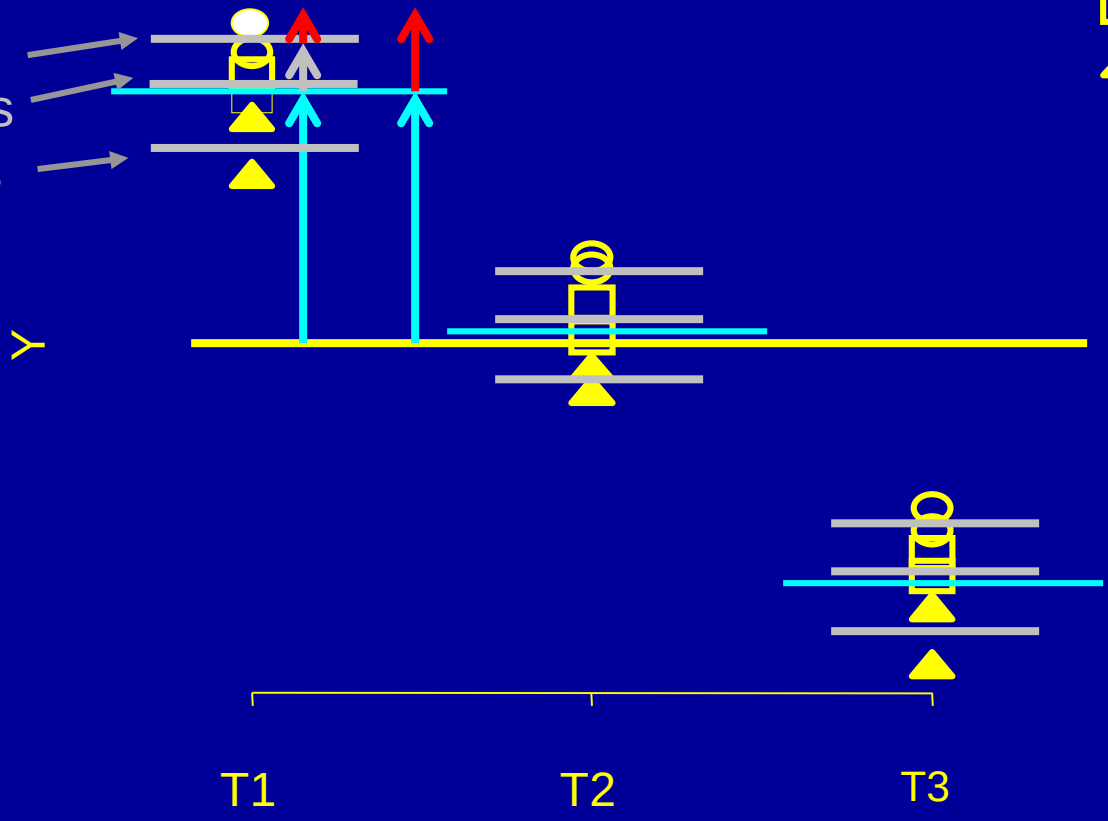
rat size shown

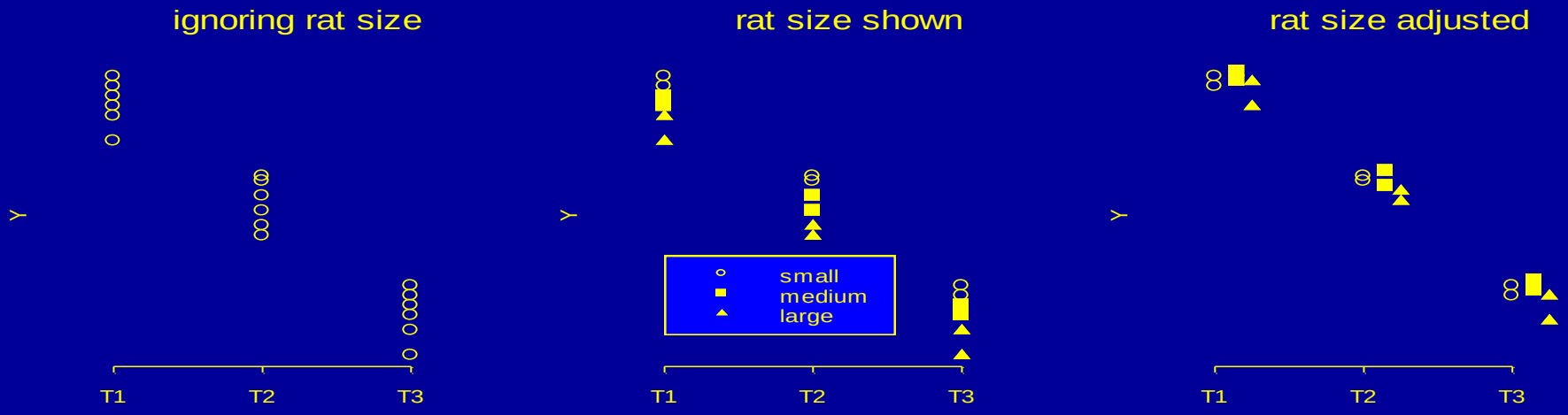
Additional info:

Mean small rats

medium rats

large rats



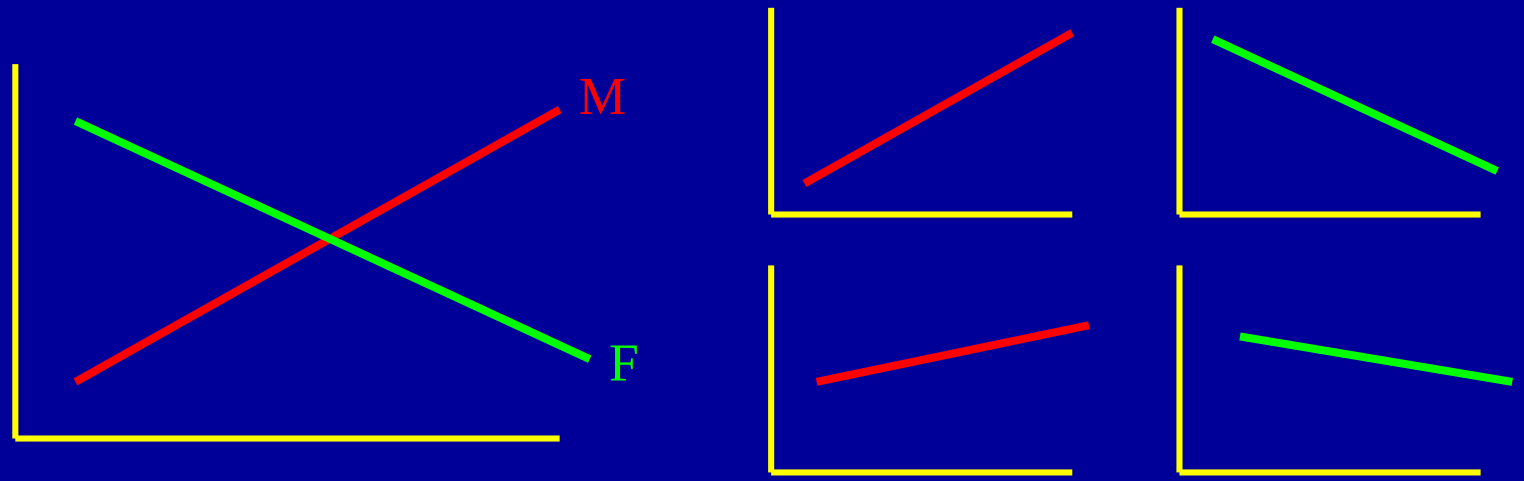


Source	DF	SS	MS	F	P
Treatment	2	2861	1430.5	13.13	0.00014
Error	24	2614	108.9		
Total	26	5475			

WITH BLOCKING

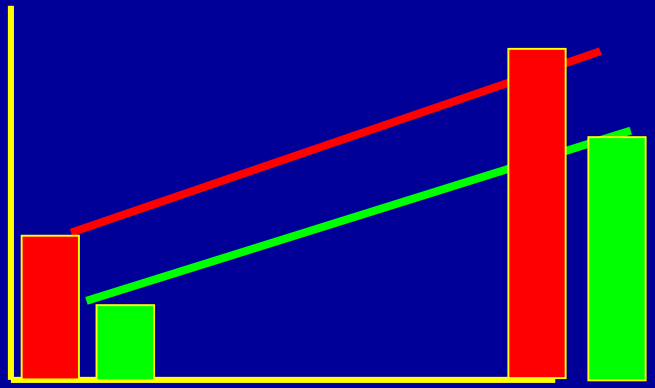
Source	DF	SS	MS	F	P
Block(=size)	2	2279	1139.7	74.94	0.0000
Treatment	2	2861	1430.5	94.06	0.0000
Error	22	334	15.2		
Total	26	5475			

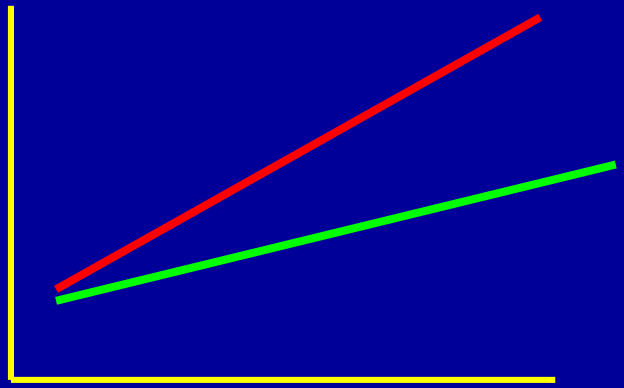
“interactions”: information for (almost) free



	df SS MS F p
Sex	(overall diff. between the sexes ?)
Drug	(overall diff. between the drug treatments ?)
Sex*Drug	**Interaction ?**

“interactions”: information for (almost) free



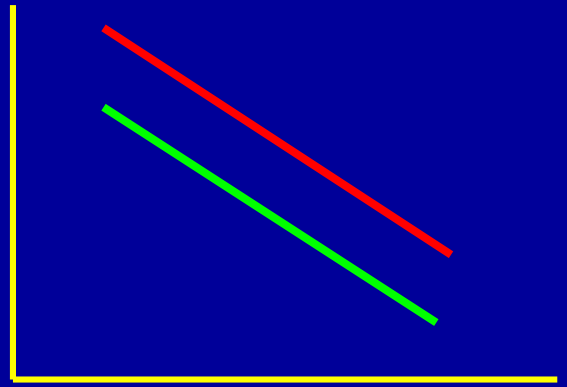


	df SS MS F p
Sex	(overall diff. between the sexes)
Drug	(overall diff. between the drug treatments)
Sex*Drug	**Interaction**

“interactions”: information for (almost) free

POLL OPEN

	df SS MS F p
Sex	Yes
Drug	Yes
Sex*Drug	No (**Interaction**)



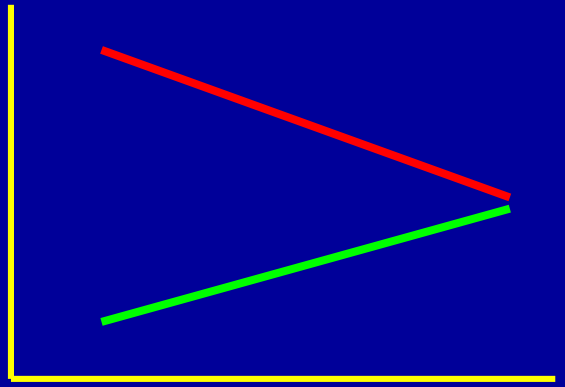
Interaction ?

1. *YES*
   4.35%
2. *NO*
   91.3%
3. *Not sure*
   0%
4. *Panic:*
   0%
5. *Please explain the principle again*
   4.35%

“interactions”: information for (almost) free

POLL OPEN

	df	SS	MS	F	p
Sex	Yes				
Drug	No				
Sex*Drug	Yes **(Interaction)**				



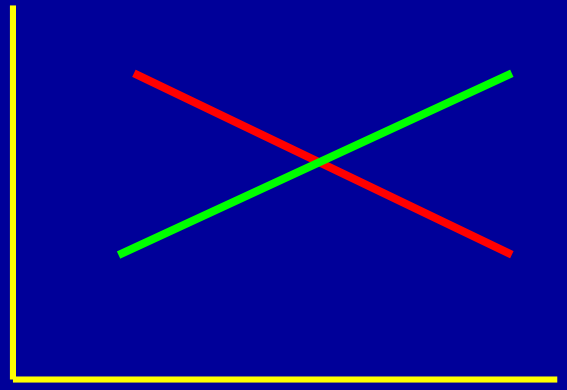
Interaction ?

1. *YES* — 93.75%
2. *NO* — 3.13%
3. *Not sure* — 0%
4. *Panic:* — 0%
5. *Please explain the principle again* — 3.13%

“interactions”: information for (almost) free

POLL OPEN

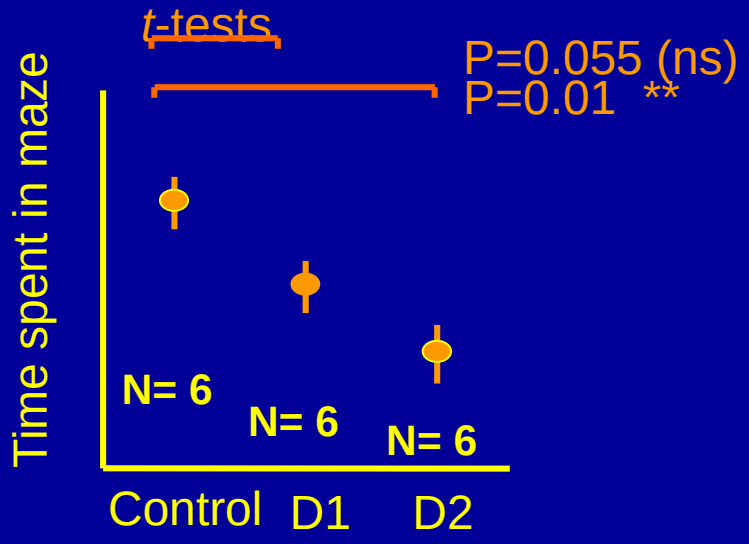
	df	SS	MS	F	p
Sex	No				
Drug	No				
Sex*Drug	Yes **(Interaction)**				



Interaction ?

1. *YES* *84.62%*
2. *NO* *7.69%*
3. *Not sure* *7.69%*
4. *Panic:* *0%*
5. *Please explain the principle again* *0%*

Multiple comparisons



Variety of tests. Most common are:

1. Comparison of selected pairs of mean: Bonferroni tests; (equivalent of multiple t tests with correction for multiplicity) but harsh (low power); not recommended for 5 groups or more.

2. One group (e.g control) against all the others = Dunnett's test

3. Compairing means of preselected groups A & B vs C&D&E = Contrast

4. All pairs of means = Tukey's or Student-Newman's test (Roughly the same: Tukey said to be more conservative ie more false negatives (Type II error), and SN more false positives (Type I error)

Statistics, Experimental Design, and Animal Experimentation.

- Searching for Truth
- Power (Calculations)
- Exp. Design & Analyses: controlling variation
- **Quick Recap**
- Refinement vs Reduction

Thinking (List) Exercises

Document 1

Final look at 1. Test of two drugs on maze ability

Searching for Truth = Experiment

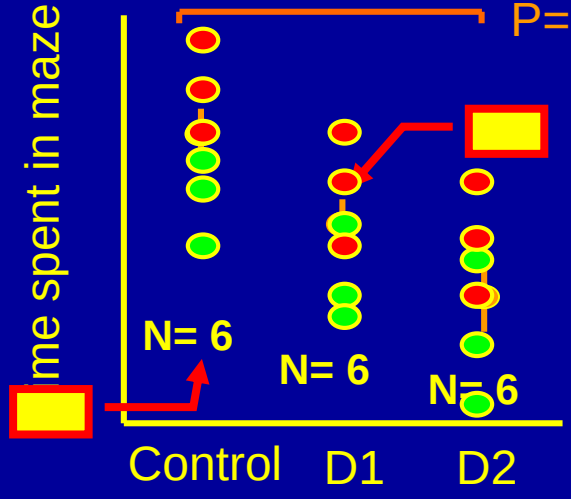
Test of drug doses on cognitive abilities (tested in maze).

Exp 1: 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

t - tests

P=0.055 (ns)
P=0.01 **

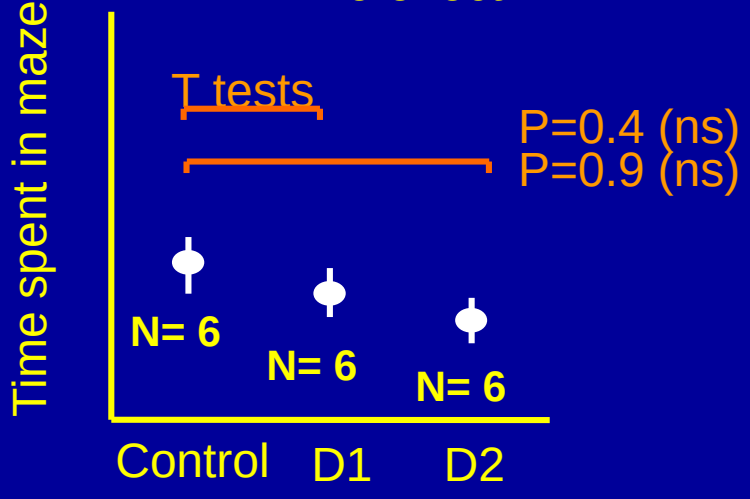


Exp.2 Repeated on females

Conclusion:
D1 no effect.
D2 no effect

T tests

P=0.4 (ns)
P=0.9 (ns)



Searching for Truth = Experiment

Test of drug doses on cognitive abilities (tested in maze).

Exp 1: 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

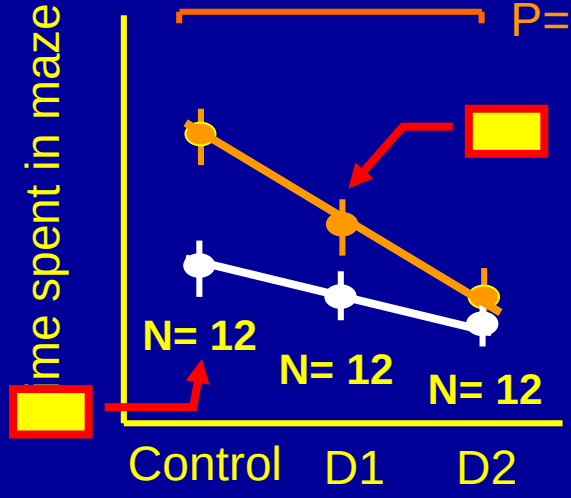
t - tests

P=0.055 (ns)

P=0.01 **

Exp.2 Repeated on females

Combine experiments



Factorial Design

	Control	D1	D2
Males	N=6	6	6
Females	6	6	6
N=	**12**	**12**	**12**

Searching for Truth = Experiment

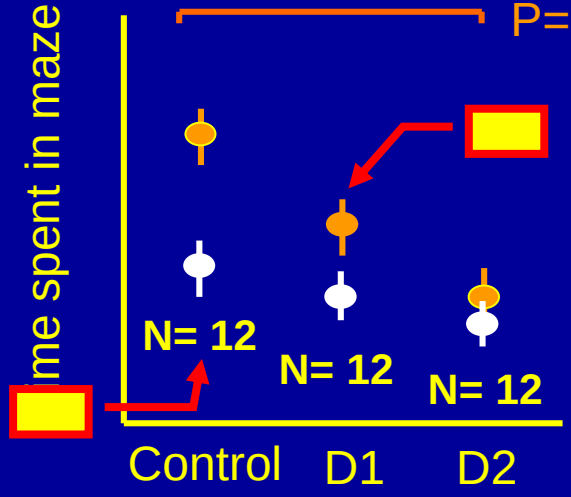
Test of drug doses on cognitive abilities (tested in maze).

Exp 1: 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

t - tests

P=0.055 (ns)
P=0.01 **



Exp.2 Repeated on females

Combine experiments

Factorial Design

		Control	D1	D2
Males	Small	N=3	3	3
	Large	3	3	3
Females	Small	3	3	3
	Large	3	3	3
N=		**12**	**12**	**12**

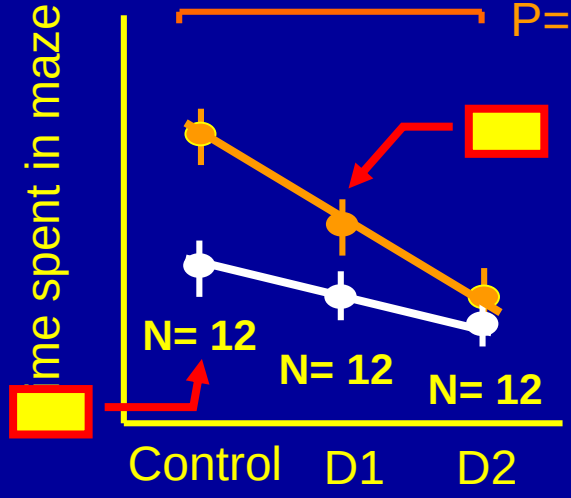
Searching for Truth = Experiment

Test of drug doses on cognitive abilities (tested in maze).

Exp 1: 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
t - tests

P=0.055 (ns)
P=0.01 **



~~**Exp.2** Repeated on females~~

Conclusion
D1 no effect.
D2 no effect

Conclusion NOW:

1. Overall: very significant effect (P<0.0001)
2. ("Rat weight": significant effect) (P<0.001)
3. D1 significant effect. (P<0.01)
4. D2 significant effect. (P<0.001)
5. Sex: Females generally quicker (P< 0.01)
6. (interaction) Effect of drug significantly different on the two sexes (P<0.001)

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

Statistics, Experimental Design, and Animal Experimentation.

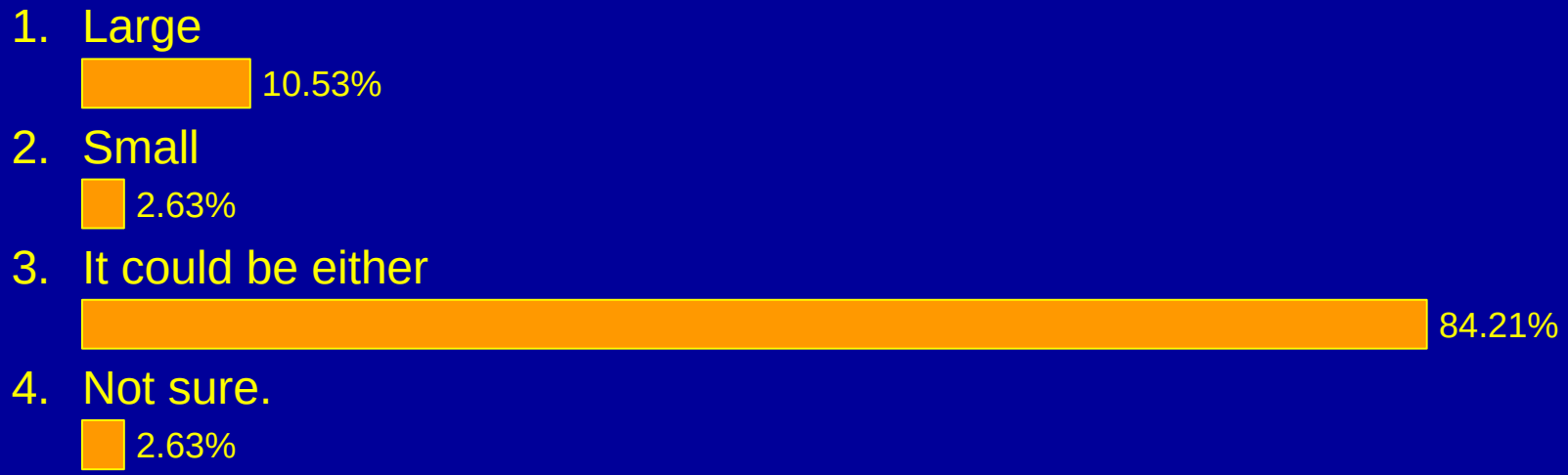
- Searching for Truth
- Power (Calculations)
- Exp. Design & Analyses: controlling variation
- Quick Recap
- Refinement vs Reduction

Quick Post course Quiz

9 questions

POLL OPEN

A highly statistically significant effect
(e.g. p <0.001) means that the biological effect has to be:



POLL OPEN

When comparing groups, twice as much noise in the data means that we require how many animals to show the same statistical significance?

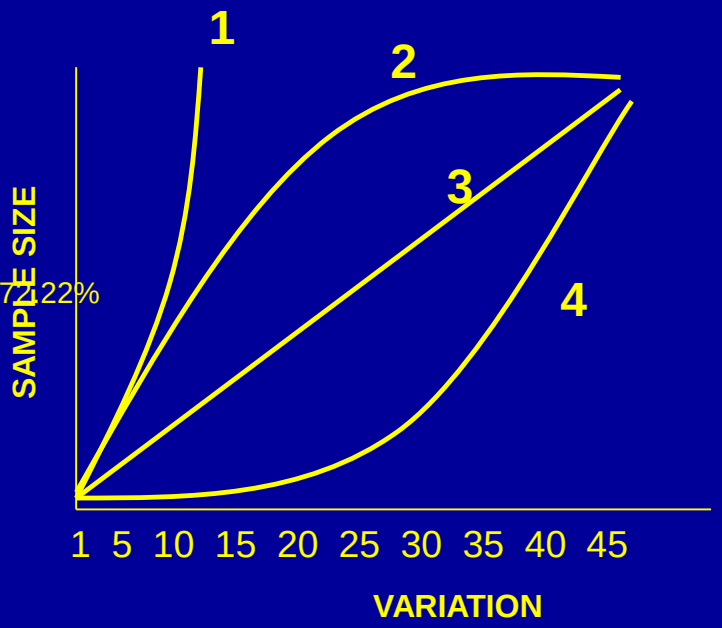
1. Same number
   0%
2. 50% more animals
   0%
3. Twice as many
    2.56%
4. Four times as many
    97.44%
5. Not sure
   0%

POLL OPEN

What is the effect of Variation on Sample size (needed to obtain statistical significance).

WRONG : 1 is the right answer !

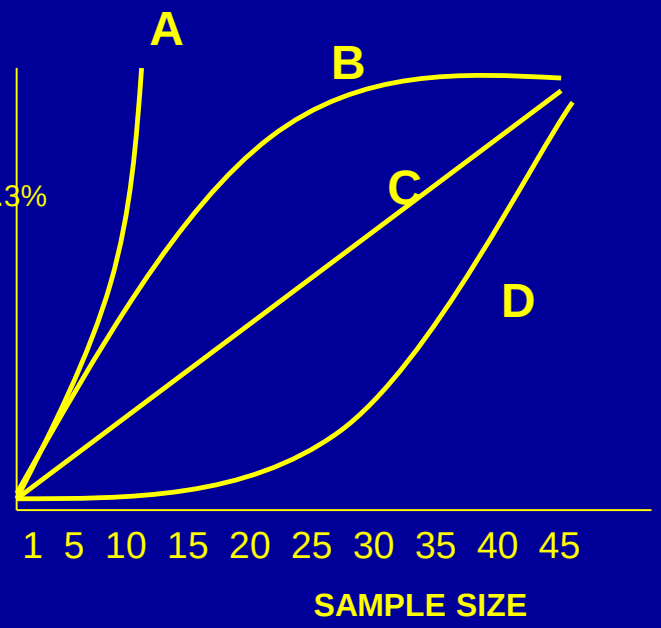
1. 1
   13.89%
2. 2
   72.22%
3. 3
   0%
4. 4
   11.11%
5. Don't know
   2.78%
6. Don't understand the question.
   0%



POLL OPEN

What is the relationship between Sample Size and the Power of an experiment

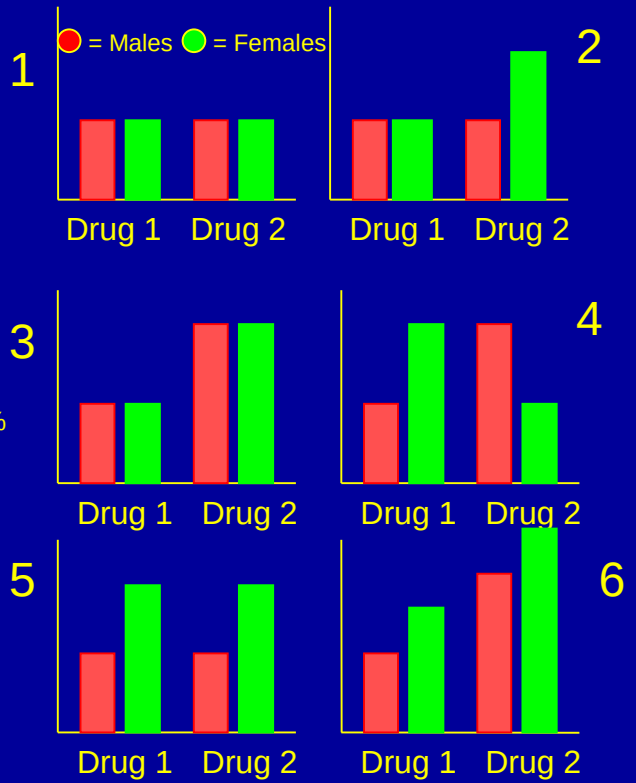
1. A
   0%
2. B
3. C
   2.7%
4. D
   0%
5. Don't know
   0%
6. Don't know what Power means
   0%



POLL OPEN

Statistically speaking, which of these results show a drug * sex interaction (multiple answers allowed)

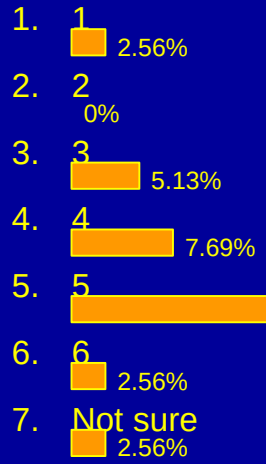
1. 1
   0%
2. 2
   75%
3. 3
   0%
4. 4
   93.75%
5. 5
   31.25%
6. 6
   37.5%
7. Not sure
   0%
8. Don’t know what interaction means
   0%

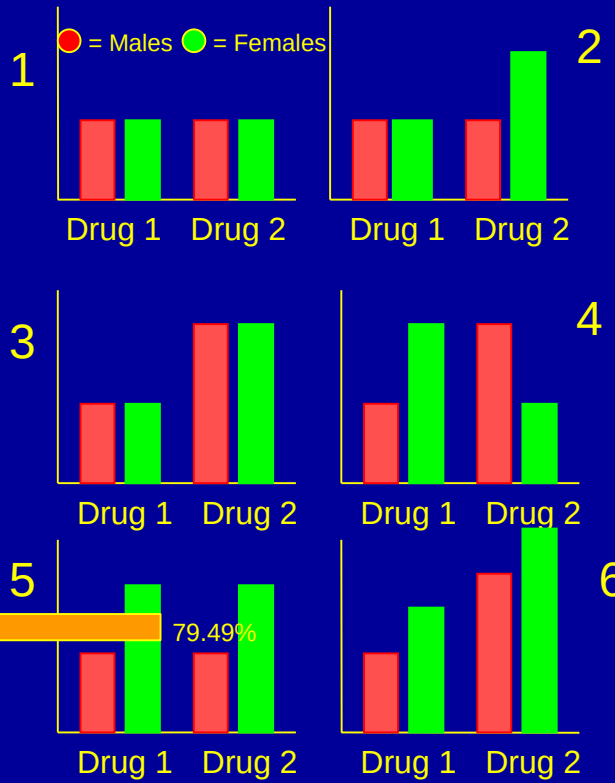


POLL OPEN

The analysis table below corresponds to which graph?

Source	DF	SS	MS	F	P
Sex	1	100	100	25	0.001
Drug	1	5	5	0.2	0.7
Sex*Drug	1	3	3	0.1	0.8
Error	33	108	3.3		
Total	37	208			





POLL OPEN

What is the relationship between SE, SD and N ?

1. 1. SE = SD * √N
   0%
2. 2. SE = SD * N
   0%
3. 3. SE = SD / N
   0%
4. 4. SE = SD /√N
   
   100%
5. 5. SE = √N / SD
   0%
6. 6. Not sure
   0%

What is a type 1 error about?

1. Chance of obtaining a false positive

2. Chance of obtaining a false negative

2.7%
3. Fundamental error at the data collection stage
0%
4. Fundamental error at the analysis stage
0%
5. Fundamental error at the experimental design stage
0%
6. Not sure (but it sounds worse than a type 2 error !)
0%

ET VOILA

Thank you