# Three kinds of lies- a brief introduction to statistics

**Dr Richard Mead**
Kenneth A Snowman - MND Association Lecturer
Dept. Neuroscience, University of Sheffield

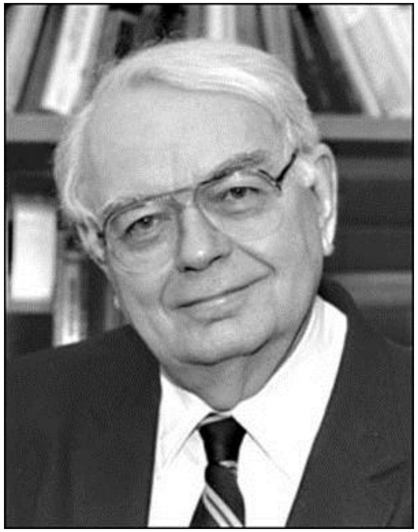
The University Of Sheffield.

SITraN
Sheffield Institute for Translational Neuroscience

# "There are three kinds of lies: lies, damned lies and statistics"

Attributed to Benjamin Disraeli (1804-1881) by Mark Twain

"Its easy to lie with statistics, but it's easier to lie without them" Charles Frederick Mosteller (statistician)



# Why do we need statistics

- To separate what is a real difference (in an experiment) and what is caused by 'noise' (random biological and experimental variation)
- Human brains are built to find patterns – need something objective to confirm our judgement

The University Of Sheffield.

SITraN Sheffield Institute for Translational Neuroscience

# Basic idea of statistics

- Using limited amounts of data to draw general conclusions
- Sample versus population
- Gather data from a sample of a population
- Use rules of probability to draw conclusions about the whole population
- What is the 'sample' and 'population' in the following?
  - Clinical trial
  - Enzyme assay on tissue cultured cells

# Need for independent samples

- Statistical tests assume that each experimental unit (variable) was sampled (measured) independently of the others
- How many independent experimental units:
  - You are measuring blood pressure in animals. You have five animals in each group, and measure the blood pressure three times in each animal.
  - You have done a biochemical experiment three times, each time in triplicate.

The University Of Sheffield.

SITraN
Sheffield Institute for Translational Neuroscience

# Types of data (levels of measurement)

- Nominal scale (categories)
  - e.g gender, strain, genotype
- Ordinal Scale (ranks)
  - scoring system (e.g. Glasgow coma scale)
- Interval scale
  - E.g. Temperature in celcius
- Ratio scale
  - Most of the things we measure (e.g. weight, enzyme activity)

Increasing information

The University Of Sheffield.

# Why is this important?

SITraN Sheffield Institute for Translational Neuroscience

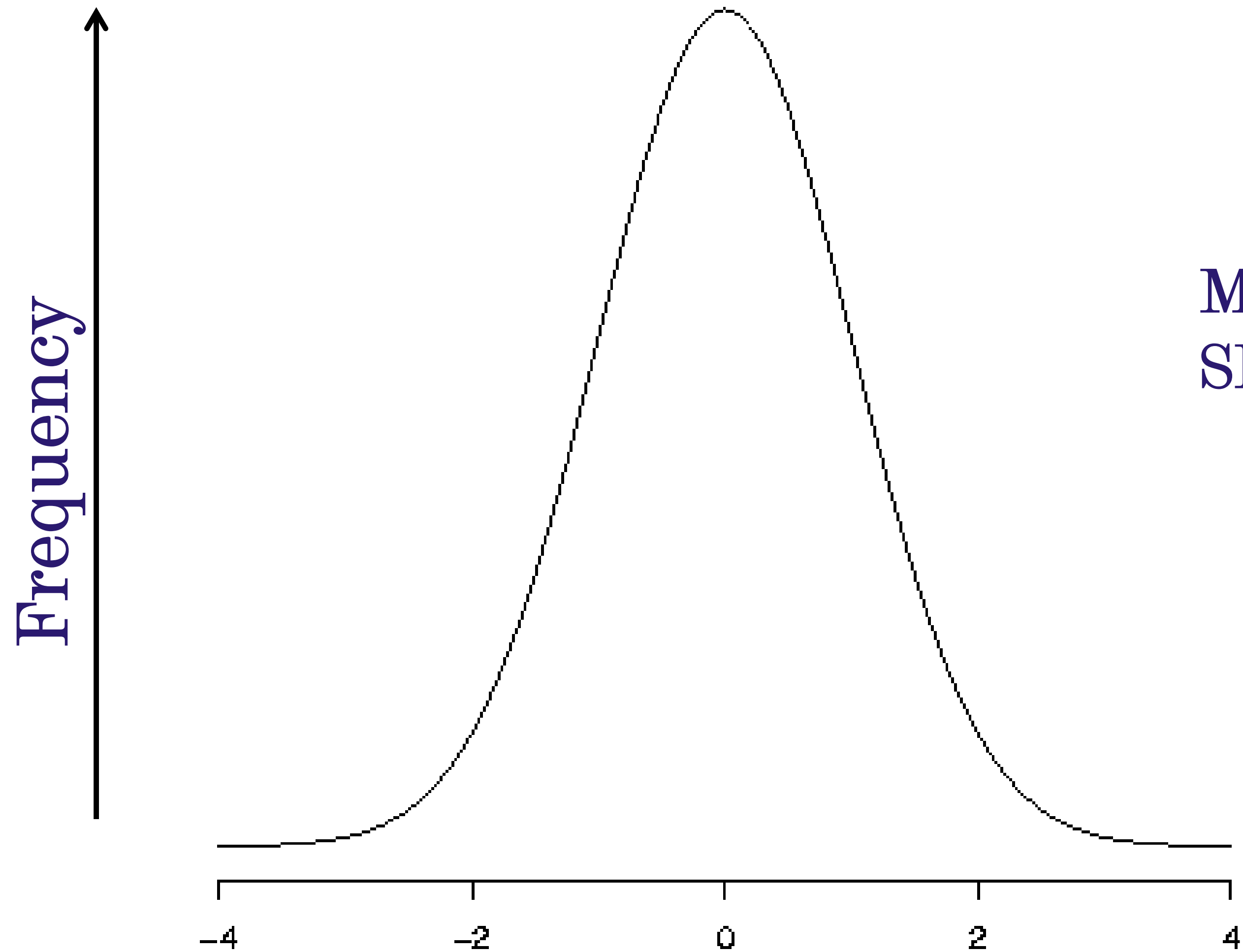
- Experiments designed to find relationships between variables
- Dictates how to manipulate and present your data
- Dictates what statistical tests are appropriate

| **OK to compute:** | **Nominal** | **Ordinal** | **Interval** | **Ratio** |
|---|---|---|---|---|
| Frequency distribution | Yes | Yes | Yes | Yes |
| Median and percentiles | No | Yes | Yes | Yes |
| Add or subtract | No | No | Yes | Yes |
| Mean, standard deviation, SEM | No | No | Yes | Yes |
| Ratio, or coefficient of variation | No | No | No | Yes |
| **Statistical tests:** | | | | |
| Non-parametric | Yes | Yes | Yes | Yes |
| Parametric | No | No | Yes | Yes |

# Parametric vs Non-parametric

- Parametric
  - Assumes the data you have sampled has come from a particular probability distribution
- Non-parametric
  - Probability distribution free
- What distribution is typically used in biomedical applications?

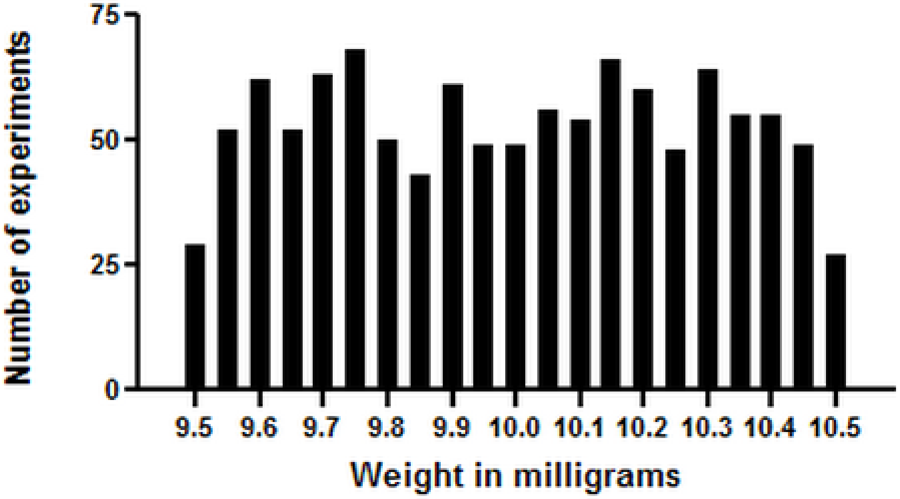
# The Normal (or Gaussian) Distribution

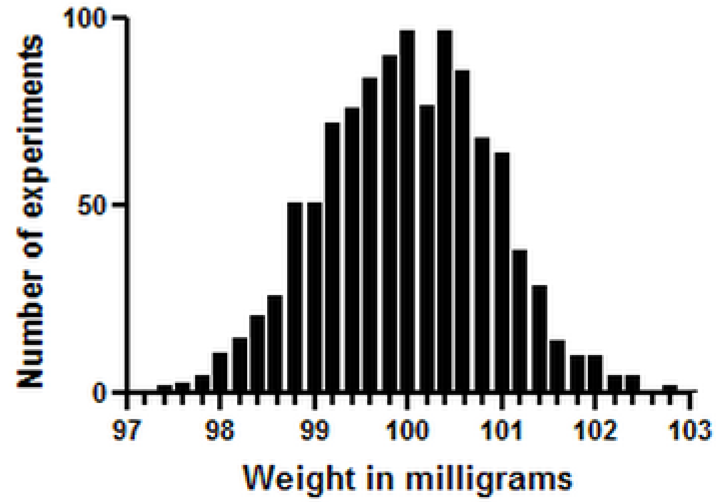


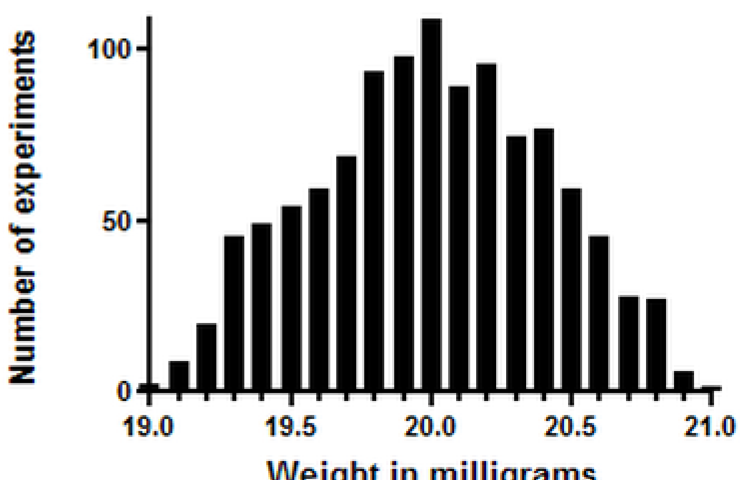

Mean = 0

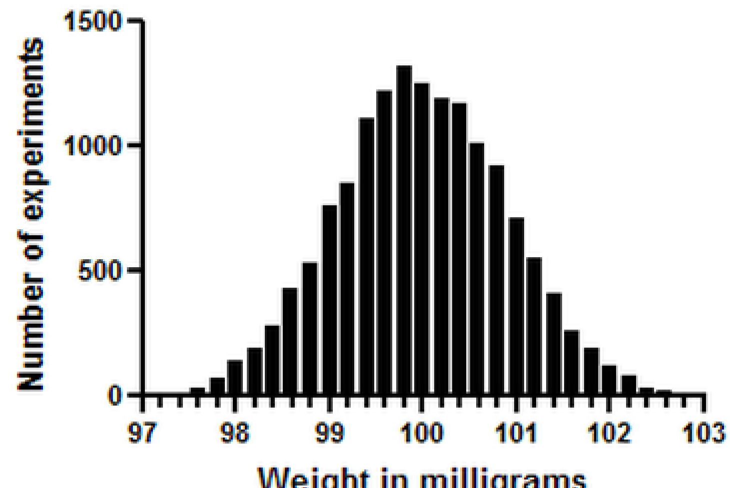
SD =1

# The Normal (or Gaussian) Distribution

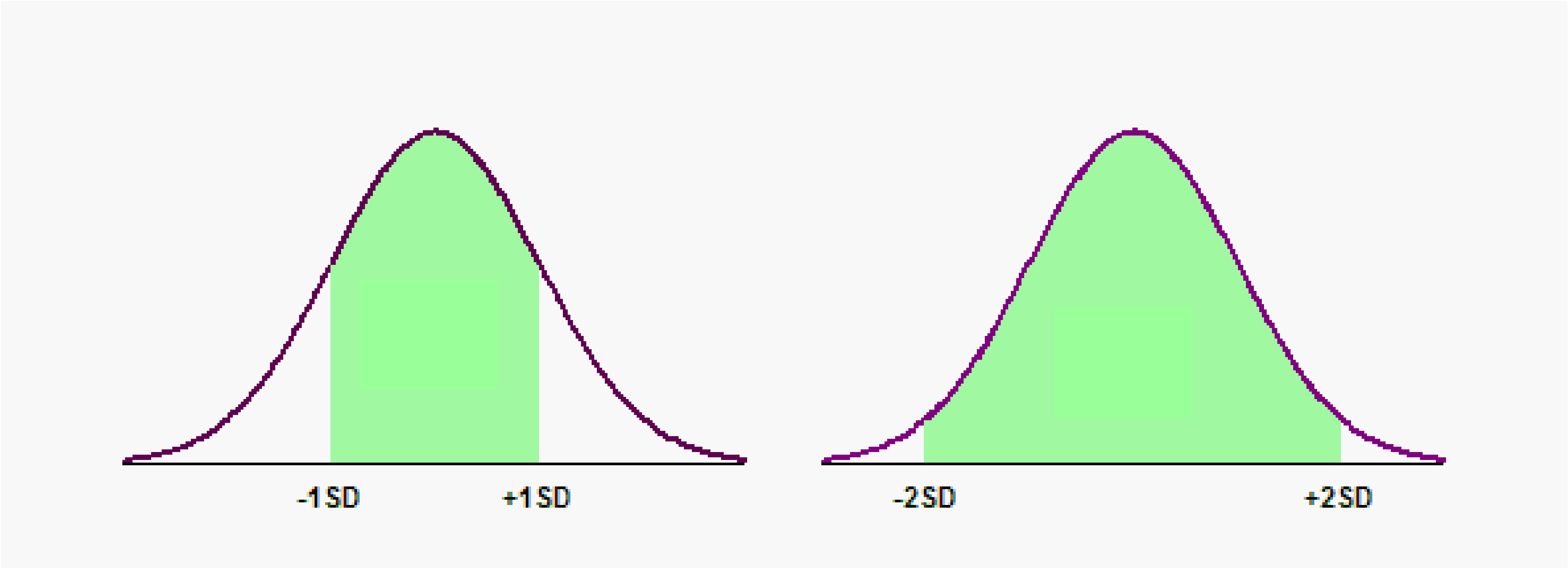


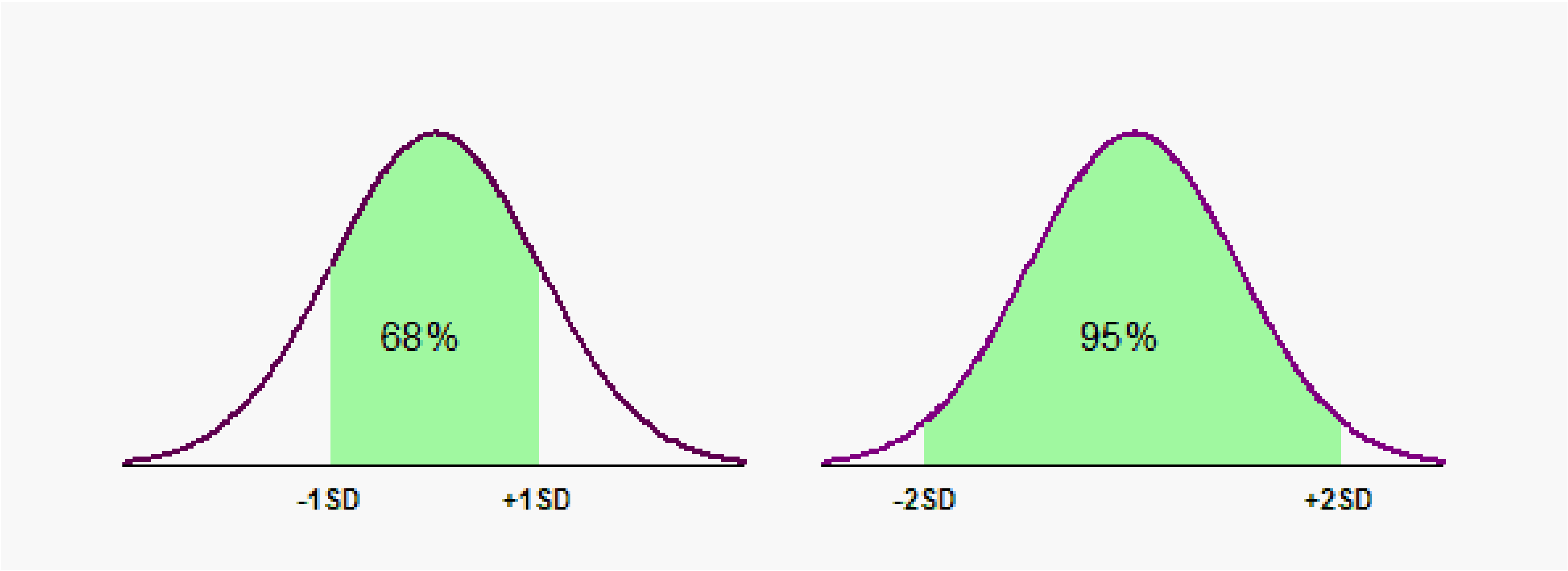



# Standard deviation (SD) and normal distribution



# Standard deviation (SD) and normal distribution



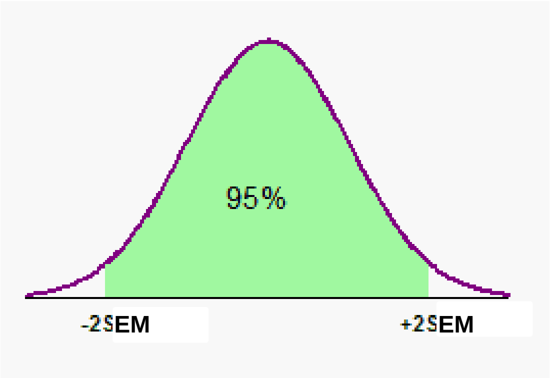

$$\sigma = \sqrt{\frac{\sum(\bar{x} - x)^2}{n - 1}}$$

- Subtract each value from the mean and square it
- Add those up
- Divide by the number of values less 1

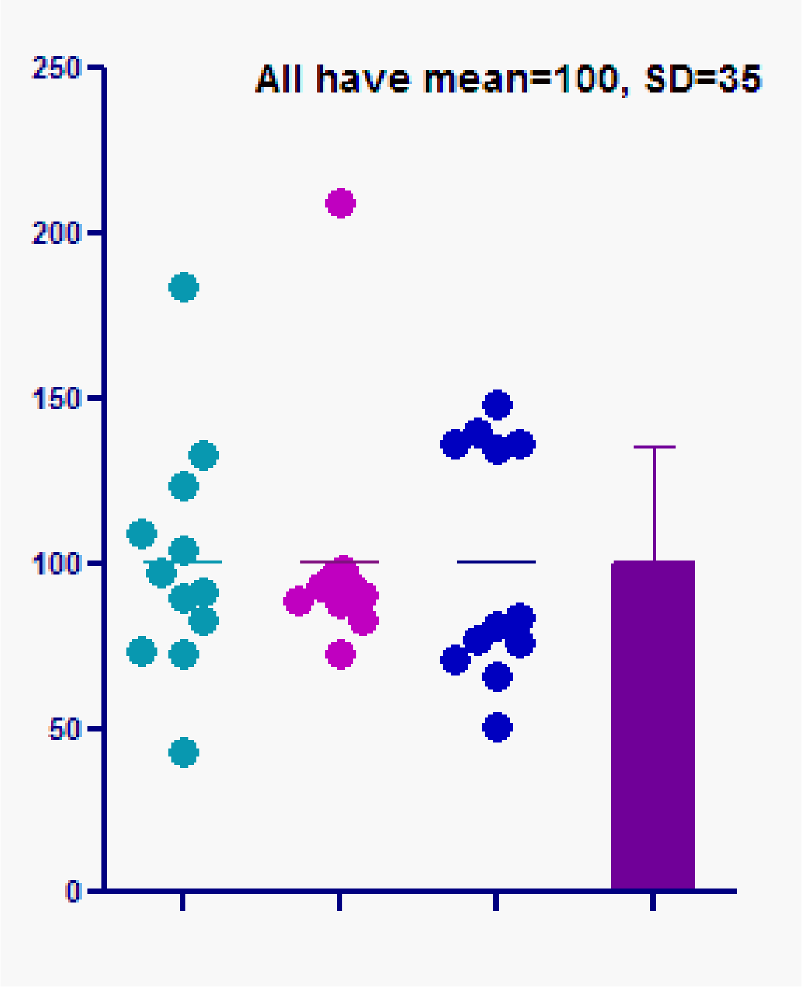
# Standard Error of the Mean (SEM)



- Does this measure variability within your sample?
- No- quantifies the precision of the mean
- If you run the experiment again, the SEM gives an indication of where mean will fall
- Can use to generate confidence intervals (CI)



# Beware assuming Gaussian distribution when interpreting SD



The University Of Sheffield.

SITraN
Sheffield Institute for Translational Neuroscience

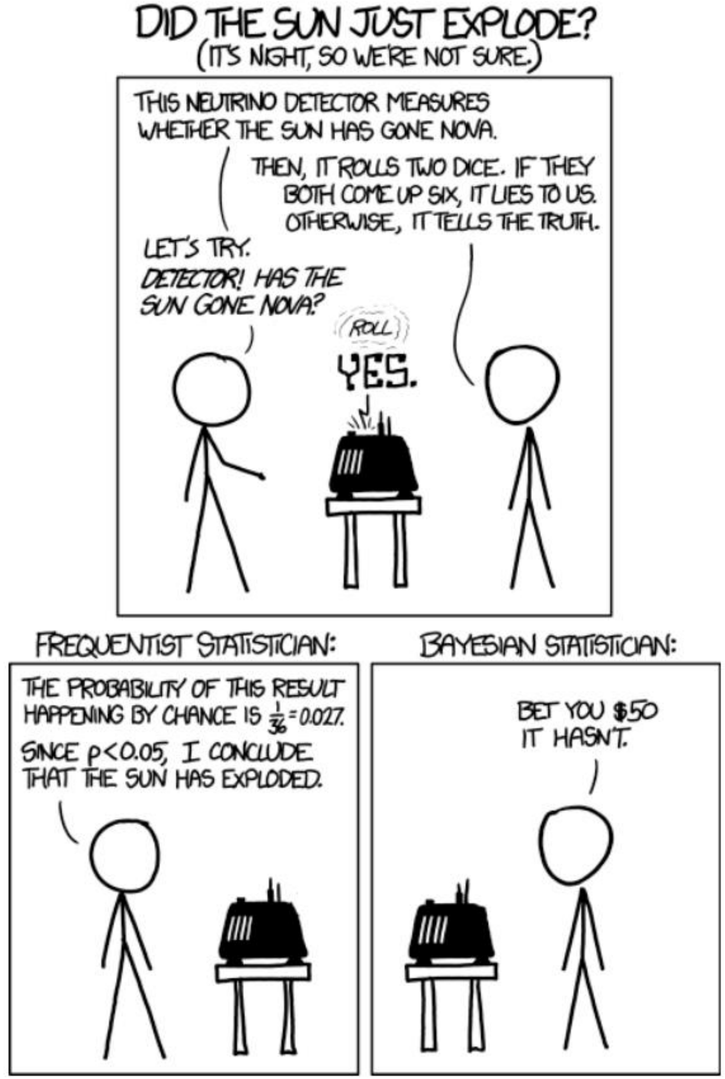
# P values and statistical testing

- When testing if two groups of data differ (e.g Student's T test)
- First hypothesise that the two groups are identical (come from the same normal distribution)
- Called the 'null hypothesis'
- Assuming null hypothesis is true, what is the probability (p-value) of observing that data
- If p is large – null hypothesis is likely to be true
- If p is small– null hypothesis is likely to be false
  - Reject null hypothesis at given significance level (usually $p<0.05$)

# What does p<0.05 mean?

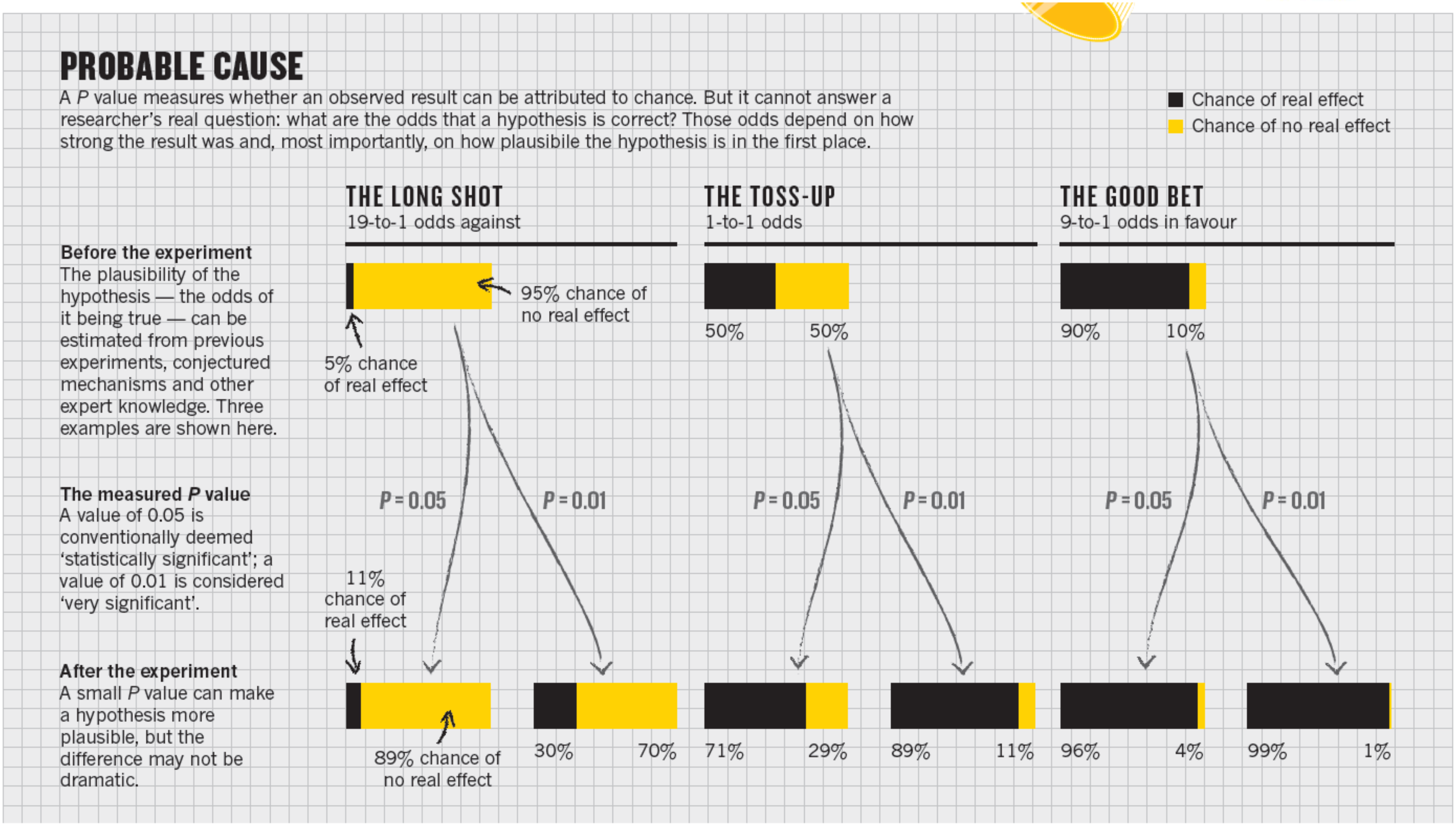
- Arbitrary value
- Means that the observed difference had a 5% chance of being observed if null hypothesis was true
- OR 'there is a 5% chance or less of observing a difference as large as you observed even if the two population means are identical (the null hypothesis is true).

# FREQUENTISTS VS. BAYESIANS



The University Of Sheffield.

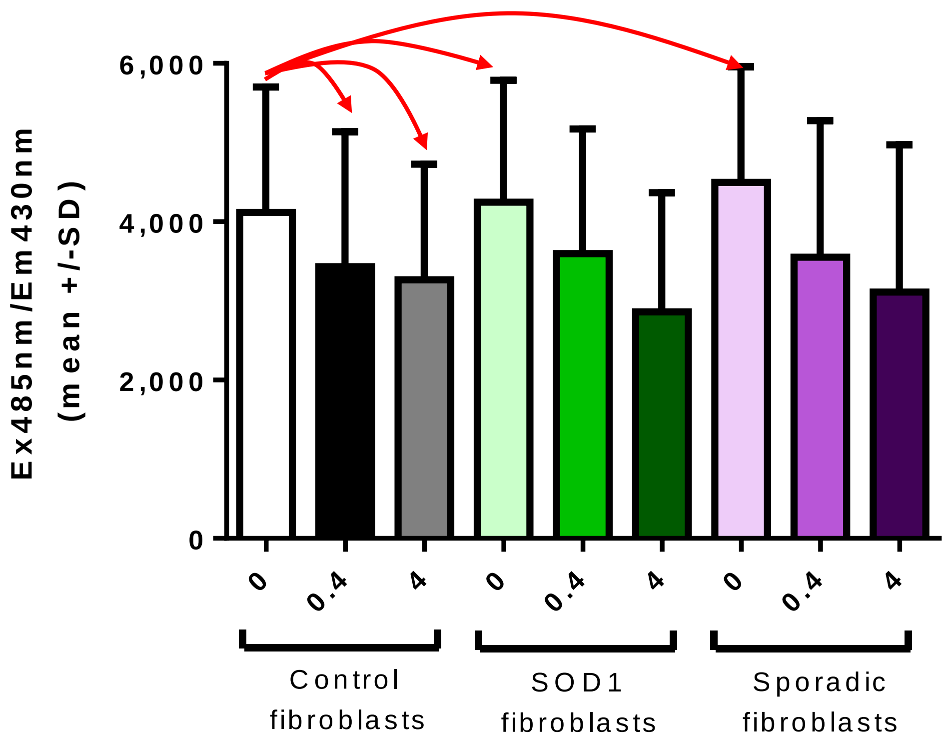
SITraN
Sheffield Institute for Translational Neuroscience




http://www.nature.com/news/scientific-method-statistical-errors-1.14700

# Beware multiple testing

- Student's independent T Test is a parametric test for comparing two independent samples
- What if you had more than two samples?




18 comparisons

18 * 0.05 = 0.9

# Multiple comparisons

- Bonferroni correction. Simply divide threshold p-value by number of comparisons being made

| Age (days) | 35 | 42 | 50 | 56 | 63 | 70 | 77 | 84 | 91 | 98 | 105 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Intensity - fore paws | 0.034 | NS | 0.0207 | 0.0083 | NS | NS | NS | NS | 0.0175 | 0.0323 | NS |
| - hind paws | NS | NS | NS | NS | 0.0202 | 0.0453 | 0.0317 | 0.0043 | < **0.0001** | < **0.0001** | < **0.0001** |
| Print Width - fore paws | NS | 0.0078 | 0.0172 | NS | NS | NS | NS | 0.0217 | 0.0088 | 0.0016 | 0.0002 |
| - hind paws | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | 0.0185 |
| Print Length - fore paws | NS | 0.0011 | < **0.0001** | 0.0005 | NS | NS | NS | NS | 0.0492 | 0.0244 | 0.0109 |
| - hind paws | NS | 0.0408 | NS | NS | NS | NS | NS | NS | NS | NS | NS |
| Print Area - fore paws | NS | NS | 0.0213 | 0.0014 | NS | NS | NS | 0.0326 | 0.0399 | 0.0034 | **0.0001** |
| - hind paws | NS | NS | NS | 0.0169 | NS | 0.002 | 0.0113 | NS | 0.0463 | NS | NS |
| Stand time - forelimb | NS | NS | < **0.0001** | NS | < **0.0001** | NS | < **0.0001** | 0.001 | **0.0001** | 0.0009 | < **0.0001** |
| - hindlimb | NS | NS | 0.0074 | NS | < **0.0001** | NS | 0.0016 | NS | **0.0001** | 0.0021 | **0.0001** |
| Paw angle - fore paws | NS | NS | NS | NS | NS | NS | NS | 0.0477 | NS | NS | NS |
| - hind paws | NS | NS | NS | NS | NS | NS | NS | 0.0465 | NS | NS | NS |
| Swing - forelimb | NS | NS | NS | NS | 0.0314 | NS | NS | NS | 0.0278 | 0.0097 | 0.002 |
| - hindlimb | NS | NS | NS | NS | NS | NS | 0.0061 | NS | NS | 0.05 | NS |
| Stride length - front | NS | NS | NS | NS | 0.039 | NS | NS | 0.0036 | 0.0045 | 0.0002 | < **0.0001** |
| - back | NS | NS | NS | NS | 0.0339 | NS | NS | 0.0005 | 0.0003 | < **0.0001** | < **0.0001** |
| Duty cycle - forelimbs | NS | NS | < **0.0001** | 0.0195 | 0.0073 | NS | NS | **0.0001** | < **0.0001** | < **0.0001** | < **0.0001** |
| - hindlimbs | 0.0415 | NS | NS | NS | 0.0181 | NS | NS | 0.0053 | < **0.0001** | 0.0007 | **0.0001** |
| Max contact at % - fore | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | 0.0088 |
| - hind | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | 0.0055 |
| Swing speed - fore limb | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS |
| - hind limb | NS | NS | 0.0228 | NS | NS | NS | NS | 0.0193 | NS | NS | NS |
| Stand index - fore limb | NS | NS | 0.0011 | NS | 0.0044 | NS | NS | 0.0117 | 0.0202 | 0.0416 | 0.0193 |
| - hind limb | NS | NS | 0.0053 | NS | NS | NS | NS | NS | NS | NS | 0.0052 |
| Duration | NS | NS | NS | NS | NS | NS | 0.0322 | NS | NS | 0.0412 | 0.037 |
| Step pattern # | NS | 0.0173 | NS | NS | NS | NS | NS | 0.0248 | 0.0034 | NS | 0.0051 |
| Step pattern Ca | 0.0414 | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS |
| Step pattern Cb | NS | NS | NS | 0.0372 | NS | NS | NS | NS | 0.0041 | 0.0039 | 0.0088 |
| Step pattern Aa | NS | NS | NS | NS | NS | NS | 0.0211 | NS | 0.0019 | 0.018 | **0.0002** |
| Step pattern Ab | NS | NS | NS | NS | NS | NS | 0.0353 | NS | **0.0001** | 0.0005 | < **0.0001** |

42 different measures at 11 different timepoints (462 comparisons) P-value cut-off <0.00011

# ANOVA

- Analysis of variance
- Can cope with multiple comparisons
- Compares the variances within groups to the variances between them

# ANOVA example

| | Group1 | Group2 |
|---|---|---|
| Replicate 1 | 2 | 6 |
| Replicate 2 | 3 | 7 |
| Replicate 3 | 1 | 5 |
| Mean | | |
| Sum of squared differences from mean | | |
| Overall mean | | |
| Overall sum of squares | | |

'Sum of squared differences from the mean' is an estimate of the variance (used to calculate SD earlier)

$$\Sigma (\bar{x}-x)^2$$

# ANOVA example

| | Group1 | Group2 |
|---|---|---|
| Replicate 1 | 2 | 6 |
| Replicate 2 | 3 | 7 |
| Replicate 3 | 1 | 5 |
| Mean | 2 | 6 |
| Sum of squared differences from mean | 2 | 2 |
| Overall mean | 4 | |
| Overall sum of squares | 28 | |

Power of ANOVA is that it can take account of multiple variables and calculate interaction effects between these variables and so is statistically more powerful than a T-test

# What about non-parametric tests?

- Mann-Whitney U-Test
- Doesn't assume data come from any distribution
- Works on *ranking* the data
- The magnitude of the data values is therefore irrelevant
- Use for ordinal data or if sample sizes small/not normally distributed

The University Of Sheffield.

# Mann-Whitney U test

SITraN Sheffield Institute for Translational Neuroscience

| | Width of leaf / cm | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Sunlight** | 6.0 | 4.8 | 5.1 | 5.5 | 4.1 | 5.3 | 4.5 | 5.1 |
| **Shade** | 6.5 | 5.5 | 6.3 | 7.2 | 6.8 | 5.5 | 5.9 | 5.5 |

# Put data in rank order then sum up the ranks

| Sunlight | Rank | Rank | Shade |
|---|---|---|---|
| 4.1 | 1 | | |
| 4.5 | 2 | | |
| 4.8 | 3 | | |
| 5.1 | 4.5 | | |
| 5.1 | 4.5 | | |
| 5.3 | 6 | | |
| 5.5 | 8.5 | | |
| | | 8.5 | 5.5 |
| | | 8.5 | 5.5 |
| | | 8.5 | 5.5 |
| | | 11 | 5.9 |
| 6.0 | 12 | | |
| | | 13 | 6.3 |
| | | 14 | 6.5 |
| | | 15 | 6.8 |
| | | 16 | 7.2 |
| $R_1$ = | **41.5** | **94.5** | = $R_2$ |

# Mann-Whitney U test

- Calculate U value for each group (normalises data depending on sample size)
- U value can then be read off statistical tables
- Note that these types of tests are less statistically powerful than parametric tests
- For more than two groups use the Kruskal-Wallis test

# Summary

- Plan your experiment before you start
  - Choose statistical methodology
- Understand the basis for the chosen test
- Be confident you can defend your choice