Presentations, Scientific Method, Statistics and Graphs

Katrina Attwood, based on slides by Anna Bramwell-Dicks and Susan Stepney

Outline

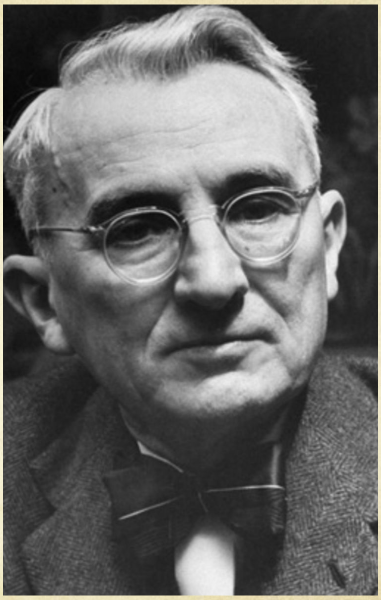
- Presentations
- Scientific method
- Descriptive Statistics
- Inferential Statistics
- Graphs

Presentations

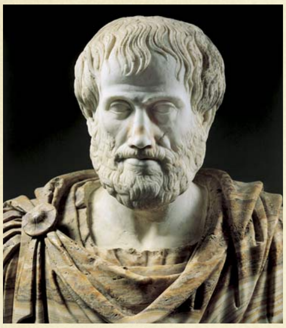
The Rule of Three

Tell the audience what you're going to tell them, then tell them, then tell them what you just told them.

Dale Carnegie (1888-1955), Author of *How to Win Friends and Influence People*



Some More Sage Advice



Advice distilled from Aristotle, *The Art of Rhetoric* (4th century BCE); Roman portrait bust of Aristotle from the British Museum

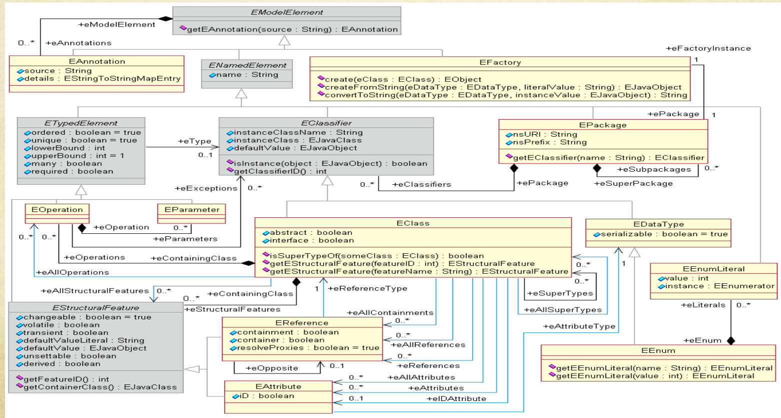
- **Ethos** – your personal credibility and the impression you make. Why should people listen to you?
- **Pathos** – help the audience 'feel' your presentation. Control of your information and involvement in the subject
- **Logos** – the way you actually present your information. Are your Powerpoint slides engaging and interesting, legible, cluttered?

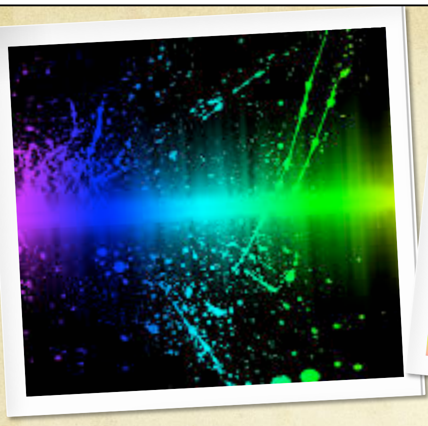
What not to do …

- Don't try to say too much
- Don't speak too fast
- Don't worry about being nervous
  - We all are – especially the person giving feedback!
- Don't worry about pausing to remember what you were trying to say
- Don't just read the slides

More things not to do

- Don't go over the time limit
- Don't put complex images on the screen





Please don't …

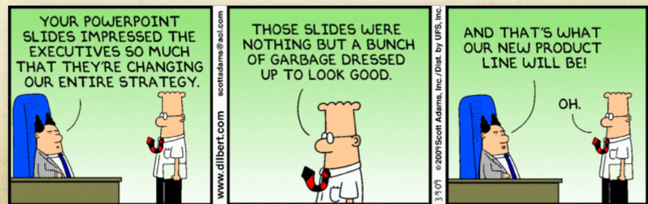
Use distracting backgrounds

What to do

- Be yourself, and be authoritative
- Look at the audience
- Emphasise your "take-home message"
- Speak loudly, slowly and calmly
- Make jokes
- Practice, practice, practice...
- Include references slide

Title: at least 36 point (this is 40)

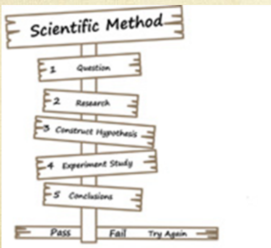
- Use a clear, unfussy font
- Text: minimum of 24-point
  - This is 24-point (you will have seen from earlier slides that smaller is more difficult to read...)
- Provide handouts for any complex diagrams/images/ datasets/graphs



Project Presentations

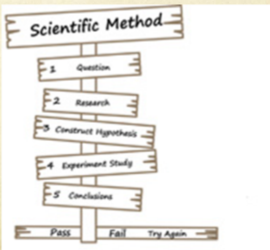
- You will be giving *two presentations* as part of your project
  - One next summer, which is assessed (more details nearer the time)
  - One next term, which is a practice one, for which you will be given constructive feedback
    - Various sessions, in the week beginning 23rd January 2017
    - In groups of 8 or so
    - Please check the VLE for details of your group, and the time/place of the session
    - Each required to attend for one hour
- What your project is about, and why it is important
  - 7-minute presentation, using Powerpoint
    - No more than 6 slides recommended
  - Title slide with your name and the title of your project
  - Outline of your problem/research question
  - Refer to the literature

Scientific Method



Have a clear question

- Clear statement of what you are trying to build or discover
- Posed in such a way that you can tell if you have answered it



Experimental Design

- Need for a controlled experiment
- Design of the *controls* is difficult:
  - Placebo effect
  - Observer effect

Null Hypothesis (H_0)

- General statement or default position that there is *no relationship* between two measured phenomena, or *no difference* between groups
- H_0 is generally believed to be true until data proves otherwise
  - Science is about trying to disprove or to confirm the null hypothesis – i.e. demonstrating there is a relationship, or that there is no relationship
    - And possibly using this inference to build a new hypothesis
  - See Darrell Huff, *How to Lie with Statistics* (Penguin 1991)

Significance Testing Approach

- Ronald Fisher, *Statistical Methods for Research Workers* (1925)
- Two sets of data are compared
- H_0 = there is no relationship between the datasets
- A threshold value for a level of significance in the relationship shown by the data is pre-determined (various tests)
- If there is a *statistically significant* similarity between the datasets, the null hypothesis is rejected
- Think of a law case, where H_0 = “not guilty”

Hypothesis Testing Approach

- Jerzy Neyman and Egon Pearson, 'On the Problem of the most Efficient Tests of Statistical Hypotheses', *Philosophical Transactions of the Royal Society* vol 231 (1933)
- Null hypothesis is contrasted with an alternative hypothesis
- Data collected, with error rates
- Hypotheses compared as likely explanations

Example (NB: please do not try this, or tell the ethics committee!)

- H_0 = red-haired people feel no more pain than non-red-heads
- H_1 = red-haired people are more susceptible to pain than non-red-heads
- Equipment: 25 red-headed volunteers who have given informed consent, 25 non-red-headed volunteers who have given informed consent, chair, table, several candles, cigarette lighter, ruler
- Method: sit person on chair behind table, put candle on table and light it, tell person to hold hand 15cm above candle for 3 seconds, ask them how much pain they are in on scale of 1-10
- Results: two data distributions, one for red-haired people, one for non-red-heads

Data Analysis

- Choose a suitable statistic (more about this later)
  - Average amount of pain recorded by each group
  - Number of people in each group scoring pain over a certain value
  - Most commonly-reported pain score in each group
- Compare the results
- Determine whether any variance is statistically significant (more about this later)
- http://news.bbc.co.uk/1/hi/magazine/8195177.stm
- **http://sciencenordic.com/redheads-feel-different-kind-pain**

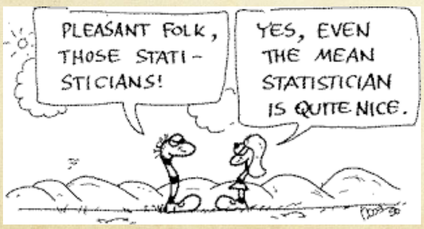
Compare Like with Like

- "My algorithm is better than X's algorithm"
  - My highly-optimised algorithm is better than X's prototype algorithm
  - My algorithm worked better than X's did the first time I ran it
  - My algorithm worked better than X's did on this artificial problem
    - This particular (tailored) data set
    - This processor
  - My algorithm worked better on this problem than X's algorithm worked on that problem

Results

- Descriptive Statistics
- Inferential Statistics

Descriptive Statistics



Descriptive Statistics: Purpose

- To describe, quantitatively, the main features of a collection of data
- Finding a number which 'best characterises' the data
- Two types:
  - Measures of central tendency – mean, median, mode
  - Measures of variability – standard deviation, quartiles, min-max values

Example

- Sample mean
  - {-30, 1, 2, 3, 4} = -4
  - {0, 1, 2, 3, 4) = 2

$$m = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Median
  - {-30, 1, 2, 3, 4} = 2
  - {0, 1, 2, 3, 4} = 2
- Sample standard deviation
  - {-30, 1, 2, 3, 4} = 14.6
  - {0, 1, 2, 3, 4} = 1.6

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x-\bar{x})^2}$$

- Lower quartiles:
  - {-30, 1, 2, 3, 4} = 1
  - {0, 1, 2, 3, 4} = 1
- Upper quartiles:
  - {-30, 1, 2, 3, 4} = 3
  - {0, 1, 2, 3, 4} = 3

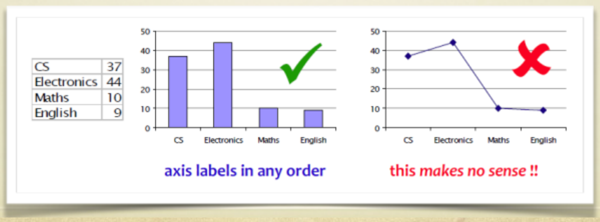
Data Types

- Choice of appropriate descriptive statistic depends on data type
- What are they?

- Nominal/Categorical
- Ordinal/Ranked
- Numerical
  - Interval
  - Ratio

Nominal/Categorical Data

- Data grouped in **categories**
  - on/off, male/female, butcher/baker/candlestickmaker
- Typical Operations: equality of category
- Descriptive Statistic: mode



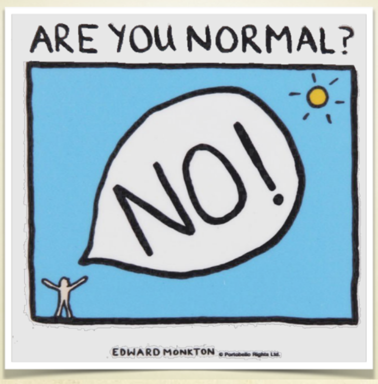
Ordinal/Ranked Data

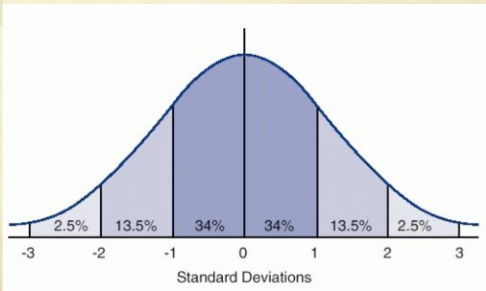
- Data can be ranked: a < b < c
  - Very poor < poor < satisfactory < good < very good < excellent
- Typical Operations: comparisons of ranks
- Descriptive statistic: median, quartiles

Numerical data

- Interval data
  - Numerical data, with an arbitrary zero point
    - E.g. temperature in degrees Celsius; dates
  - Typical Operations:
    - Comparisons of values
    - Arithmetic: a + b, a – b
- Ratio data
  - Numerical data, with an absolute zero point
    - E.g. temperature in degrees Kelvin; length; mass
  - Typical Operations:
    - Comparisons of values
    - Arithmetic: a ± b, a / b

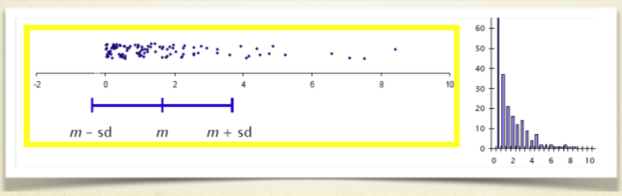
Descriptive Statistic: Need to ask …





Mean / Standard Deviation

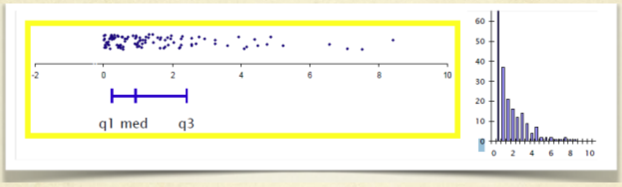
- Is it normal?



- Mean = 1.69; standard deviation = 2.14
- Majority of data < m; no recognition of skew
- m – s.d. is outside the possible range of data!

Median / Quartiles

- Is it normal?

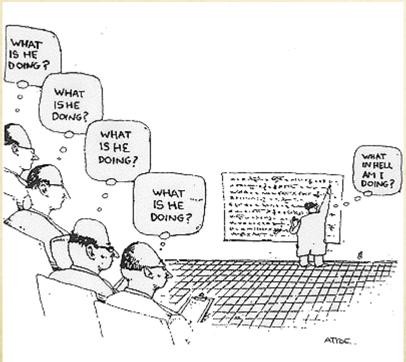


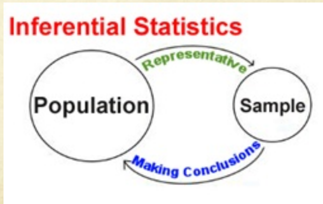
- Median = 0.98; 1st quartile = 0.34; 2nd quartile = 2.32
- Half of the data is < median (by definition)
- Quartiles are inside data range

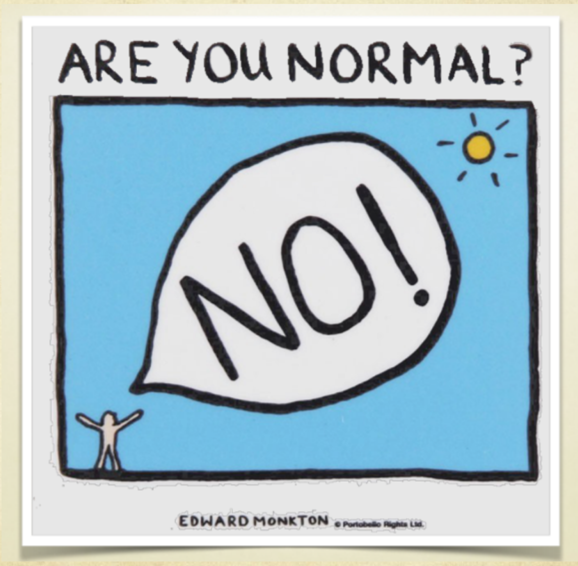
Inferential Statistics

Inferential Statistics: Purpose

- Use data to make inferences about the population
- Null Hypothesis Significance Testing







Why assume normality?

- Normal distributions are relatively common
  - But so are: long-tailed distributions, power laws, skewed data
- Statistics was an area of study well before computers
  - Normalisation useful, as it made volume of data tractable
- Now we have computers, we can analyse actual distributions rather than approximations
- Non-parametric tests should be your first choice
  - Otherwise you have to demonstrate that your data is normal

Statistical Testing (1)

- Make sure you know what your test is comparing!
  - Mann-Whitney U-Test compares medians of different samples
    - H_0: Samples X and Y have same medians
    - If they don't, then the distributions are different
  - Kolmogorov-Smirnoff Test (KS test) compares distributions of different samples
    - H_0: Samples X and Y have the same distributions

Statistical Testing (2)

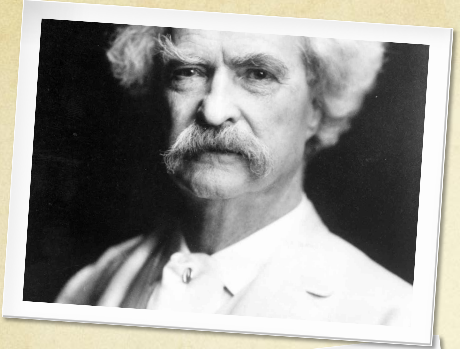
- Other tests are available, suited to other situations and purposes
  - Particularly paired tests, e.g. sets of 'before' and 'after' data
- NB: Statistical tests not available for all situations
- Very important to check that there is a suitable test before you run your experiment
  - If there is not one, then change the experimental design

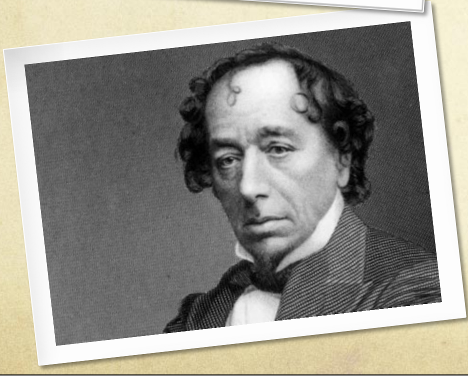
Statistical Argument

- We have talked a lot about the argument of your project
- Use inferential statistics as *evidence* to support your argument

"[data analysis] should make an interesting claim; it should tell a story that an informed audience will care about, and it should do so by intelligent interpretation of appropriate evidence from empirical measurement or observations."

Robert Abelson





"Figures often beguile me, particularly when I have the arranging of them myself; in which case the remark attributed to Disraeli would often apply with justice and force: "There are three kinds of lies: lies, damned lies and statistics.""

Mark Twain, *Chapters from my Autobiography* (1906)

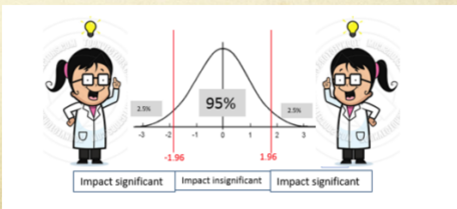
NB: There is no trace of this remark in Disraeli's published works, and the earliest occurrences were after his death…

…maybe it is a lie too…

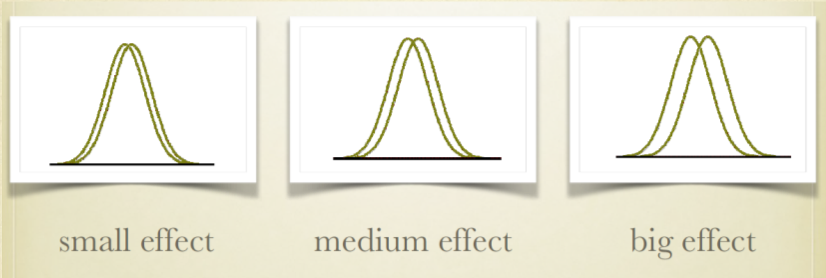
Photo of Mark Twain in 1907, The Mark Twain House and Museum; Benjamin Disraeli c. 1878, Wikipedia

"Significant" ≠ "Important"

- The old algorithm has a success rate of 52.38%
- My algorithm has a success rate of 52.41%
- But it has improved significance at the 99.9% confidence level
- Effect size is small
- Result is not significant



Effect Size



Observed effect size

- Many different ways to calculate
- Means/distances between means
- You may also come across:
  - Cohen's d
  - Eta squared η^2
  - Partial eta squared η^2_p
  - Omega squared ω^2
  - etc…

Inferential Statistics

- You don't need to calculate values yourself!
- Use:
  - R; SPSS; Matlab …
- Lots of good, helpful books out there
- And lots of helpful academics in the Department who have a wealth of knowledge and experience and are eager to share

Results

- You've done a well-designed, controlled experiment, comparing like with like and have evaluated it properly
- Outcome 1: You get statistically significant, repeatable results!
  - My algorithm really is better than X's is!
  - You hope you are going to get a good project mark …

Results

- You've done a well-designed, controlled experiment, comparing like with like and have evaluated it properly
- Outcome 2: You *don't* get statistically significant, repeatable results!
  - I have no evidence that my algorithm is better than X's is!
  - You fear you are going to get a terrible project mark…

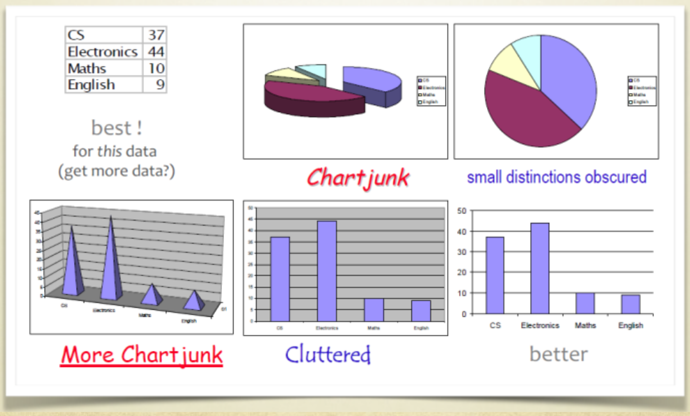
Results

- The truth is that, whatever the results, they are only interesting if you can say why
  - Negative findings are as interesting as positive ones
- Leads to new questions and hypotheses
- You become a giant, for the next generation of dwarves…

A pile of £2 coins (inscription after Newton); Eduardo Paolozzi, *Shoulders of Giants* (1995), British Library; US Library of Congress, MS Rosenwald 4, bl. 5r (a medieval collection of allegorical and medical drawings)

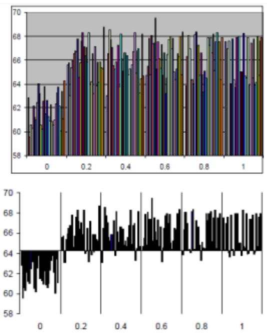
Graphs

Graphing your data



Graph to … expose structure

0.00	0.20	0.40	0.60	0.80	1.00
62.86	65.61	63.06	64.70	63.85	67.54
59.58	65.74	66.51	65.48	66.32	64.31
60.59	63.06	68.55	67.86	68.05	66.96
60.26	65.35	67.73	64.77	66.96	67.60
62.13	65.93	67.03	68.05	64.31	64.05
61.20	66.51	65.55	67.35	64.38	63.65
61.00	66.77	65.22	67.41	64.05	67.79
62.46	66.71	65.87	69.50	68.11	64.38
64.05	67.79	64.83	65.22	68.36	63.72
63.19	64.57	67.22	67.54	64.38	67.92
60.53	65.74	63.79	66.13	67.22	64.64
60.19	67.60	68.17	65.29	64.70	67.98
62.53	68.30	66.13	67.28	66.64	65.03
63.79	67.16	65.09	64.70	64.77	65.03
61.47	66.39	66.64	66.64	63.32	63.79
62.60	67.03	65.09	63.19	65.09	68.24
61.26	65.81	66.39	64.51	64.51	64.64
60.93	68.30	64.77	66.45	64.44	64.51
62.27	63.98	65.22	67.73	67.92	67.92
60.86	64.18	64.64	67.98	67.22	63.92
60.59	66.51	65.61	66.96	67.92	63.98
61.00	63.19	66.71	67.86	66.58	67.41
62.33	67.16	68.30	65.03	68.17	64.51
63.26	63.85	64.90	64.38	65.16	67.79
63.72	65.81	66.77	65.22	67.54	63.92
62.13	64.11	66.96	66.51	68.30	64.11
60.05	65.81	63.26	66.32	64.44	67.98
63.39	65.35	64.44	67.98	67.54	64.70
61.06	65.22	68.17	64.77	64.38	67.60
64.24	68.74	64.64	68.36	67.92	67.92

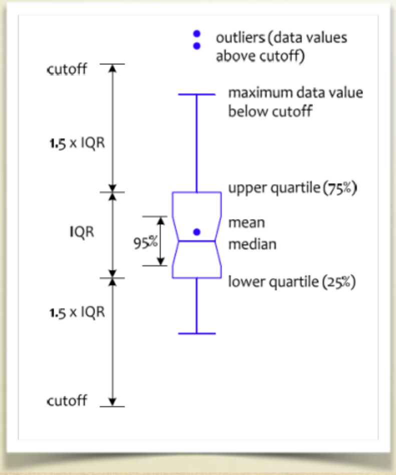


[Clark, Jacob, Stepney. Secret Agents Leave Big Footprints. *GECCO 2003*]

Graph to … see outliers

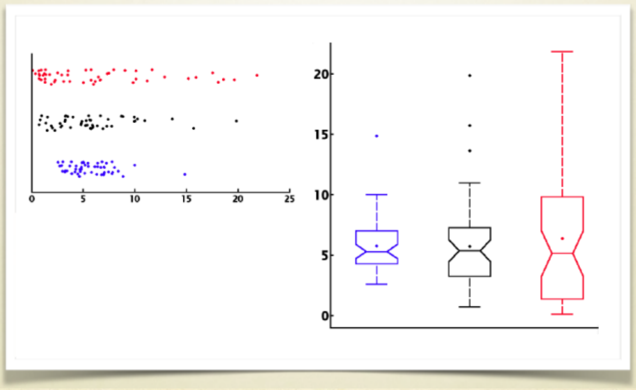
- Don't just discard as an "outlier"
  - Is it just a statistical fluctuation?
  - Is it an error in experimental design or implementation?
  - Is it an interesting unexpected effect?
    - Or the thing you are actually looking for?

Enhanced Box-and-Whisker Plot



- Highlights outliers
- Plot mean (to highlight skew)
- "notches" at median ± 1.58 (IQR / √n)
- If notches on separate bars, do not overlap, ~95% confidence medians are different

Example



3 data distributions – use enhanced box and whisker plot to compare

Summary

- Experimental design is important
- Use descriptive statistics to describe your data
- Use inferential statistics to make statements about data and (hopefully) claims about the wider world
- Use graphs to tell your story
- Make use of tools and available expertise to help you plan, collect, analyse and present your statistics