Experimental Design and Data Analysis

Nilesh B. Patel, PhD
Dept Medical Physiology
University of Nairobi
Kenya

What is science?

- The systemic study of the structure and behavior of the physical world, involving experimentation and measurement and the development of theories to describe the results of these activities.

- Cambridge International Dictionary of English

What is an experiment?

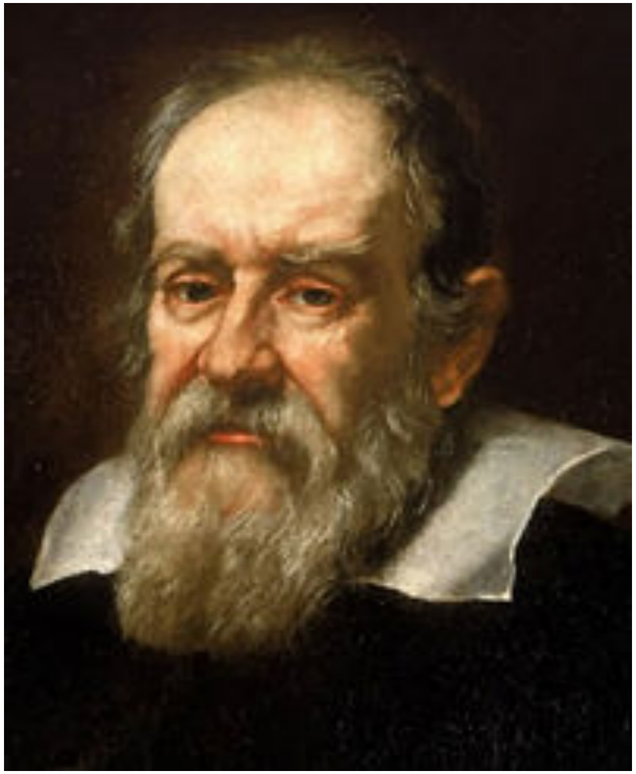
- A test done in order to learn something or to discover whether something works or is true *(plausible).*
- Cambridge International Dictionary of English
- Experiments are done to find out if the hypothesis is false.
- Only falsification is possible with certainty; proving something can never be done with certainty

What can be studied with the methods of science?

1. The capital of Kenya is Nairobi.
2. Holland is twice the size of Kenya.
3. All humans are mortal.
4. All rabbits are grey.
5. All ziwats are blue.
6. Mentally ill people are possessed by an evil spirit.
7. Mentally ill people are possessed by an evil spirit that cannot be detected by any known means.

Adapted from: Circadian Physiology. Roberto Refinetti. CRC Press 2000

Galileo Galilei
(1564-1642)



- Father of modern physics
- Father of modern observational astronomy
- Father of modern science
- Pioneered the use of quantitative experiments whose results could be analyzed with mathematical precision

How many of you accept that the sun goes round the earth?
How many of you accept that the earth goes round the sun?

Types of Experiments

- Experiments are done to test a hypothesis; a prediction.
  - Observational
  - Experimental
  - Quasi-experimental

Experimental Design

- Whether or not a study's findings are useful or not depends crucially on design.
- No matter how ingenious or important an idea for an experiment might be, if the study is badly designed, it's worthless.
- The virtue of the experimental method for doing science is that it is an excellent procedure for determining cause and effect.

Three Aims of Research

- Validity
  - Results actually show what it is that you intent them to show
- Reliable
  - Potentially replicable by yourself or anyone else
- Important
  - Subjective
- Research can possess all the above qualities but still be essentially trivial.
- Conversely, a research cannot be important if its findings are unreliable or invalid.

Score or Measurement

1) A true measure of the thing we want to measure
2) A measure of other things
3) Systemic (non-random) basis: measuring other things inadvertently.
4) Random (non-systemic) error, which should cancel out over large number of observations.

- We want the score or measurement to consist of as much “true score”, as possible and little of the other factors (validity and reliability).

Evidence for the intellectual inferiority of women

- Paul Broca (19th Century) make careful measurement of brain weight and found Caucasian men have larger brains than Caucasian women, who in turn have a larger brain than negros or any other non-Caucasian for that matter.
- Modern brains were supposedly heavier than medieval brains, and French brains were heavier than German brains.
- The brain weights differences were considered to reflect differences in intelligence between these different groups.

Experimental Design

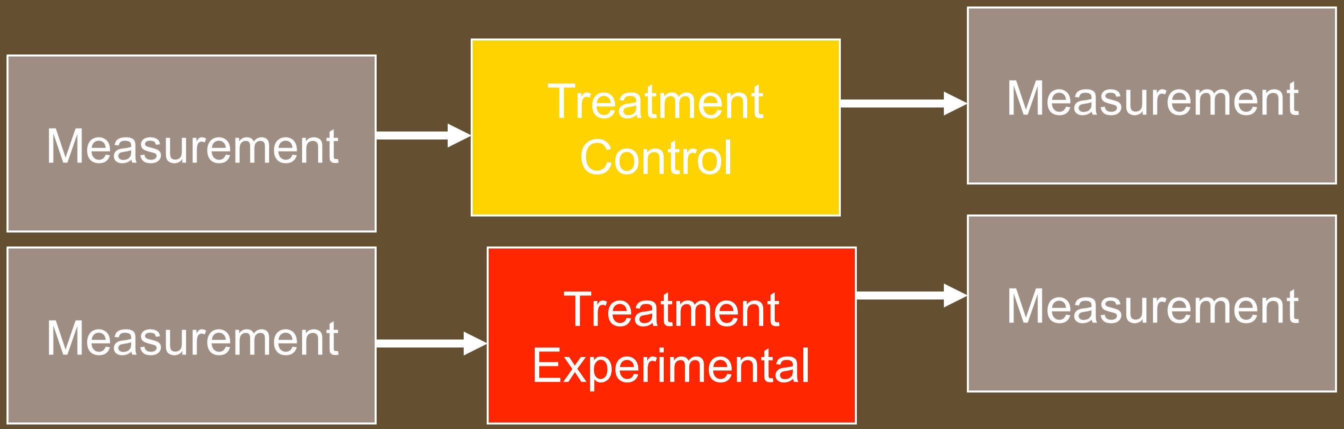
- This is important so that the results can be interpreted and other causes can be excluded from the most likely conclusion.
- The results have to be valid, reliable and generalizable (important)

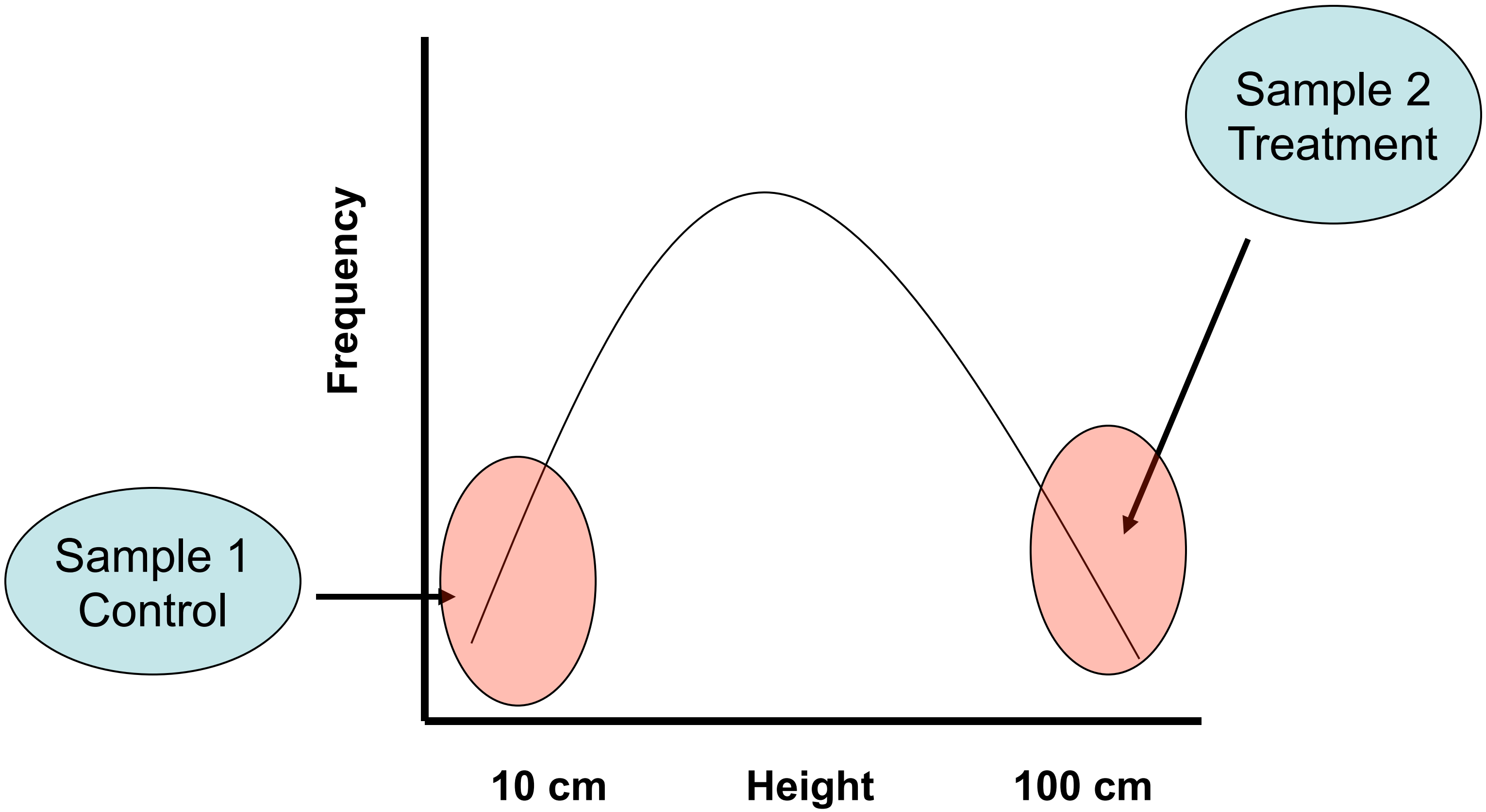
Simplest Experimental Design

There is a difference between pre and post-treatment.
So it is shown that the treatment has an effect!

Importance of Control

Experiment with Control

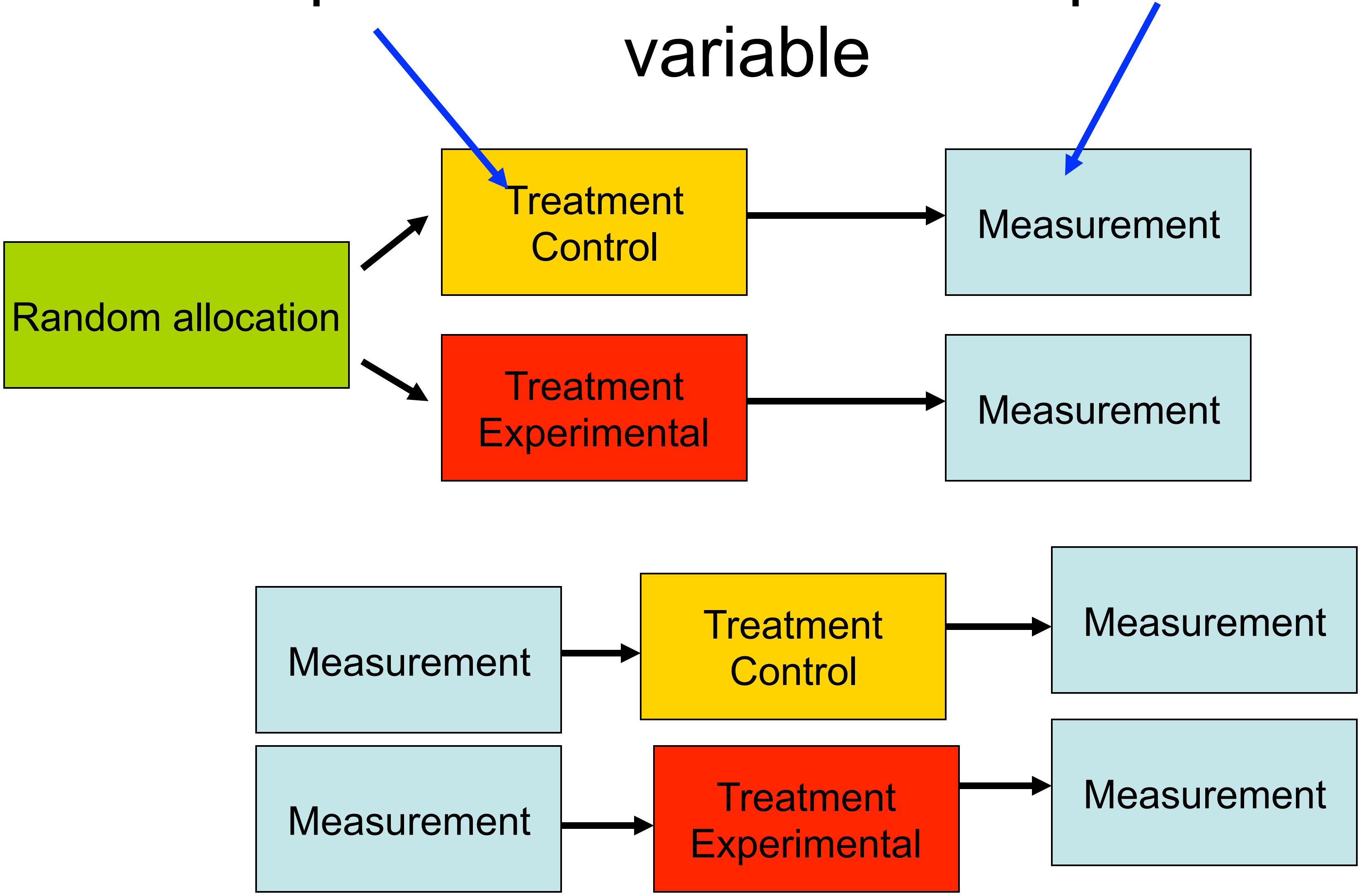




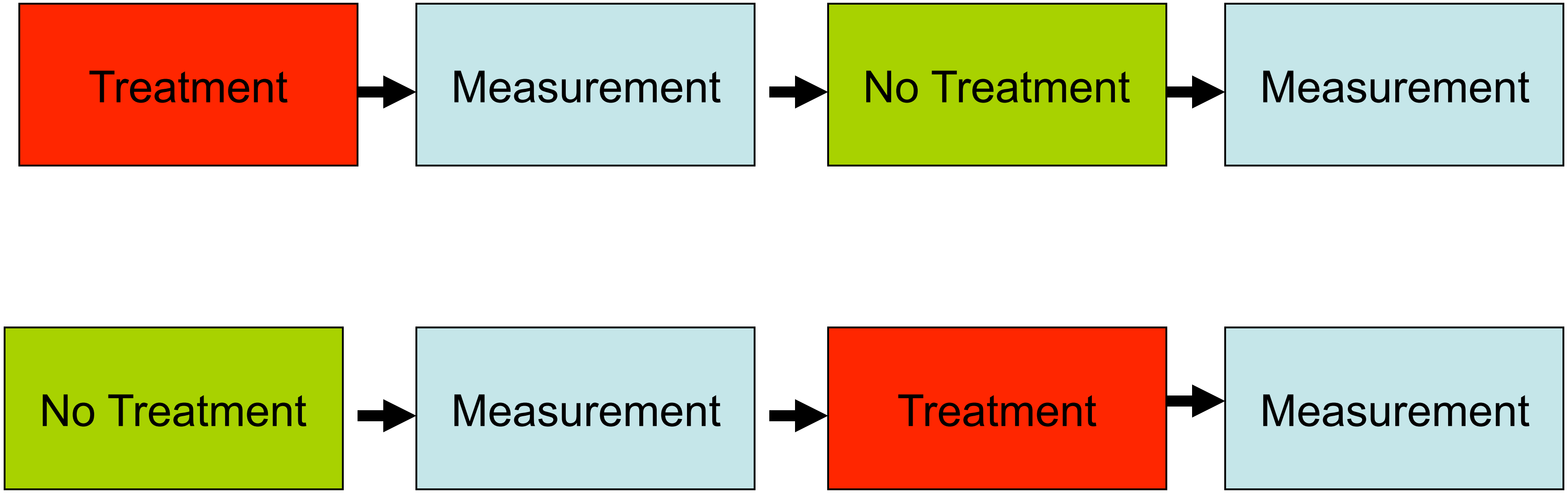
The mean for control was 12 cm and the treatment group was 98 cm so my treatment was significant.

Importance of randomization in sample selection

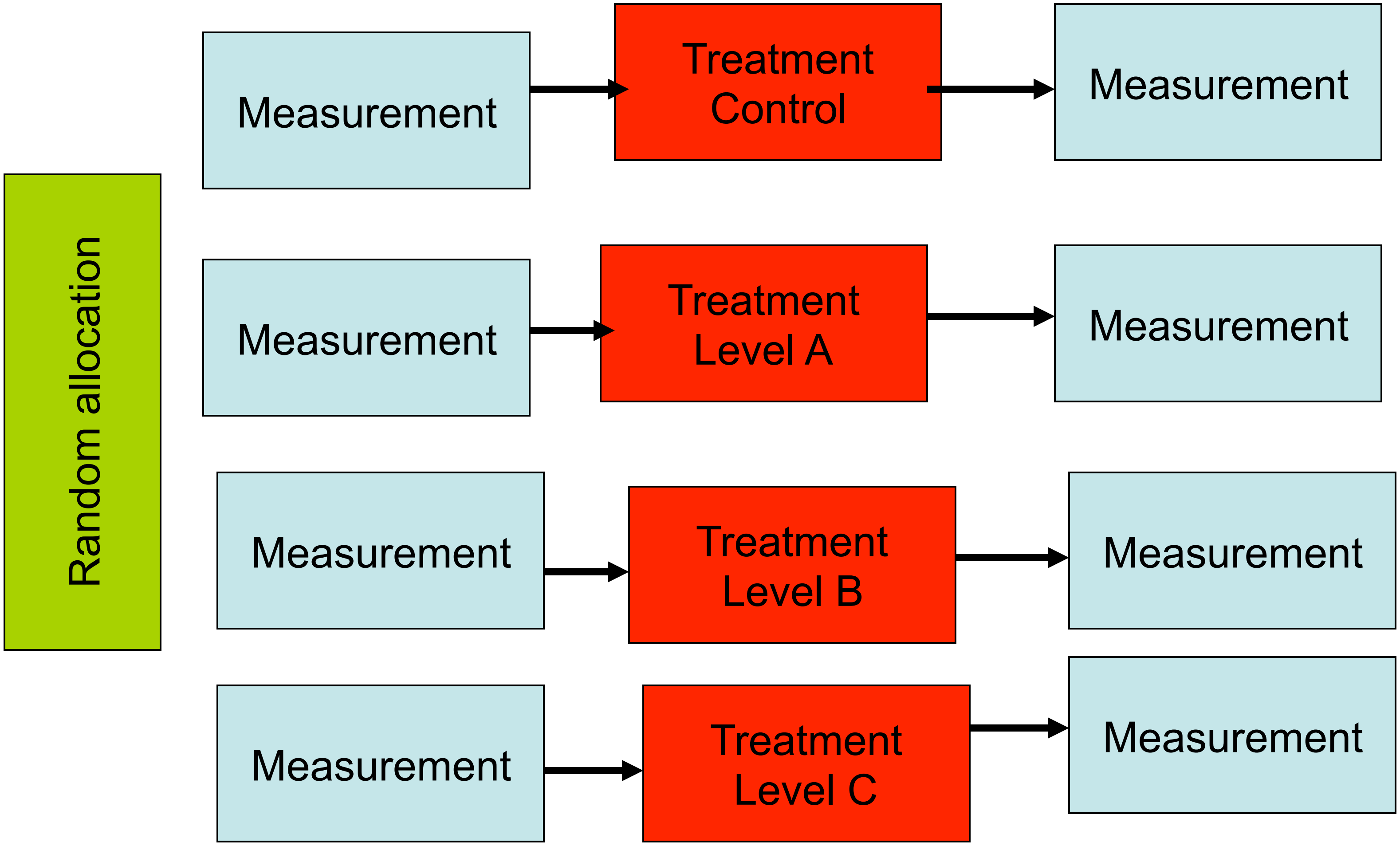
Post-test/control group
Independent variable and dependent variable



Two-Condition Repeated-Measure



Multiple Levels of Independent Variable
Measurement of dependent variable



Multi-Factorial Design

- Two or three independent variables in the same study
- Gender and treatment
- Time and treatment

Measurement

- Types of measurements
  - Nominal
  - Ordinal
  - Interval
  - Ratio
- Is what you measured, a real measure of what you are interested in?
  - It is a measure what you are interested in.
  - Random (non-systematic) errors.
  - Non-random errors (systemic) errors.

Why Inferential Statistics?

- Descriptive statistics does not help to answer research questions.
- Most of modern research is hypothesis driven.
- There are two common hypothesis (predictions) inherent

1. Experimental hypothesis (H_A)

   The experimental manipulation will have an effect.

2. Null hypothesis (H_0)

   The experimental manipulation will have no effect

So what if there is a difference between the means?

- Two samples from the same population will have different means.
- If the samples are taken from the extremes of the distribution, the difference between the means can quite large.
- So large differences suggest that the means are from two different populations, but how can we be certain that it not due to sampling from the extreme of the distribution.
- Application of inferential statistic.

The question

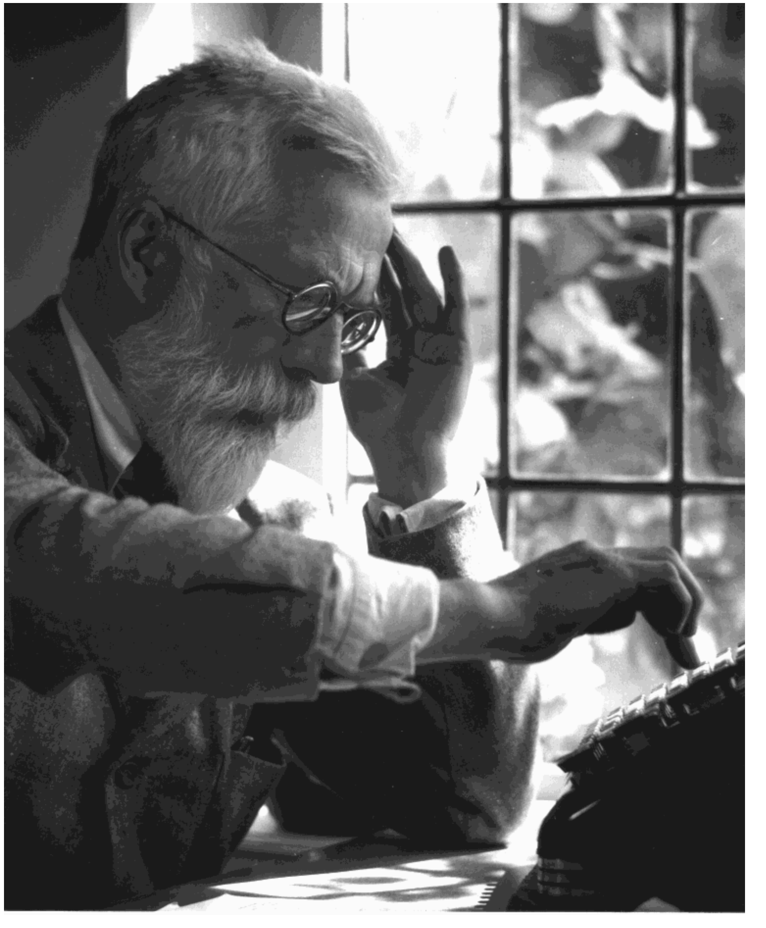
- "what is the probability that the difference between the means is due to chance and not due to my manipulation?"

Carry out tests of significance

- Give the probability of obtaining the difference between the means by chance.
- Choice of test depends on the type of measurement and the experimental design.
- Whichever test you use you will end up with a number – **the test statistic,e.g. t, z, F**.
- What is the probability of obtaining that value of the test statistic for the means and the spread of data that I have?
- On what basis do I or anyone else accept that my manipulation did indeed have an effect?

Ronald Aylmer Fisher
1890-1962

- **"To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of."**



Fisher tea test

- Is the milk added after or before the tea is added?
- The level of significance $p < 0.05$.
- There is a 5% or less probability that the difference between the means is due to chance.
- I am therefore willing to accept with a 95% confidence that the difference I obtained was due to my manipulation and not due to chance.
- My results are significant but are they important?

Commonly used tests of significance

Parametric tests

- Student's *t*-test
  - Independent t-test
  - Dependent *t*-test
- Analysis of Variance (ANOVA)
  - One-Way ANOVA
  - Two-Way ANOVA
  - Repeated measures ANOVA
  - Mixed ANOVA

Non-parametric tests

- Mann-Whitney test
  - = independent t-test
- Wilcoxon Signed-Rank test
  - = dependent t-test
- Kruskal-Wallis test
  - = one-way ANOVA
- Friedman's ANOVA
  - = one-way repeated measures ANOVA

Which test to use?

- Sample size
- Data distribution
- Homogeneity of variance
- Independent measurements or repeated measurements
- Comparison between 2 groups
- Comparison between 3 or more groups
- Type of data

Basic tests

- Mann-Whitney test
  = independent t-test
- Wilcoxon Signed-Rank test
  = dependent t-test
- Kruskal-Wallis test
  = one-way ANOVA
- Friedman's ANOVA
  = one-way repeated measures ANOVA

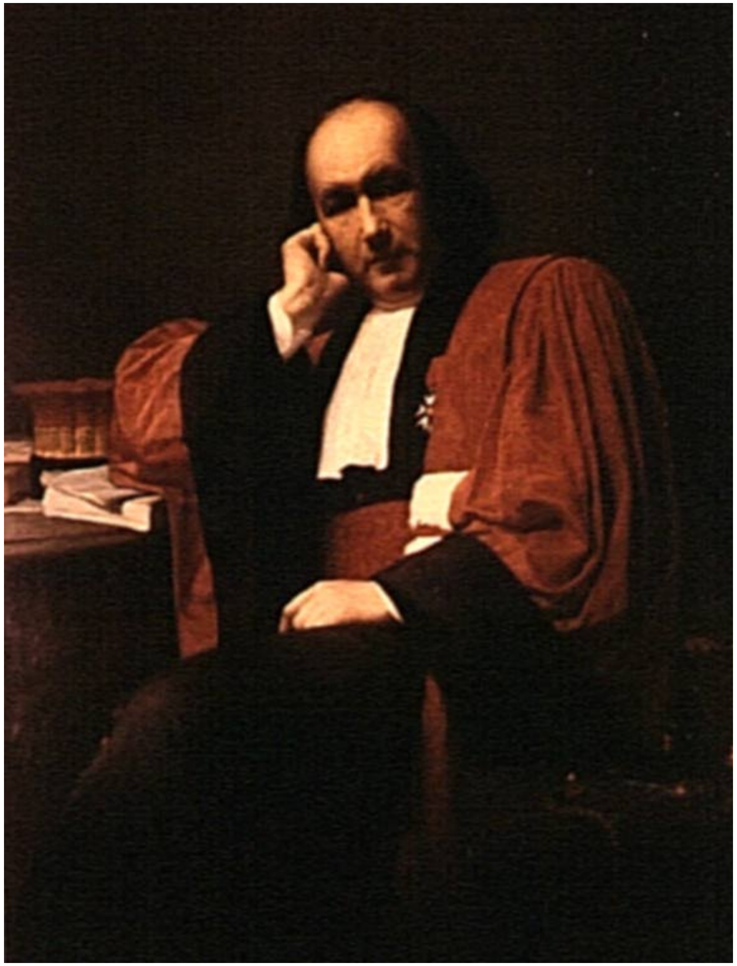
Mann-Whitney Test

- = independent t-test
  - Two conditions and different participants in each condition
- Men (M dn = 27) did not seem to differ from dogs (M dn = 24) in the amount of dog-like behaviour they displayed ($U = 194.5$, *ns*)
- Graphs are usually box and whisker plots.
- Use of the median
  - Non-parametric tests are testing the difference in ranks; not the difference in means.
  - Means are biased by outliers; whereas ranks are not.

Kruskal-Wallis Test

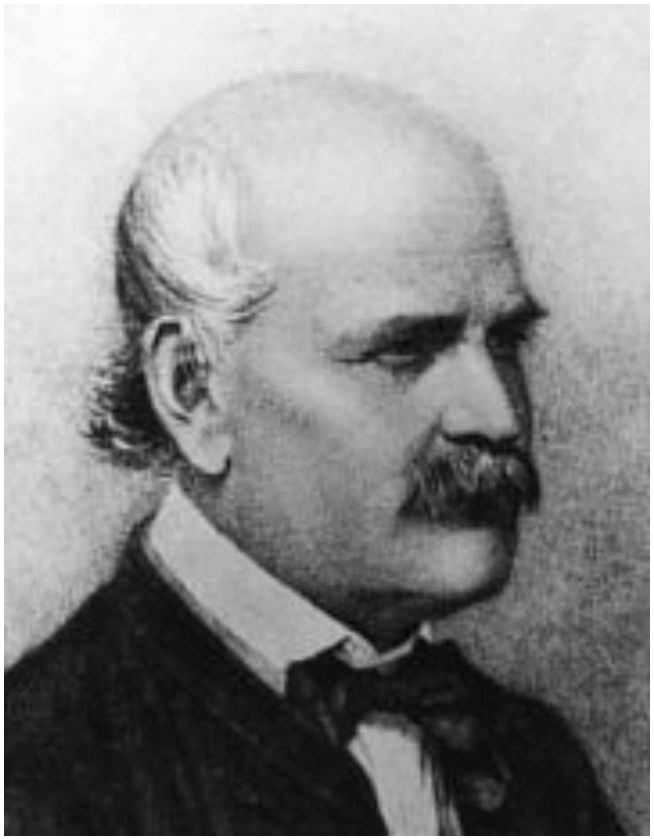
- = one-way ANOVA
  - More than 2 conditions and different participants have been used in each condition.
- The test statistic is H and has chi-squared distribution
- Children's fear beliefs about clowns were significantly affected by the format of information given to them ($H(3) = 17.06$, $p < 0.01$). Mann-Whitney test were used to follow-up this finding. A Bonferroni correction was applied and so all effects are reported at a 0.167 level of significance. It appears that fear beliefs were significantly higher after the adverts compared to the control, $U = 37.50$, $r = -.60$.

Claude Bernard (1813-1878)



- Those whose minds are bound and cramped. They make poor observations, because they choose among the results of their experiments, observation, and reading only what suits their object, neglecting whatever is unrelated to it and carefully setting aside everything which might tend toward the idea they wish to combat.

Ignaz Semmelweis (1818-1865)



- *Savior of mothers*
- 10-35% mortality
- Death of colleague Jacob Kolletschka (1847)
- Introduced lime washing of hands
- 12.24% to 2.38%.
- The Etiology, Concept and Prophylaxis of Childhood Fever (1861).

Semmelweis Reflex

- Dismissing or rejecting out of hand information, automatically, without thought, inspection, or experiment.

Correlations

- Relationship between two variables.
- Smoking and lung cancer.
- NOT A CAUSE AND EFFECT

Regression and correlation

- Regression coefficient
- Product-moment correlation coefficient (Pearson's coefficient) - r.
- Spearman's coefficient of rank correlation – r'.
- A strong correlation merely indicates a statistical link. Not to be confused with cause and effect.
- Weekly deaths = a Temperature + b Relative humidity + c Index of air pollution

Statistical Significance

- What does it mean?

Proof

- A statistician's view of "proof" is rather different to that of the mathematician's. A statistical proof is closer to the legal notion of a conclusion being proved "beyond reasonable doubt".
- All statistical judgments are given in association with a measure of probability of how confident we are that the pattern or difference is a real one and not just a fluke result due to chance.

How to test a hypothesis?

- The hypothesis must be set up in a quantitative manner.
  - Discrete or categorical variables
  - Continuous variables.
- The hypothesis tested statistically is the null hypothesis, H_o. This is tested against the alternate hypothesis H_A.
- We test the null hypothesis because it reduces the number of comparisons to between H_o and H_A to one rather than many.

- It is impossible to prove the null hypothesis.
- All statistical tests can do is reject the null hypothesis or fail to reject it.
- We do not prove hypothesis by gathering affirmative or supportive evidence because no matter how many times the experiment showed a difference close to zero, we cannot be sure that the next time we did the experiment, we would not find a huge difference that was nowhere near zero.

Reasoning

- I will assume the hypothesis that there is no difference is true.
- I will collect the data and observe the difference between the two groups.
- If the null hypothesis is true, how likely is it that by chance I would get results such as these.
- If it is not likely that these results could arise by chance under the assumption the null hypothesis is true, then I will conclude it is false, and I will accept the alternate hypothesis.

Types of error

- We can never be 100% certain that we are right in either accepting or rejecting the null hypothesis. We run the risk of making one of two kinds of error:
- Type I error
  - Rejection of the null hypothesis.
- Type II error
  - Failing to reject the null hypothesis.

Significance level

- We cannot eliminate the risk of making one of the two errors.
  - The probability of making type I error is know as **the significance level of statistical test**. (Also called the alpha value)
  - The probability of making type II is called the beta value.
- Mostly in experiments and surveys, **type I error is the concern** and probability of making this error is set low ($p < 0.05$ or 5%)

	Drug has no effect	Drug has effect
REJECT H_o	Type I error	No error
DO NOT REJECT H_o	No error	Type II error

Consequence of making either Type I or Type II error

- Type I Error
  - If we believe the null hypothesis we will not use the drug.
  - Consequence: Since the drug is beneficial, by withholding it, we will allow patients to die who might otherwise have lived.
- Type II Error
  - If we reject the null hypothesis, we will use the drug.
  - It will be of no benefit to the patients. Also since we think we have found the cure we may no longer test other drugs.

- We can never know the “True State of Nature”, but we can infer it on the basis of sample evidence and the use of statistics.
- We will reduce the probability of making a mistaken conclusion.

Once there was a King who was very jealous of his Queen. He had two knights, Alpha who was very handsome, and Beta who was very ugly. It happened that the Queen was in love with Beta. The King however suspected the Queen was having an affair with Alpha and had him beheaded. Thus the King made both kinds of errors: he suspected a relationship (with Alpha) where there was none, and he failed to detect a relationship (with Beta) where there really was one. The Queen fled the kingdom with Beta and lived happily ever after, while the King suffered torments of guilt about his mistake and fatal rejection of Alpha.

Test of Significance or Hypothesis tests

- There are a variety of tests to chose from, the particular choice depending on the nature of the data and the context in which the comparison is being made.

Charlie Brown was addressing his baseball team at the end of the season. He recited numerous dismal statistics such as: Runs scored by us 12, by opponents 125.

At the end of the speech he yelled out: "And what are we going to do about it?"

"Get a new statistician!“ shouted the team

References

- Andy Field and Graham Hole. How to Design and Report Experiments. Sage Publications. London. 2006.
- Martin Bland. An Introduction to Medical Statistics. 2nd ed. Oxford Medical Publications. 1996.
- Norman T. J. Bailey. Statistical Methods in Biology. 3rd Ed. Cambridge University Press. 1995.
- Beth Dawson and Robert G. Trapp. Basic and Clinical Biostatistics. Lange Medical Books/ McGraw-Hill. 3rd ed. 2001.