Let's start with some equations…

Natural Scientists + Equations

=

Natural Scientists + Statistical Theory

=

Natural Scientists + Lectures on Computer Coding

=

So…

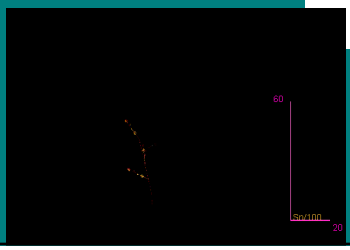
- As few equations as possible
- As little statistical theory as possible
- The best way to learn the software and methods is to USE THEM!
- So plenty of practice in labs and after them…
- No sleeping – questions welcome!

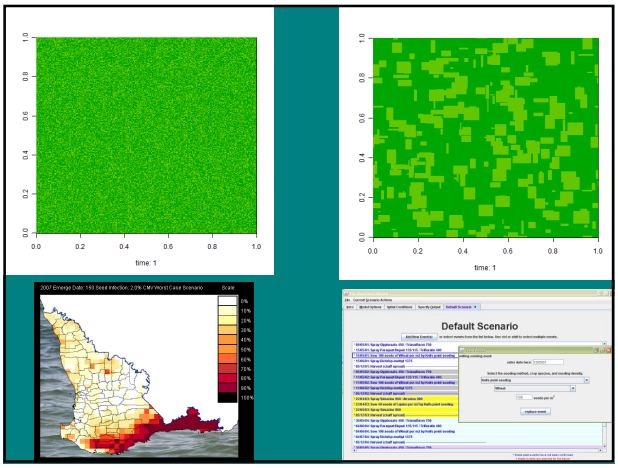
Ground Rules

- Open
- Interactive
- Informal
- Please ask questions!

Me

- I can balance on my elbows
- I never studied statistics at university
- I travelled across the Middle East when I was 18
- I never studied biology at university
- I spent a year modelling apple trees in France





And lots of stats...

You?

- Statistics background?
- Feeling for statistics?

A Warning!!

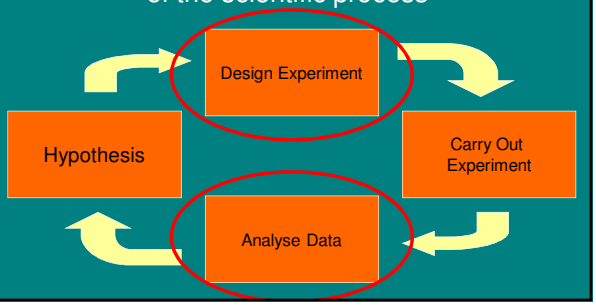
Stats is an art,
as much a science…
There are NO
'right' answers
or absolutes

or approaches that always work.
The field of stats is alive and dynamic
and changing and controversial,
just like science!

It may not be possible to simply follow a recipe

Another Warning!

Statistical analysis is an integral part of the scientific process

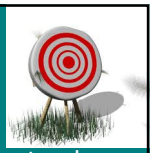


- To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of.
  ~ Sir Ronald Aylmer Fisher

  The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data.
  ~ John Tukey

This Course

Aims



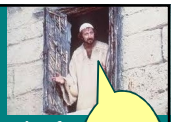
- for you to gain an overview and understanding of some of the statistics required for research in natural and agricultural science,
- learn to use a powerful and free software package for data management and statistical analysis (R),
- to get you to think seriously about the design of your project and how you will analyse the data you collect – *from the beginning!*

Approach

- graduate course: mostly self-guided independent study (Lots!?!?)
- gener… …ing, rather… …etc – you need t… …would be applie… …heck that it work… …interpret its res… …natical details involved…

- ongoing assessment to ensure that you are keeping up

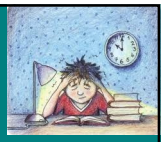
Assumptions



- you have all done some statistics before undergraduate courses or a 'catch-up' course
- BUT everyone will have different backgrounds, knowledge and experience
- Only YOU will be able to know what you need to work on to pass the course
- You will work this out through independent study and ongoing assessment

You are all different !!!!

Assessment



- 20% each week
- Quizzes (5% each)
- Assignment Reports (15%/20% each)

- My assessment will be done online using LMS/Moodle
- Quiz and assignment questions provided on LMS
- Should be answered before doing assessment
- You will be able to take each quiz up to three times
- You will have limited time to enter answers so you will need to study lots before (or do assignment thoroughly)
- There will be a closing date
- Check email and LMS

Quiz 1

- Opens today!
- Closes before Day 1!
- Preparation / revision
- Internet Explorer may not work!!

Project Prep

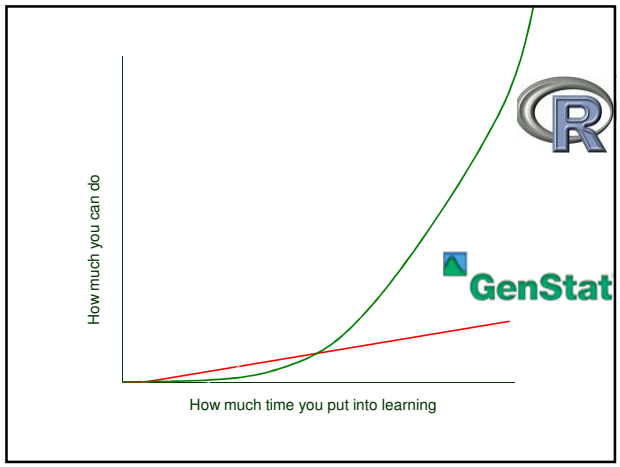
- An exercise focussing on preparation for the stats for your research project is on LMS – strongly recommended!

Resources

- Lectures and lecturers
- Lab and demonstrators
- CAST – Biometrics workbook
- Online videos on R (LMS - FF not IE, google)
- Library – Books (eBooks)
- Internet - you should check a few sites for cross-checking
- List of books and internet resources on LMS
- Your supervisor(s) (for project specific)

Why R?

- Free, open-source
- Many, many free libraries
- Scripted
  - Difficult to get started
  - Very powerful
- Manuals & Books
- Etc, etc, etc



Installing and Running R

- Great idea to install it on your own computer (bring laptop?)
- Type 'R' in google to find main R website
- Follow links to download installer…
- Run installer
- Click icon to run
- Or in labs find it in the menu
- Any problems??

Interface

- The windows – console, R script(s), graphics
- Sending stuff from the script to the console – copy/paste, crtl R, menu
- Writing and saving a script – why it's awesome!
- GUI options like Rcommander – why we're not teaching them
- Rstudio

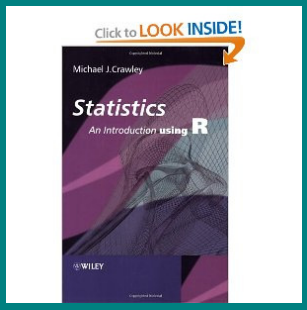
Demo Basics

Guided T-test Example

Help and Documentation

- Scripts – mine and yours
- Help search and help command – look at T-test in detail
- Pdfs that come with R

R Books



The R Book

- Available online...
- For Day 1:
  - The first few pages of chapter 2
  - Chapter 3
  - Chapter 4
  - Chapter 5
  - Chapter 8

Help and Documentation

- Scripts – mine and yours
- Help search and help command – look at T-test in detail
- Pdfs that come with R
- Books – see list
- Free online tutorials eg CSIRO – see list
- Google - forums
- People – love, tutoring, co-authorship, fees
- Courses – online, Maths Dept, others
- Maths Dept Stats Clinics
- Ecomod Group
- R users group???

Project Preparation Exercise

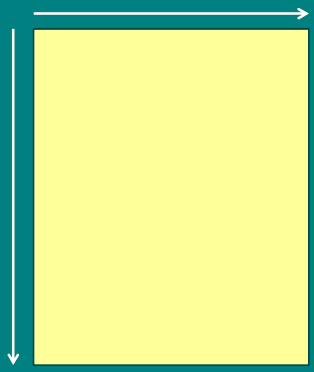
1. Give a brief background/overview for your study (no more than two or three paragraphs)
2. What are the experimental hypotheses and/or questions? (a few sentences)
3. What is the experimental design? Discuss treatments, sampling, factors, levels of factors, covariates, repetitions and/or randomisation etc (no more than two paragraphs, plus figures perhaps… you can chose to describe only a subset of your experiment if it is too big to fit.)
4. What kind of data will you get? (no more than one paragraph)
5. How will you present this data? Why? (no more than one paragraph, plus example figures perhaps)
6. What are the formal 'statistical' hypotheses you will test (dot points)
7. What statistical tests or methods will you use to test these hypotheses? Why? (no more than one paragraph)
8. What software will you use to carry out these tests?
9. How will you present the results of these tests?

R Error Messages

- Error: object 'x' not found
- Error: unexpected ',' in "a <- 2,“
- Error in a + 2 : non-numeric argument to binary operator
- Error: could not find function "stdev“

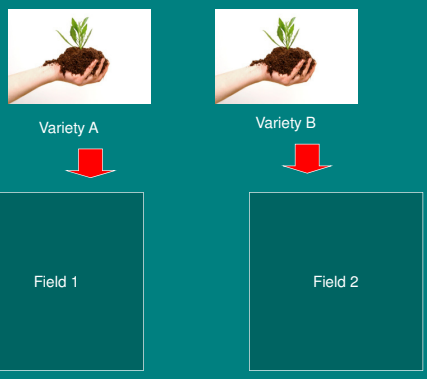
$

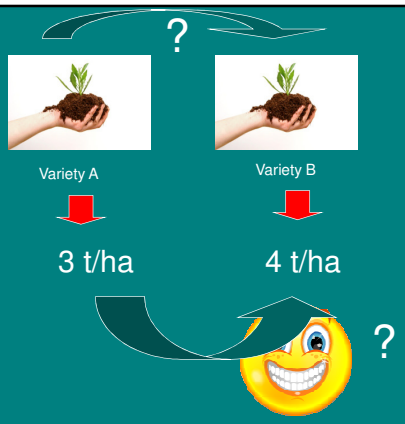
my.data.frame$biomass



- R does not read excel files directly
- csv
- text (separated by spaces)

What are the sources of variability in experimental measurements?



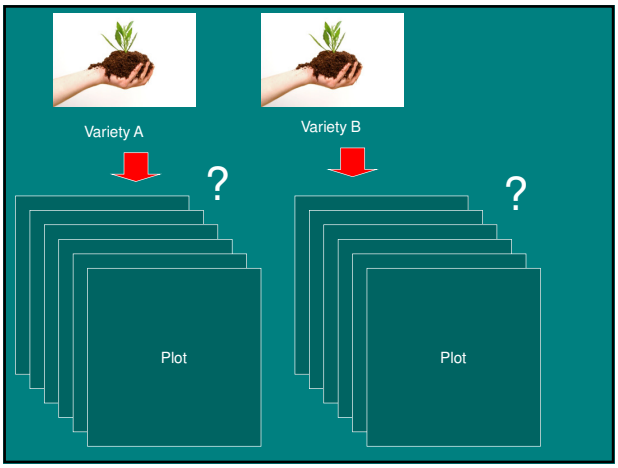


What else could be causing this difference?

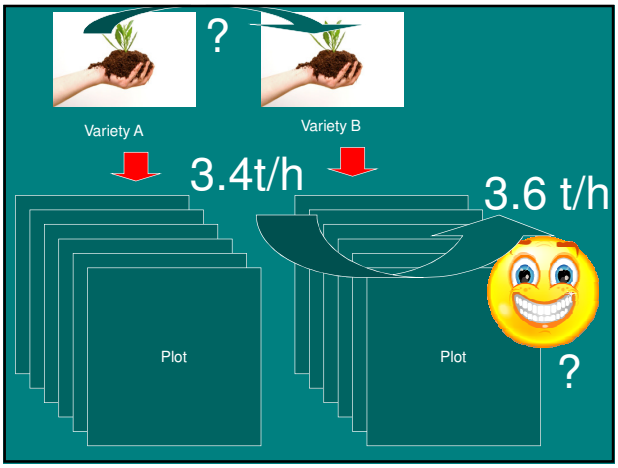
- Soil
- Insects
- Weeds
- Sowing
- Measurement
- Micro-climate

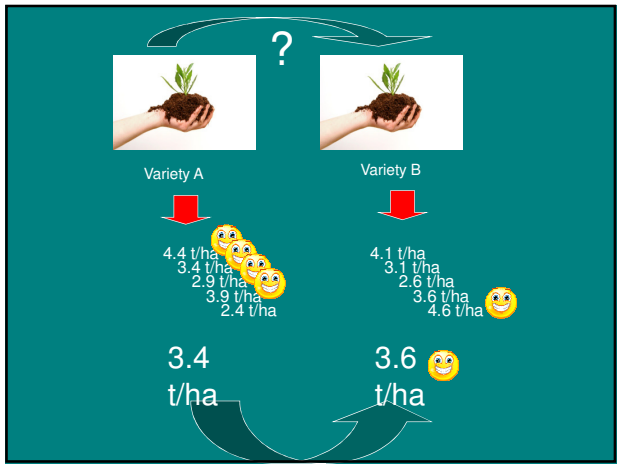
- So how can I make the test fair?

Replicates!



So I take the average! (mean)





Statistical Tests

… give us a fair way to test whether the differences we see between cultivars, treatments etc…

…could have just occurred by chance…

…OR…

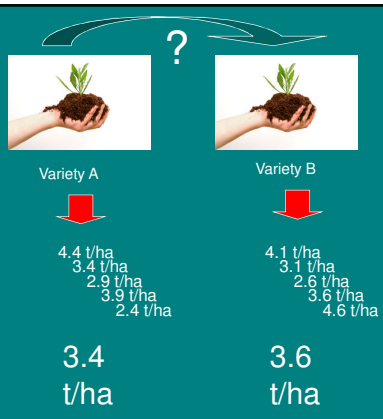
…are REALLY likely be due to differences between cultivars, treatments

Populations and Samples

- Our real questions usually concern large populations
  - Are fish in a marine reserve bigger than fish outside the reserve?
  - Is variety A better than variety B?
- But we usually measure a small sample
- Why do we take samples?
- What is the problem with a sample?
- How big a sample is big enough???

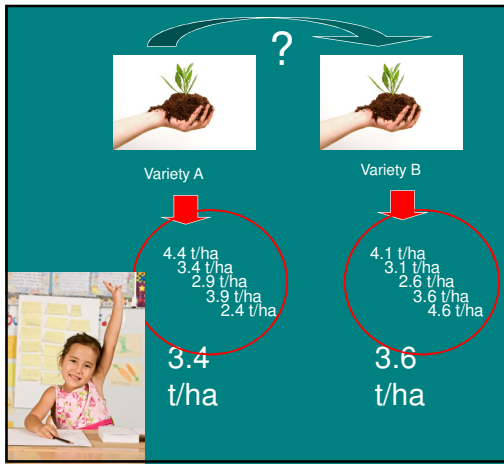
So what tests are there?

?



T-test

- For testing whether two treatments, cultivars etc are different…



Results of T-Test

t = -0.4, df = 8, p-value = 0.6996
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-1.353002 0.953002
sample estimates:
mean of x mean of y
3.4 3.6

P Value

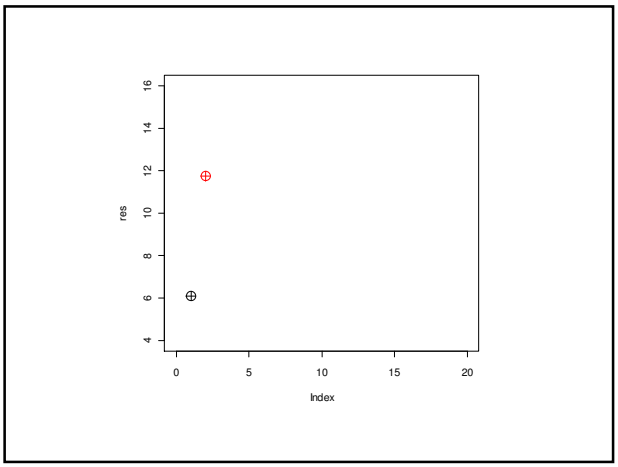
- The probability that data with this difference could have happened if there isn't really any difference…
- If it's small, normally p<0.05, then probability is small

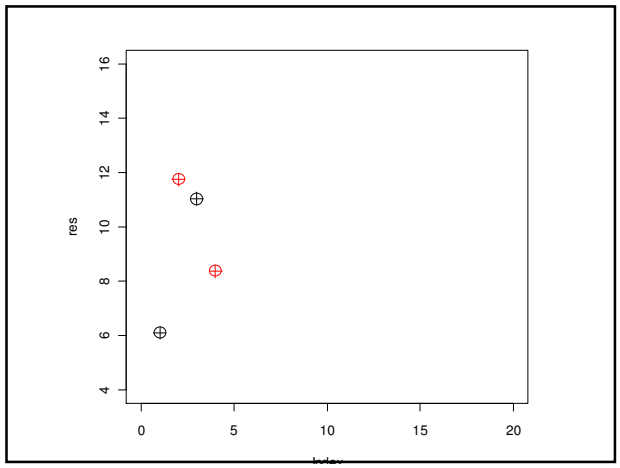
…So probably is a REAL difference

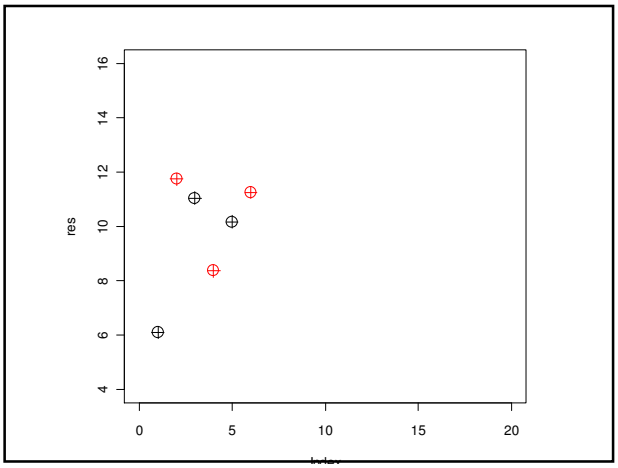
…So difference is 'significant'!

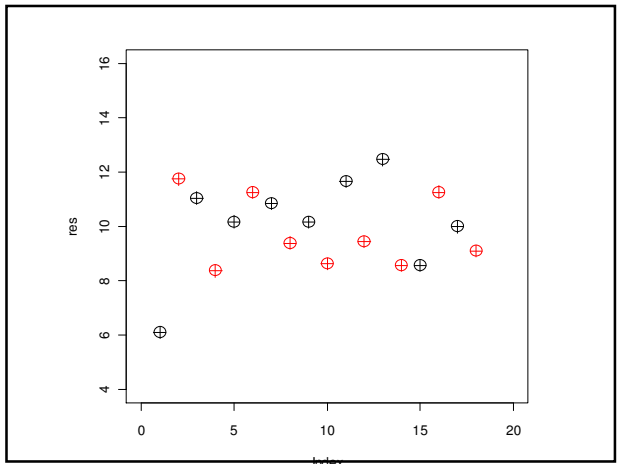
Note confidence interval

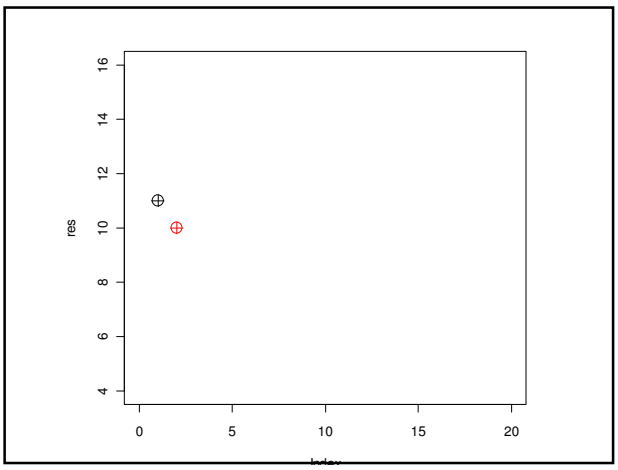
What does it mean if p>0.05??

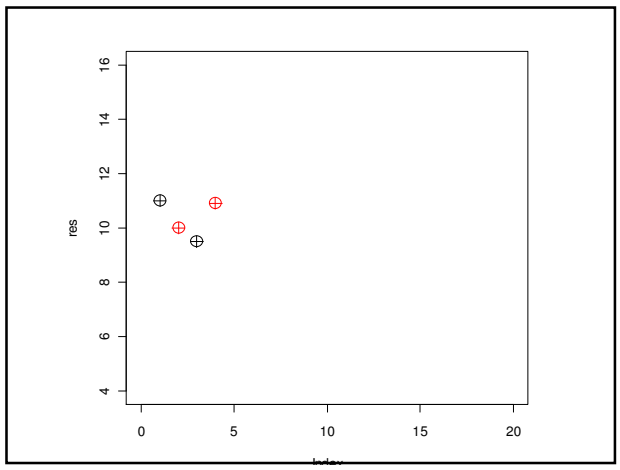

res
Index

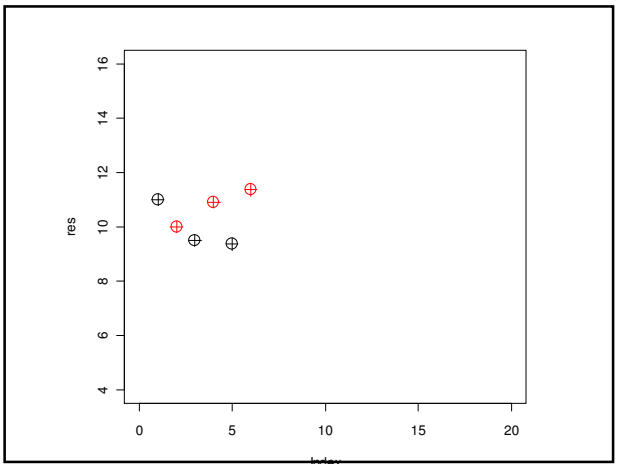

res

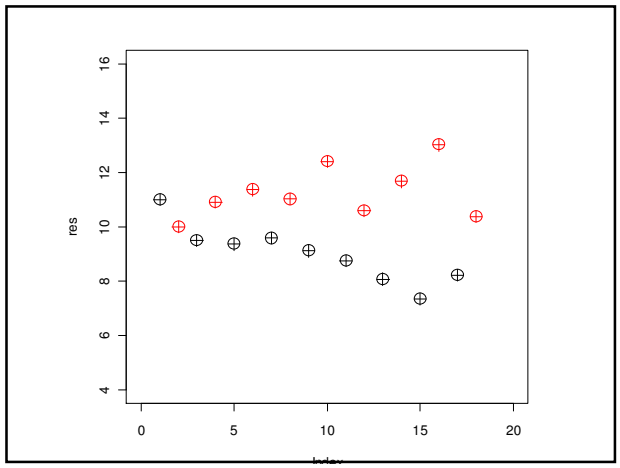

res

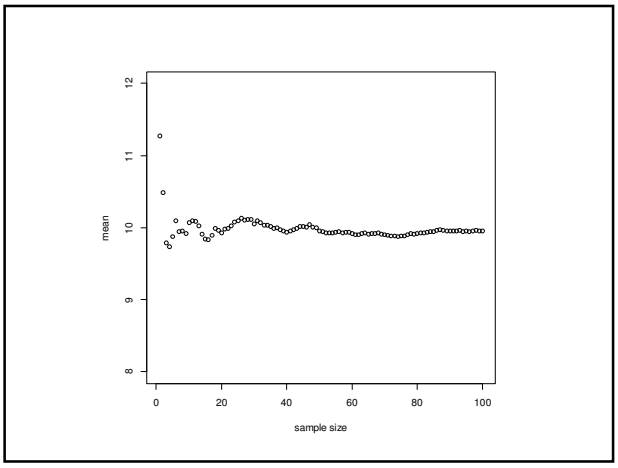

res

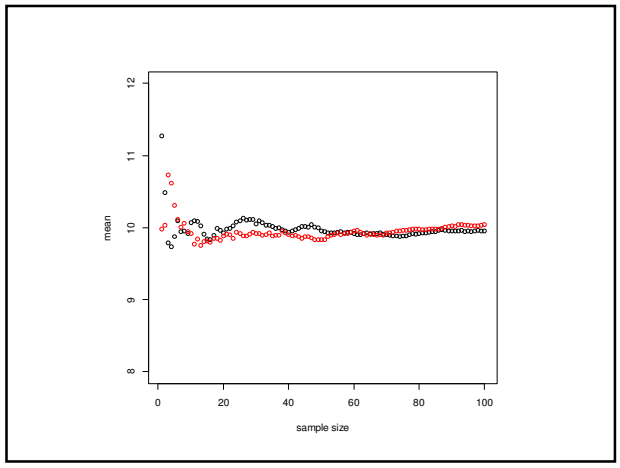

res

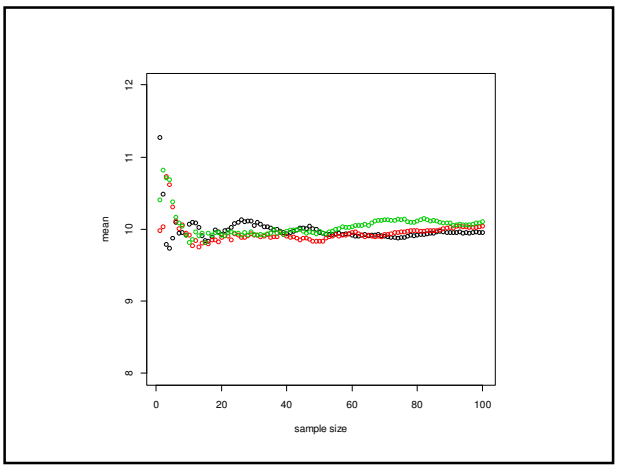

res

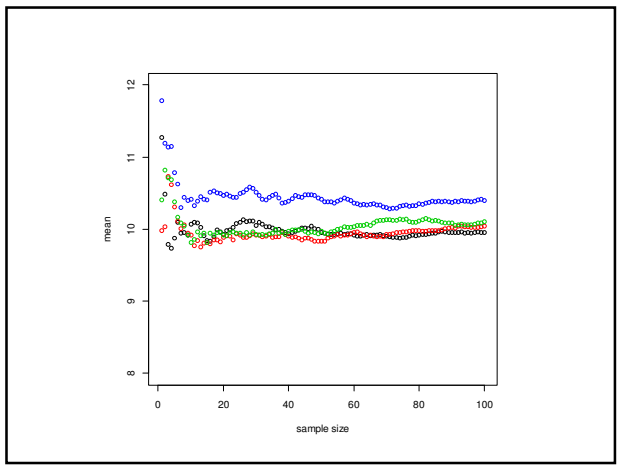

res
index


res
index


mean
sample size


mean
sample size


mean
sample size


mean
sample size

T-Test

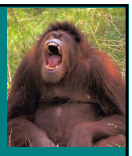
Assumptions

- Normality (??)
- Equality of variance (?? not necessary)

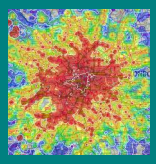
Types

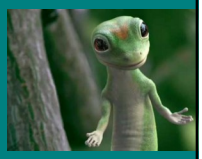
- Paired and unpaired (??)
  - Think of an example of each
- One-sided or two-sided
  - Think of an example for each

Types of Data

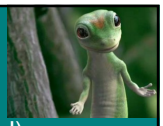


- Categorical data
- Count data
- Continuous data
- Binomial data
- Time series
- Survival times
- Spatial data





Counts of different types



- Two different categories (binomial)
- Two different categories with two different categories
- .....
- Multiple categories with multiple categories

	Site1
Male	12
Female	34

	Site1	Site2
Male	12	78
Female	34	93

	Site1	Site2	Site3	Site4
Species1	12	78	31	134
Species2	34	93	53	145
species3	23	25	54	147

Let's look at some examples in R