statstutor

community project

encouraging academics to share statistics support resources

Statistical Methods
1. Introduction

Based on materials provided by Coventry University and Loughborough University under a National HE STEM Programme Practice Transfer Adopters grant

www.**stats**tutor.ac.uk

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Summary

- What is statistics?
- What is a mean?
- Data types
- The research study process
- The statistical analysis process
- Some basic statistical concepts
- Benefits of good study design
- Comparison of two study designs

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Activity: What is statistics?

1 minute:

- Write down your own definition

2 minutes:

- Discuss it with your neighbour and agree on a definition

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

What is statistics?

The word “statistics” is used in 3 main ways:

1. **Common meaning**: factual information involving numbers. A better word for this is **data**.
2. **Precise meaning**: **quantities** which have been derived from sample data, e.g. the mean (or average) of a data set
3. **Common meaning**: an academic subject which involves **reasoning about statistical quantities**

⇒ In order to use statistics properly you need to be able to **think about statistics** in the right way

www.**statstutor**.ac.uk

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

The three main areas of the subject of statistics

1. **Descriptive statistics** – describing and summarising data sets using pictures and statistical quantities – see Workshop 3
2. **Inferential statistics** – analysing data sets and drawing conclusions from them – see Workshops 8 to 12
3. **Probability** – the study of chance events governed by rules (or laws) – see Workshop 6

Inferential statistics is based on probability because it often uses **random samples** of data sets drawn from a population (a chance event)

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

What is a mean?

The mean of a data set is a measure of its middle value.

Example: The number of nuclear power stations in various countries in 1989.

Country	Number	Country	Number
Canada	22	Spain	10
France	52	Sweden	12
Japan	43	UK	41
South Korea	9	USA	119
Soviet Union	73	West Germany	23

To calculate the mean, add all the data values together and divide by the number of values.

$$\overline{X} = \frac{22 + 52 + 43 + 9 + 73 + 10 + 12 + 41 + 119 + 23}{10} = \frac{384}{10} = 38.4$$

www.**stats**tutor.ac.uk

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Data types

In statistics it is vital to understand what types of data you are working with.

There are three main types:

- **Nominal** – **categories** that do not have a natural order, e.g. gender, eye colour, types of building
- **Ordinal** – **categories** which have a natural order but are not numerical, e.g. Likert scales
- **Scale/continuous** – numerical data ordered against a constant scale, e.g. date, temperature, length, weight, frequency

Peter Samuels
Birmingham City University

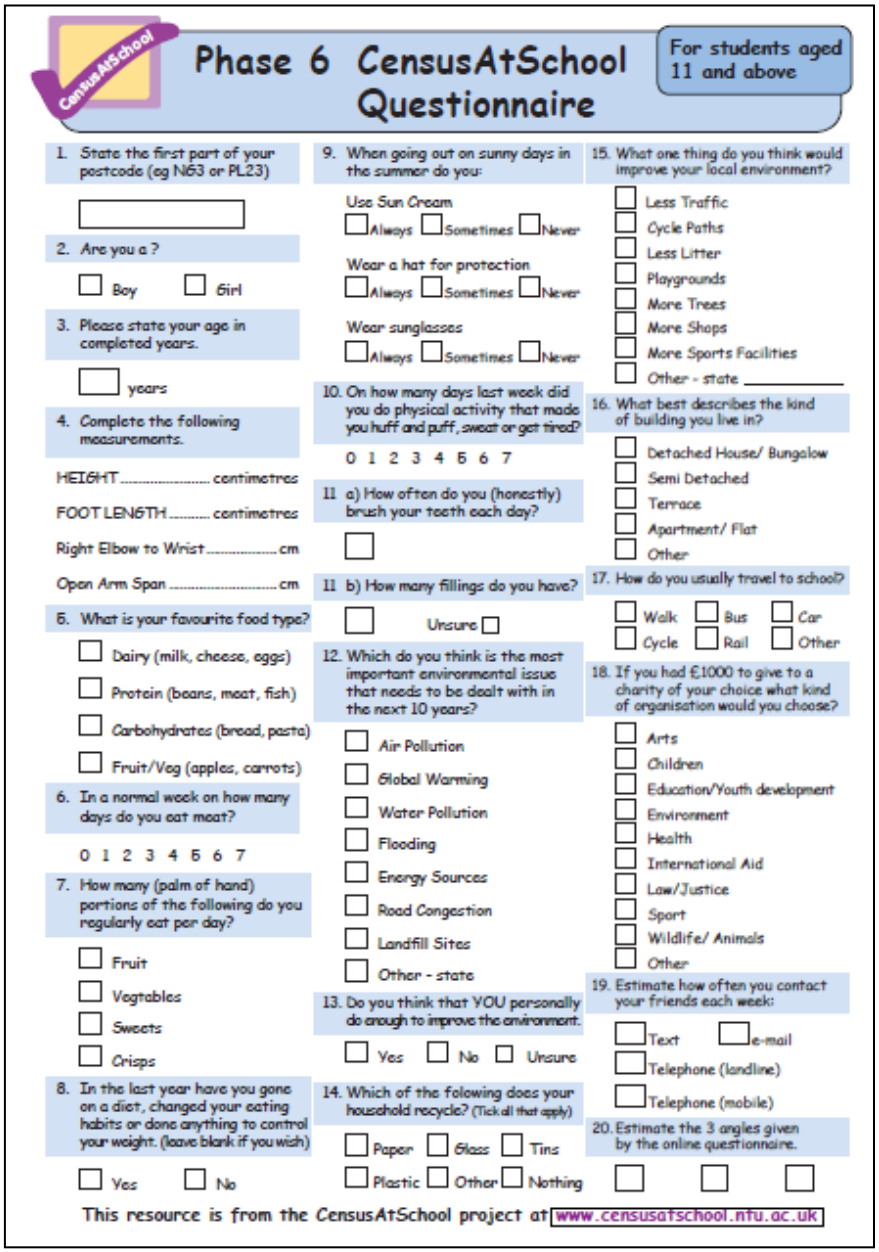
Reviewer: Ellen Marshall
University of Sheffield

Activity: CensusAtSchool Phase 6 Questionnaire

Available from:

http://www.censusatschool.org.uk/images/phases/phase6-questionnaire.pdf

Discuss with your neighbour the data type of each question



CensusAtSchool

Phase 6 CensusAtSchool Questionnaire

For students aged 11 and above

1. State the first part of your postcode (eg N63 or PL23)

2. Are you a ?
☐ Boy ☐ Girl

3. Please state your age in completed years.
☐ years

4. Complete the following measurements.
HEIGHT ..... centimetres
FOOT LENGTH ..... centimetres
Right Elbow to Wrist ..... cm
Open Arm Span ..... cm

5. What is your favourite food type?
☐ Dairy (milk, cheese, eggs)
☐ Protein (beans, meat, fish)
☐ Carbohydrates (bread, pasta)
☐ Fruit/Veg (apples, carrots)

6. In a normal week on how many days do you eat meat?
0 1 2 3 4 5 6 7

7. How many (palm of hand) portions of the following do you regularly eat per day?
☐ Fruit
☐ Vegtables
☐ Sweets
☐ Crisps

8. In the last year have you gone on a diet, changed your eating habits or done anything to control your weight. (leave blank if you wish)
☐ Yes ☐ No

9. When going out on sunny days in the summer do you:
Use Sun Cream
☐ Always ☐ Sometimes ☐ Never
Wear a hat for protection
☐ Always ☐ Sometimes ☐ Never
Wear sunglasses
☐ Always ☐ Sometimes ☐ Never

10. On how many days last week did you do physical activity that made you huff and puff, sweat or get tired?
0 1 2 3 4 5 6 7

11 a) How often do you (honestly) brush your teeth each day?
☐

11 b) How many fillings do you have?
☐ Unsure ☐

12. Which do you think is the most important environmental issue that needs to be dealt with in the next 10 years?
☐ Air Pollution
☐ Global Warming
☐ Water Pollution
☐ Flooding
☐ Energy Sources
☐ Road Congestion
☐ Landfill Sites
☐ Other - state

13. Do you think that YOU personally do enough to improve the environment.
☐ Yes ☐ No ☐ Unsure

14. Which of the folowing does your household recycle? (Tick all that apply)
☐ Paper ☐ Glass ☐ Tins
☐ Plastic ☐ Other ☐ Nothing

15. What one thing do you think would improve your local environment?
☐ Less Traffic
☐ Cycle Paths
☐ Less Litter
☐ Playgrounds
☐ More Trees
☐ More Shops
☐ More Sports Facilities
☐ Other - state _____

16. What best describes the kind of building you live in?
☐ Detached House/ Bungalow
☐ Semi Detached
☐ Terrace
☐ Apartment/ Flat
☐ Other

17. How do you usually travel to school?
☐ Walk ☐ Bus ☐ Car
☐ Cycle ☐ Rail ☐ Other

18. If you had £1000 to give to a charity of your choice what kind of organisation would you choose?
☐ Arts
☐ Children
☐ Education/Youth development
☐ Environment
☐ Health
☐ International Aid
☐ Law/Justice
☐ Sport
☐ Wildlife/ Animals
☐ Other

19. Estimate how often you contact your friends each week:
☐ Text ☐ e-mail
☐ Telephone (landline)
☐ Telephone (mobile)

20. Estimate the 3 angles given by the online questionnaire.
☐ ☐ ☐

This resource is from the CensusAtSchool project at www.censusatschool.ntu.ac.uk

www.statstutor.ac.uk

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

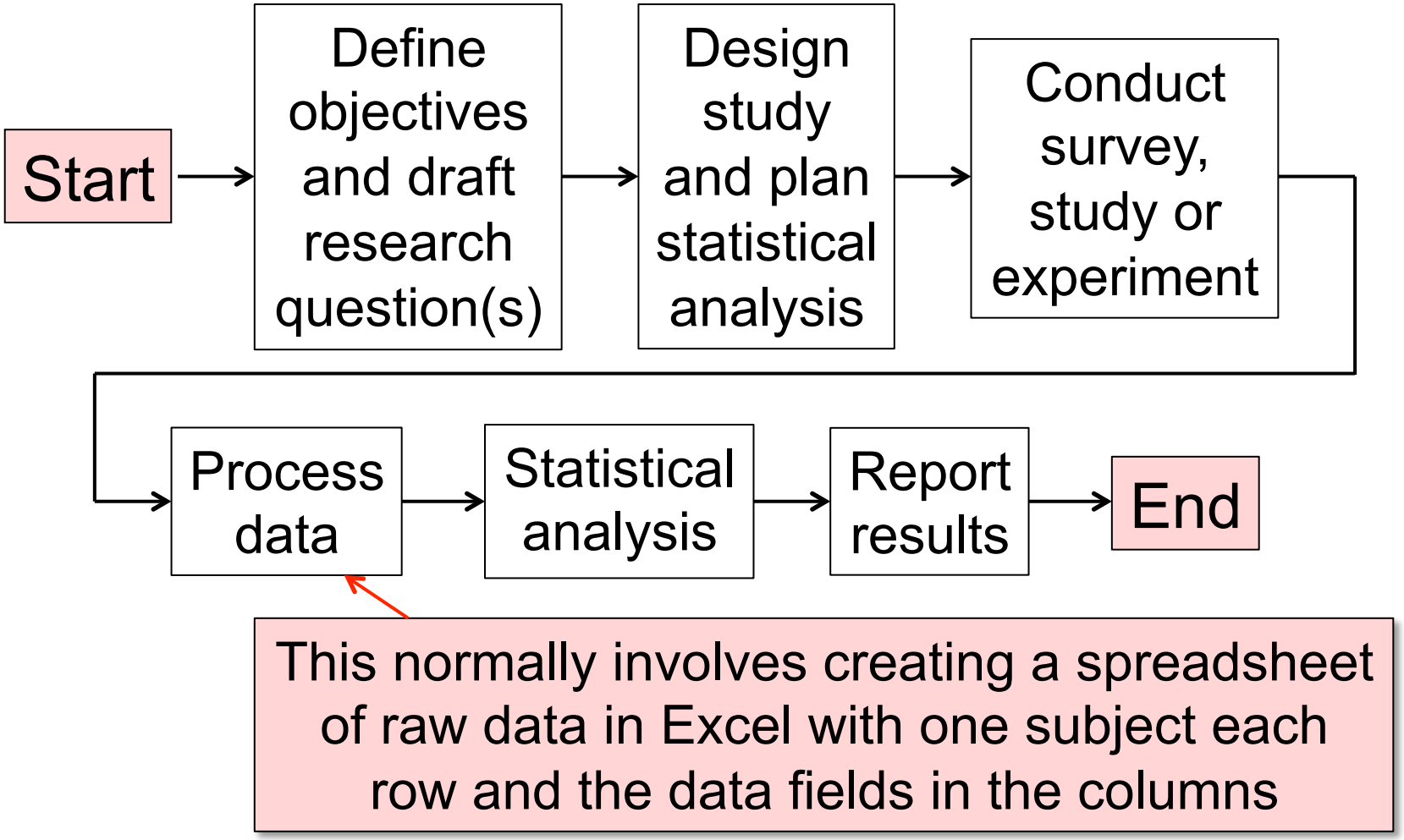
Answers

Qn. No.	Type	Qn. No.	Type
1	Nominal	11	Each: Scale
2	Nominal	12	Nominal + free text
3	Scale	13	Ordinal (unsure in middle)
4	Each: Scale	14	Nominal (multi-answer)
5	Nominal	15	Nominal + free text
6	Scale	16	Nominal
7	Scale	17	Nominal
8	Nominal	18	Nominal
9	Each: Ordinal	19	Each: Scale
10	Scale	20	Each: Scale

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

The research study process



Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Comments on the research study process

It is important at the **outset** to:

- Make objectives/research question(s) clear and unambiguous (hypothesis-driven or curiosity-led?)
- Identify what data you need
- Plan your statistical analysis **before** you collect any data

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

The statistical analysis process

Demon-strate that you are in control of the process!

- ❑ Make sure you have a good data set to start with
- ❑ Generally we advise using Excel (see Workshops 4 and 5) before using SPSS (see Workshop 7)

1. First describe and present your data, e.g. frequency distributions in tables or charts
2. Calculate basic statistics where possible, e.g. means and standard deviations
3. Start to interpret your data – what might it mean?
4. Select specific items for closer attention (based on your research hypotheses)
5. Select and carry out the right kind of test
6. Interpret your findings in terms of significance levels
7. Modify and repeat if necessary

www.**statstutor**.ac.uk

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

How statistical analysis can help you

- It allows you to make 'sense' of data
  - Descriptive (e.g. numerical or graphical, etc.)
- It allows you to evaluate uncertainty and make valid inferences
  - Make comparisons (e.g. between two groups)
  - Model orientated (e.g. model how blood pressure is affected by gender and age)

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

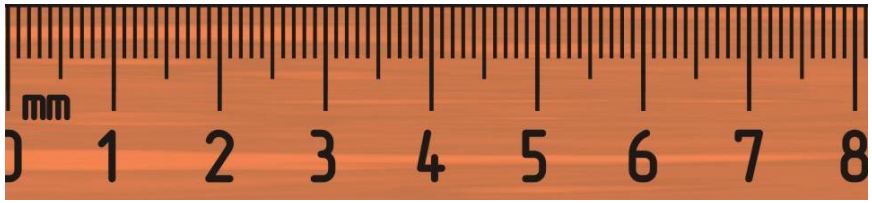
Basic statistical concepts

- Reliability and validity
- Bias and precision
- Data richness
- Populations and samples
- Parameters and estimates
- Random selection
- Robustness

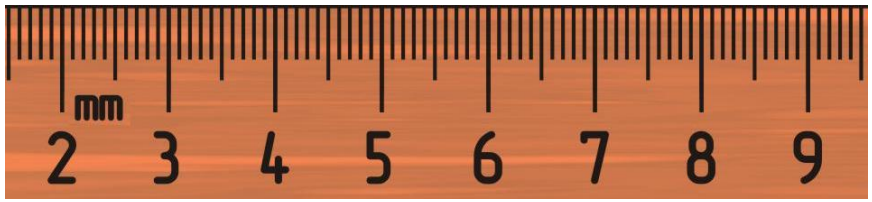
Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Reliability and validity



Valid and potentially also **reliable**, depending upon how it is used and whether the object / person being measured is always the same



Invalid as it doesn't measure what it is supposed to

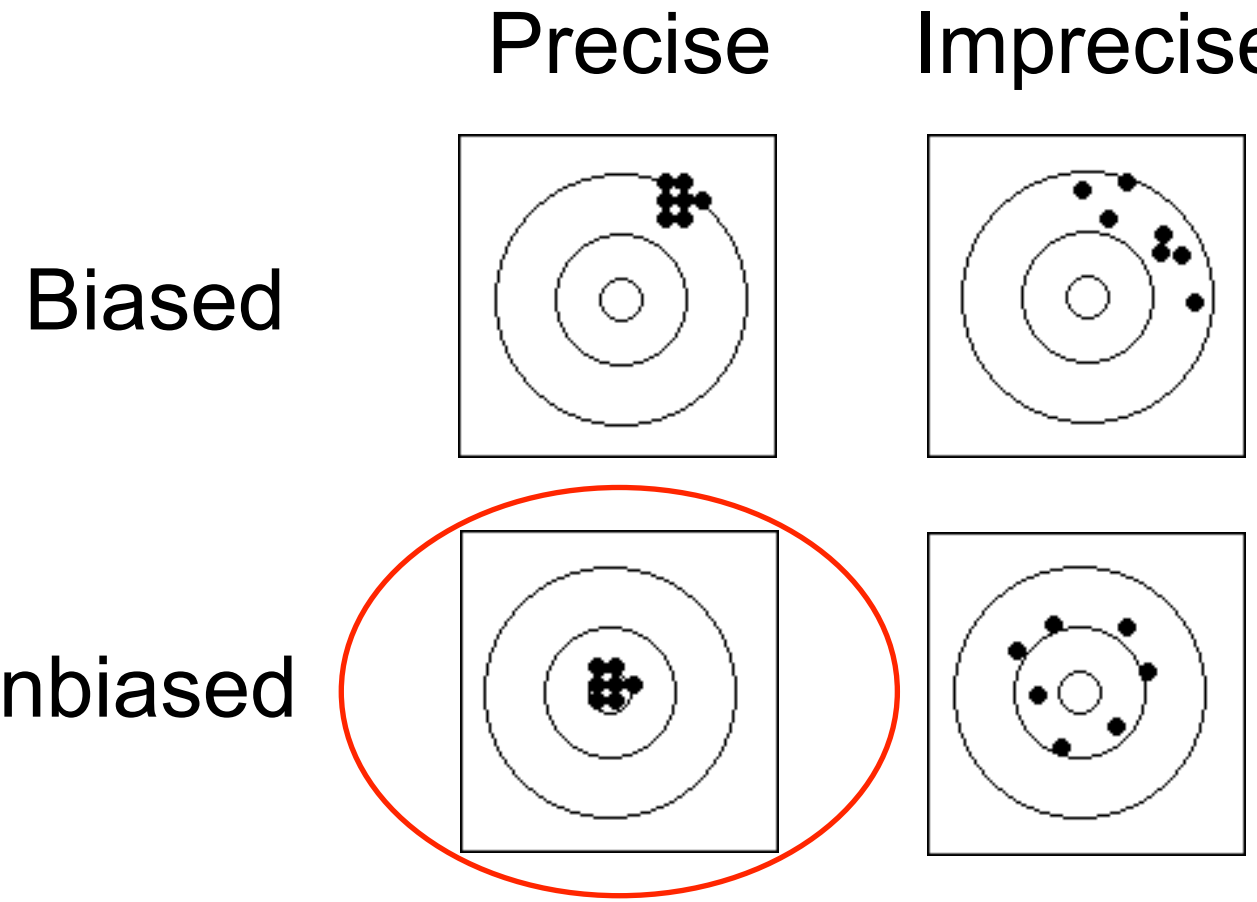
- An instrument is **valid** when it measures what it is supposed to measure
- An instrument is **reliable** if the same results are obtained when it is retested
- Standard instruments have usually already been tested for reliability and validity
- You will probably not be expected to show reliability and validity of your instrument (except possibly in Psychology)

www.statstutor.ac.uk

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Bias and precision



www.**statstutor**.ac.uk

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Data richness

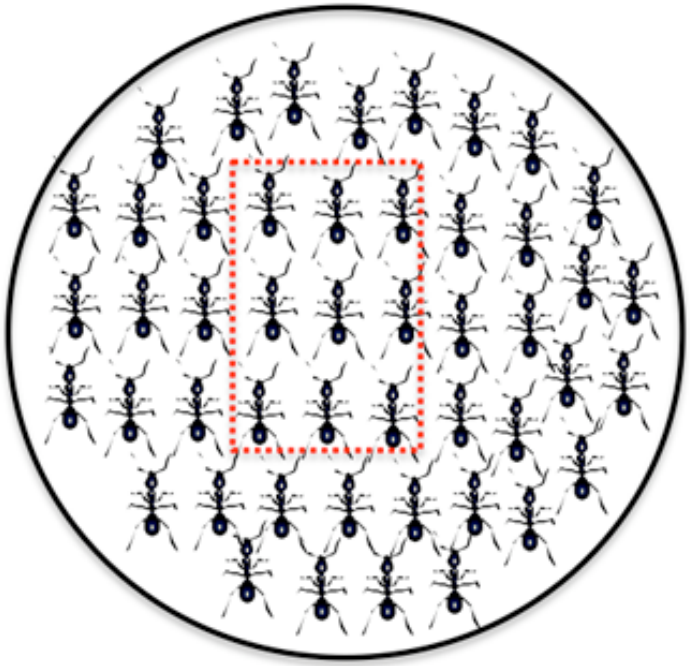
- You should always use the richest (most detailed) data available because it will give more accurate results
- Here, the Age data is richer than the Age Category data
- However, there might be ethical issues in obtaining detailed data
- Here, the respondents might feel embarrassed to give their exact age

Age	Age Category
29	25-29
50	40+
27	25-29
27	25-29
31	30-30
24	18-24
31	30-30
32	30-30
34	30-30
17	18-24

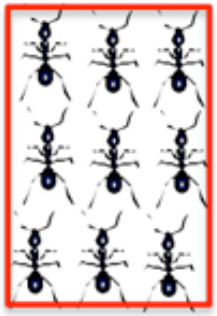
Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Populations and Samples



Population:
May be too big / expensive to study



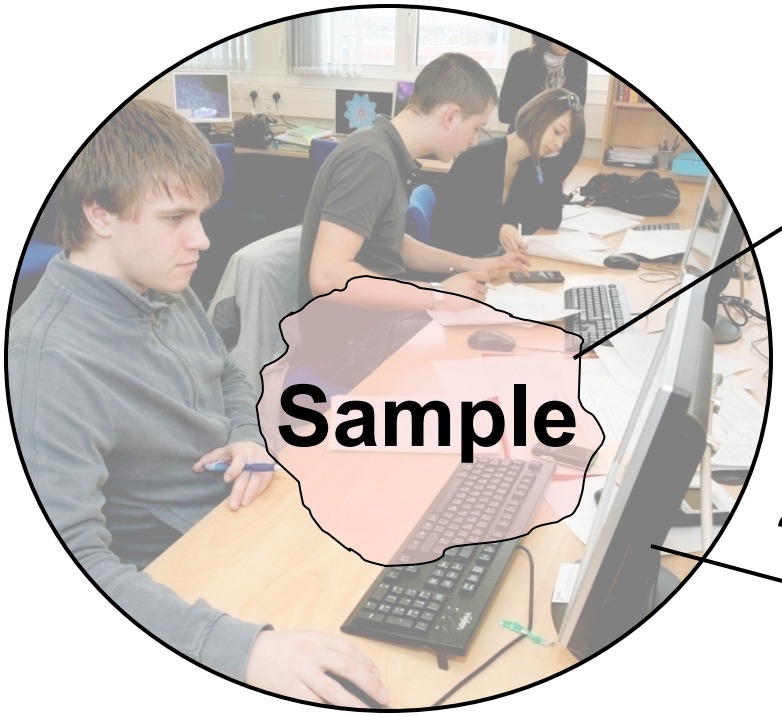
Sample:
We can learn nearly as much by studying a suitably large **randomly chosen sample** of a population as we can from studying the entire population

www.**stats**tutor.ac.uk

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Parameters and estimates



Population of students at Birmingham City University

Sample mean (e.g. age) **Estimate**

Estimates

?

Population mean (unknown) **Parameter**

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Random selection

- Most research study designs require a sample to be **randomly selected** from a population
- Research[1] suggests humans cannot generate random numbers and thus cannot make random selections
- Suggested methods:
  - Select numbered balls out of a bag (as in the National Lottery)
  - Use an online random number generator, such as www.random.org/integers
  - Use the RAND or RANDBETWEEN functions in Excel
- More details in Workshop 13

1. Bains, W. (2008) Random number generation and creativity, *Medical Hypotheses*, 70(1), pp. 186-190

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Robustness

- Parameter-based statistical tests make certain assumptions in their underlying models
- However, they often work well in other situations when these assumptions are violated
- This is known as **robustness**

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Why is study design important?

- Ensures you collect 'good' data
- Allows you to draw valid conclusions and answer your research question(s)
- Reduces potential bias
  - E.g. Staff stress survey – Perhaps staff who have been stressed are more likely to respond

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

More reasons why good study design is important

- Reduce variability in your data
  - Reduces 'noise‘
  - Enables you to see the big picture
- Improves accuracy (precision) of results
- Reduces amount of data needed
- Reduces cost (time or money)
- Surveys or observational studies **cannot** identify causes and effects
- Designed experiments **can**!

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Activity: In-car control panel design

A new type of car control panel has been developed to control various functions within a vehicle, e.g. air conditioning, heater, radio/CD etc.

Two studies were undertaken where subjects used a driving simulator, so that their mean distraction time could be measured using eye-tracking technology, whilst driving and using various control panel functions.

The idea behind the studies was to ask subjects to use the new design in the driving simulator and then repeat this using a standard design of control (i.e. one found in a large number of cars currently on the road).

The aim was to assess the research hypothesis that the new control reduces distraction times.

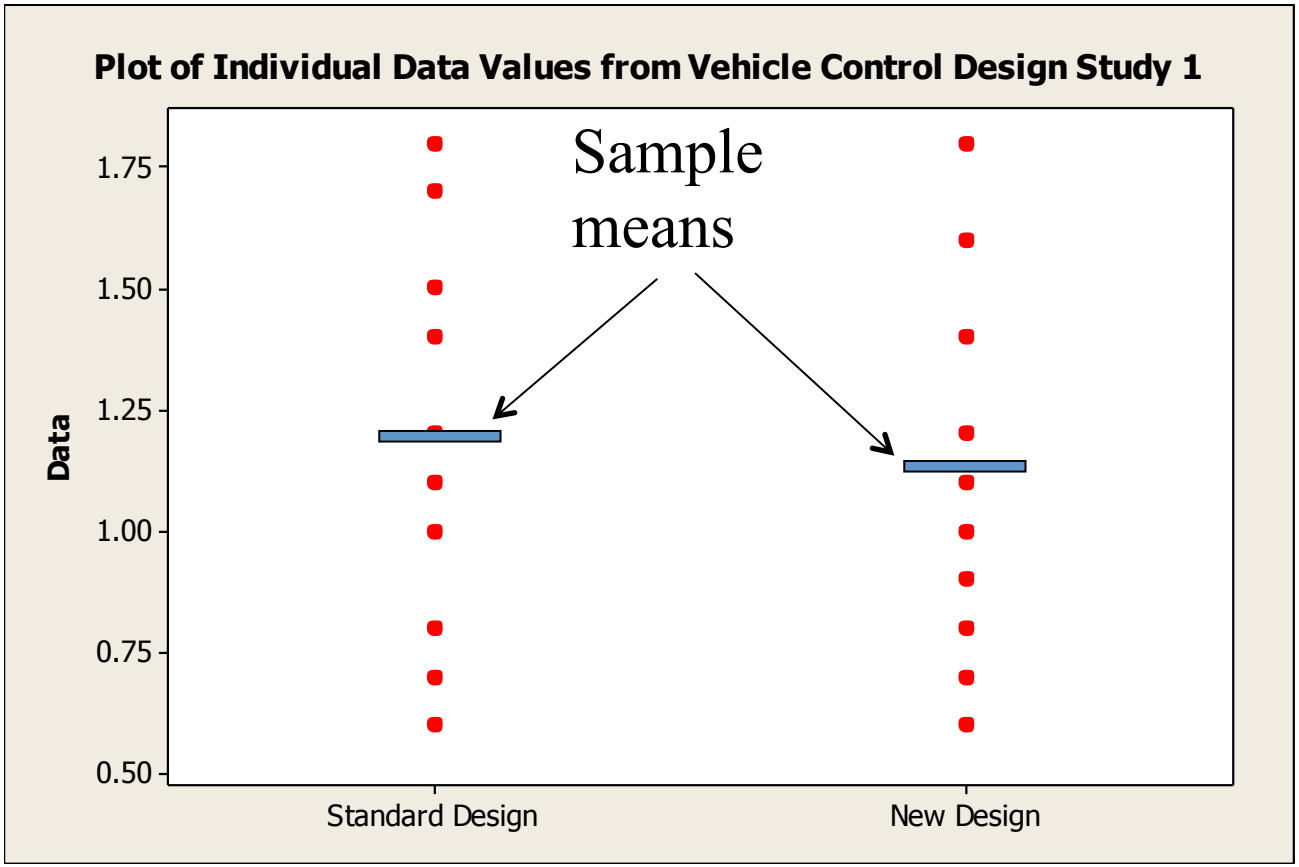
Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Study 1

Ten subjects used the new design in a driving simulator, whilst ten **different** subjects used the standard design.

A plot of their distraction times is shown on the right.



Discuss with your neighbour whether you believe this supports the research hypothesis that the new control reduces distraction times.

www.**stats**tutor.ac.uk

Peter Samuels
Birmingham City University

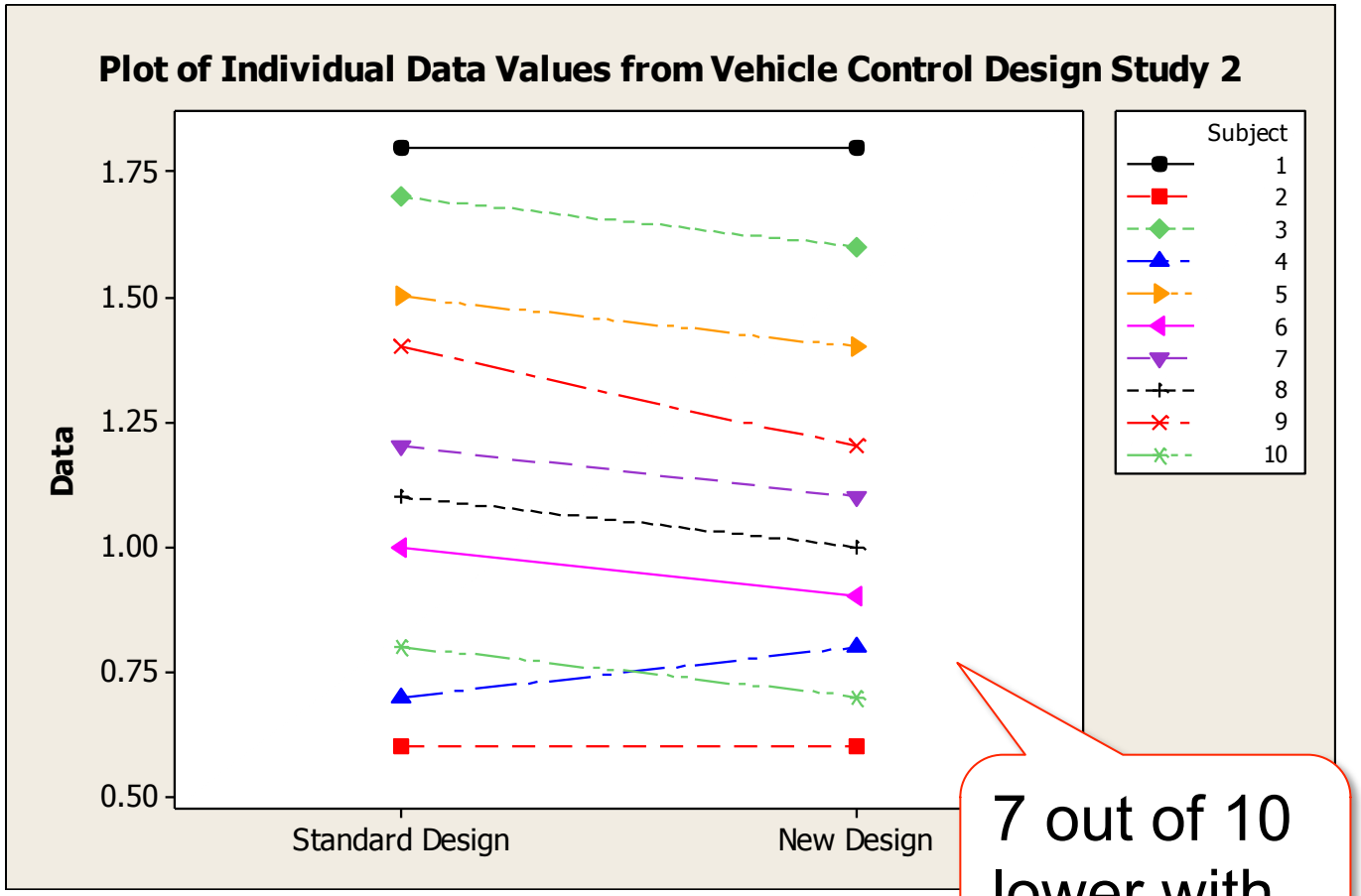
Reviewer: Ellen Marshall
University of Sheffield

Study 2

As for Study 1 but the **same** ten subjects used each design.

A plot of their distraction times is again shown on the right.

Again, discuss with your neighbour whether you believe this supports the research hypothesis that the new control reduces distraction times.



Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Recap

We have considered:

- ❑ What is statistics?
- ❑ The mean of a data series
- ❑ Data types
- ❑ The research study process
- ❑ The data analysis process
- ❑ Some basic statistical concepts
- ❑ Benefits of good study design
- ❑ Two study designs

Peter Samuels
Birmingham City University

Reviewer: Ellen Marshall
University of Sheffield

Bibliography

CensusAtSchool (2014) *2005/2006 CensusAtSchool Questionnaire*. [pdf] Available at: http://www.censusatschool.org.uk/images/phases/phase6-questionnaire.pdf [Accessed 6/01/14].

Coolican, H. (2009) *Research Methods and Statistics in Psychology*, 5th ed., London: Hodder and Stoughton.

Easton, V. J. and McColl, J. H. (n. d.) Online statistics glossary, version 1.1. Available at: http://www.stats.gla.ac.uk/steps/glossary/alphabet.html [Accessed 6/01/14].

Gonick, L. and Smith, W. (1993) *The Cartoon Guide to Statistics*, New York: HarperCollins.

Hayslett, H. T. (1991) *Statistics Made Simple*, 3rd ed., London: Made Simple Books.

Phillips, J. L. (1999) *How to think about statistics*, 6th ed., New York: Henry Holt.

Rowntree, D. (2000) *Statistics without Tears: An introduction for non-mathematicians*, New ed., London: Penguin.

Stirling, W. D. (2013) *Textbooks for Learning Statistics: Public CAST e-books*. Available at: http://cast.massey.ac.nz/collection_public.html [Accessed 6/01/14].