UXD Revision

# Week 9 - Quantitative Research

Quantitative research looks at the magnitude, size, or amount of something. E.g.:

– What is the average income of a university graduate?

– What percentage of the population own a TV?

– What is the average age of an e-sports professional?

• Quantitative research works with bigger user groups

• Data analysis is relatively quick, but requires statistical knowledge

## Measures 1: Questionnaires

Allow us to gather structured feedback from people - usually consists of a range of items (questions) and is focused on specific topic.

• Can be used independently to gather data (Surveys)

• Or as part of a user study

### Question types

|  |  |
| --- | --- |
| Closed Questions Open Questions | Closed Questions Open Questions |
| Easy to answer Some people have difficulty | Easy to answer Some people have difficulty |
| expressing themselves | expressing themselves |
| Easy to analyse data Costly to analyse data | Easy to analyse data Costly to analyse data |
| Limits number of possible options Might get too many alternative | Limits number of possible options Might get too many alternative |
| answers, might get answers that are | answers, might get answers that are |

**Common approach is to rely mostly on closed questions, with a small number of open questions.**

### Types of Closed Questions – Binary Choice

Great for simple yes or no questions.

Easy to analyse: 90% said yes.

Not very rich response.

Did you like the interface?

o Yes

o No

### Types of Closed Questions – Multiple-Choice

Respondents offered a choice of explicit nominal responses. Need to make sure all options are needed. Otherwise, include “Other” option.

How do you most often get help with the system? (tick one)

O on-line manual

O paper manual

O ask a colleague

Which types of software have you used? (tick all that apply)

O word processor

O data base

O spreadsheet

O compiler

### Types of Closed Questions – Ranking

Rank the usefulness of these methods of issuing a command

(1 most useful, 2 next most useful..., 0 if not used

\_3\_\_ command line

\_1\_\_ menu selection

\_2\_\_ control key accelerator

• Respondent places an ordering on items in a list

• Useful to indicate preferences

• Forced choice

• Limit the number of items

### Types of Closed Questions – Rating

Respondents offered a choice of from a numeric scale

The characters on the computer screen are hard to read:

1 2 3 4 5

Strongly disagree Strongly agree

**Also known as Likert scales.**

Usually agree/disagree. But can be None/Many, Never/Frequently.

Measure intensity, degree of agreement, quantity, frequency.

Important questions: How many options? Should there be a middle?

Should there be a N/A response?

### Types of Closed Questions – Semantic Differential Scales

* Bi-polar attitudes about a concept

• Pair of adjectives

The look and feel of the website is

**Exciting** 1 2 3 4 5 6 7 **Boring**

**Annoying** 1 2 3 4 5 6 7 **Pleasing**

### Types of Closed Questions – Combination

Mix between an open and closed question. The closed question gives precision, and the open question adds depth and context.

How do you most often get help with the system? (tick one)

O on-line manual

O paper manual

O ask a colleague

Why do you prefer this method?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

### Question design

Three main rules of question design:

1. The participant should be able to comprehend the question.

2. The participant should be capable of answering the question.

3. The participant should be willing to answer the question.

**Questions need to be:**

**Brief:** 20 words or less.

**Relevant:** Each question should relate to something you want

to measure.

### Question Design Tip

**Do you think the government should increase taxes and spending on welfare? X**

a) Yes

b) No

c) I don’t know

**Do you think the government should increase taxes? ✓**

a) Yes

b) No

c) I don’t know

**Do you think the government should increase spending on welfare? ✓**

a) Yes

b) No

c) I don’t know

**Bro Tip - Don’t ask double-barrelled questions.**

### Common questionnaire errors

**What is your age?**

a) Under 20

b) 20-30

c) 30-40

d) 40-50

e) 50-60

f) Over 60

Incorrect

**What is your age?**

a) Under 21

b) 21-30

c) 31-40

d) 41-50

e) 51-60

f) Over 60

Correct

**Ranges should not overlap.**

**How many jobs are available in your town?**

a) Many

b) A lot

c) Some

d) A few

**How many jobs are available in your town?**

a) A lot

b) Some

c) Only a few

d) None at all

**Make sure scales are ordinal, i.e. they have a clear unambiguous**

**order.**

**How often do you play computer games? X**

a) Rarely

b) Sometimes

c) Quite often

d) Very often

**How many times did you play computer games in the last week? ✓**

a) Never

b) Once or twice

c) About once a day

d) More than once a day

**Give frame of reference.**

**Don’t use relative terms.**

**Which of these do you use to learn about current events? X**

a) Newspaper

b) Radio

c) Television

d) Mobile browser

e) Mobile apps

f) Desktop browser

**Which of these do you use to learn about current events? ✓**

(Choose all that apply)

a) Newspaper

b) Radio

c) Television

d) Mobile browser

e) Mobile apps

f) Desktop browser

g) Other: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Anticipate all possible answers.

**Where did you go on holiday last year? X**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Did you go on holiday last year? ✓**

a) Yes

b) No

If yes, where did you go to?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Avoid making assumptions**

**about the participant.**

**Boris Johnson has wanted a trade deal with the EU that resembles that of Canada. Do you agree**

**with him?**

a) Yes

b) No

c) I don’t know

**Do you think we should have a trade deal with the EU that resembles that of Canada?**

a) Yes

b) No

c) I don’t know

**Don’t ask participants to agree or disagree with someone.**

### Pilot Questions

How long did it take to complete?

• Were the instructions clear?

• Were any questions ambiguous?

• Were any questions objectionable?

• Was the layout clear and easy to follow?

• Were any topics omitted?

**Use iterative design to improve the quality of your**

**questionnaire.**

### Questionnaires are hard

Lots of room for error

• Lots of time and effort to validate questionnaire

• How do we know the questionnaire is measuring

what we want it to measure?

e.g. What questions should we ask to measure

usability?

• Often no need to re-invent the wheel

• Using pre-existing questionnaires can help

### Pre-existing questionnaires

* Previously shown to be reliable and valid

• Can compare your results with others

• You’ve already seen one two weeks ago.

### NASA-TLX Questionnaire

Captures 6 subscales:

○ Mental Demand

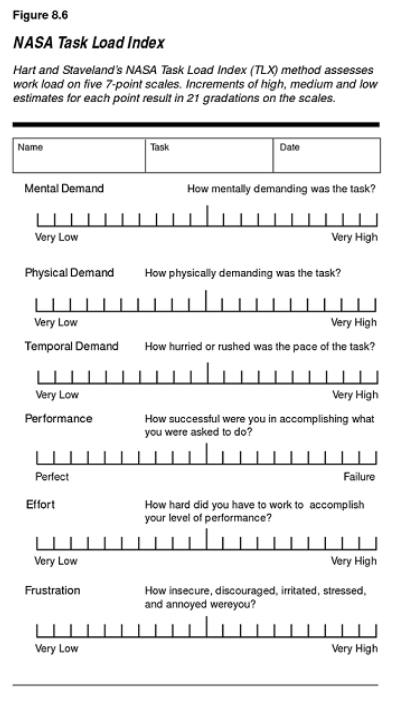
○ Physical Demand

○ Temporal Demand

○ Performance

○ Effort

○ Frustration



System Usability Scale (SUS)Created by John Brooke in 1986

• Originally created as a “quick and dirty” scale for

administering after usability tests

• Technology agnostic

– Can be used for any interactive system

– This includes computers, microwaves, mobile apps, and

prototypes.

• Industry standard

– Reliable

– Valid

– Actually not really dirty

### SUS – 10 questions, 5 point Likert

1. I think that I would like to use this system frequently.

2. I found the system unnecessarily complex.

3. I thought the system was easy to use.

4. I think that I would need the support of a technical person to be

able to use this system.

5. I found the various functions in this system were well integrated.

6. I thought there was too much inconsistency in this system.

7. I would imagine that most people would learn to use this system

very quickly.

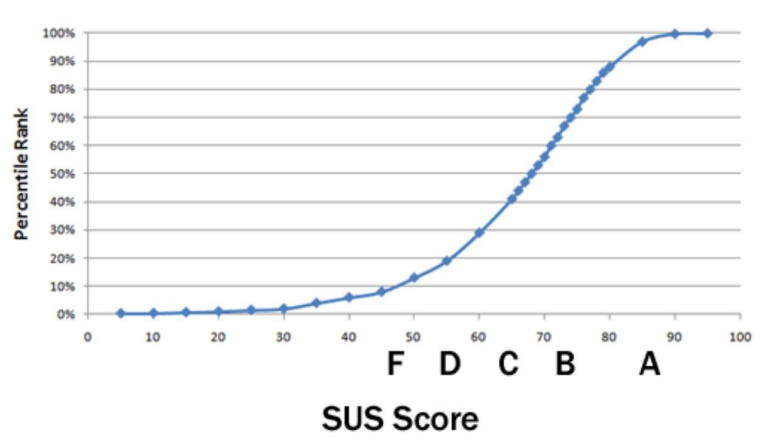
8. I found the system very cumbersome to use.

9. I felt very confident using the system.

10.I needed to learn a lot of things before I could get going with this

system.

### What is a good SUS Score?



### Player Experience Inventory (PXI)

Measures player enjoyment.

• Captures 12 subscales:

○ Enjoyment

○ Mastery

○ Curiosity

○ Meaning

○ Immersion

○ Autonomy

○ Feedback

○ Challenge

○ Audio-visual Appeal

○ Ease of Control

○ Clarity of Goals

### Summary – Questionnaires

Questionnaires can give large amounts of data.

• Require careful (and iterative) design

• Use of existing questionnaires

– E.g. SUS, NASA-TLX, PXI etc…

## Measures (2): Observation

* Observing how a person interacts with your system

– Facial expressions

– Verbal comments etc.

– More specific aspects, e.g., how user performs a

gesture on a tablet, or how player interacts with

motion-based game interface

• Challenge: We are no longer interested in quality of

experience, but in quantifiable results – how can we

record data appropriately?

Step 1: Defining quantifiable observations

• Example: Observing how a player interacts with a

gesture-based Kinect game

– Define characteristics of expected behaviour and/or

events

– Turn into observable chunks

• Basically, operationalizing observable aspects of

Interaction

Step 2: Ensuring that observations are consistent

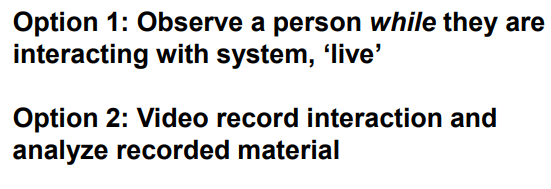
across many participants

– Develop coding scheme for observations based on

your research questions

|  |  |
| --- | --- |
| Event | Number of occurrences |
| Player smiled at partner | 10 |
| Player swore at screen | 6 |
| Player laughed | 3 |

• You need to decide how/when you will make observations:



• Option 1 takes less time, but requires you to know what

you’re looking for in advance

• Option 2 allows you to explore recordings first

### Automated Observation

• With improved technology, we can automate

observation to improve scale and reliability of detection.

• E.g. Natural language processing ➔ textual sentiment

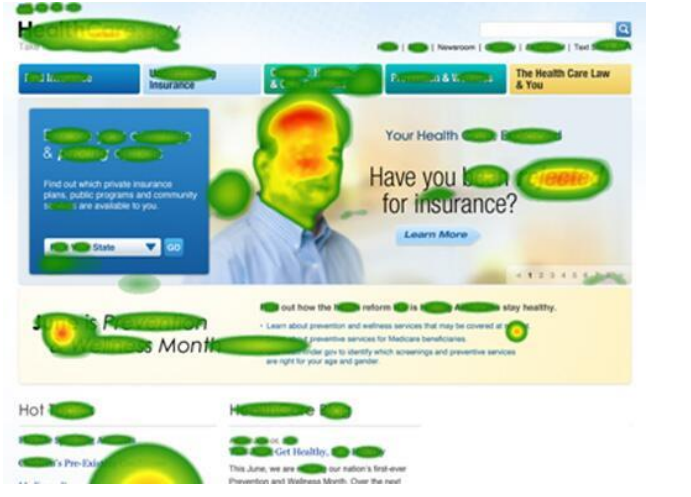
analysis.

• E.g. Computer vision ➔ emotion detection from facial

expression.

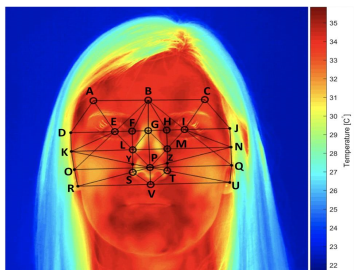
• E.g. Accelerometers ➔ physical activity detection.

**Example: Eyetracking**



## Measures (3): Physiological data

Similar to observation: real-time data

• Can give very accurate and objective measurements

based on users’ physical data.

• Most common examples:

– GSR (Galvanic Skin Response)

– Heart beat

– Brain activity

– Body temperature

## Measures (4): Metrics

• Performance metrics offer objective insights into

how user interacted with system:

– Time taken to complete task(s)

– Number of tasks completed within a set time

– Number of errors made in completing a task

– Number of times website visited

– Number of times help consulted

– Number of successful completions

## Measures (3): Metrics

• Relatively simple to present and analyse this type of

data, but recording requires preparation

• However, do not provide sufficient insights if they are sole

measure – can only tell you what happened, not why

• Good way of backing up questionnaire results and

observations – especially if findings from questionnaires

and observations contradict

### Quantitative Research: Challenge 1

Given a particular question, what data should we collect?

e.g. Does Website A offer better usability than Website B?

• Questionnaire? SUS? NASA-TLX?

• Observation? Measure frequency of frustrated

expressions?

• Metrics? Number of errors committed? Number of times

help is sought?

### Quantitative Research: Challenge 2

• How do we make sense of the data gathered?

• What can we learn about the data and what can we

generalize?

• We’ll be looking at data analysis over the coming 2

weeks.

## Summary

• Quantitative data gives us an account of the magnitude,

amount or size of something.

• Can be obtained through different means:

– Questionnaires

– Observation

– Physiological data

– Performance Metrics

• Next week we’ll look at how to analyse this type of data

# Week 10 – Data Analysis 1

## This Week?

Data Analysis, part I

– Descriptive statistics

• Distributions

• Central tendency (averages)

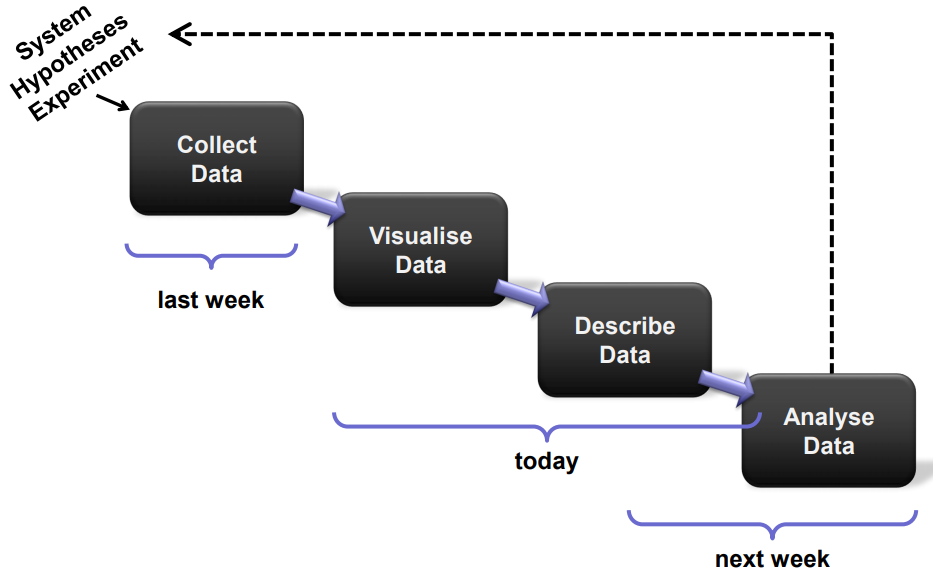
• Spread (standard deviation, CI)

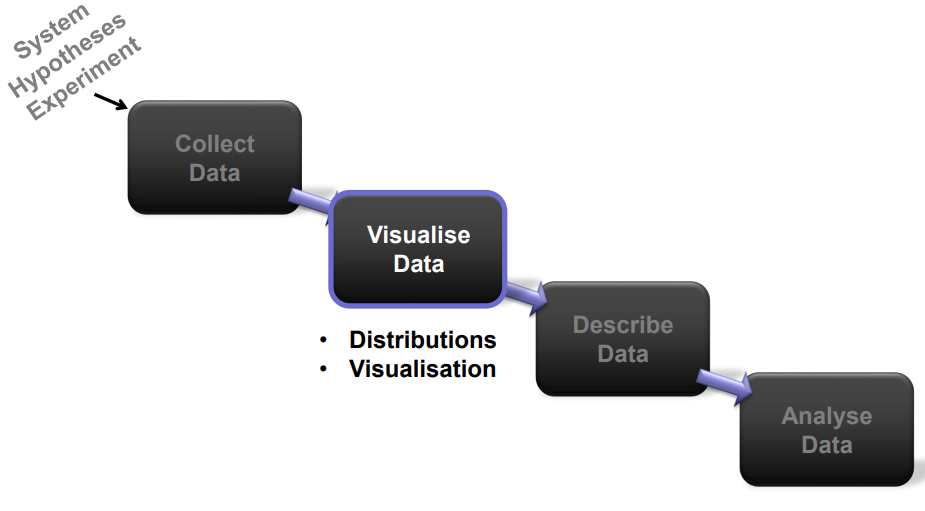
– Discovering relationships

• Scatterplots

• Correlations

## System Hypothesis: Collect Data & Visualise Data





**How do you make sense of thedata that you gathered?**

**How can you describe it in ameaningful way?**

**Imagine you asked people to ratesystem usability using a 5-point Likert item**

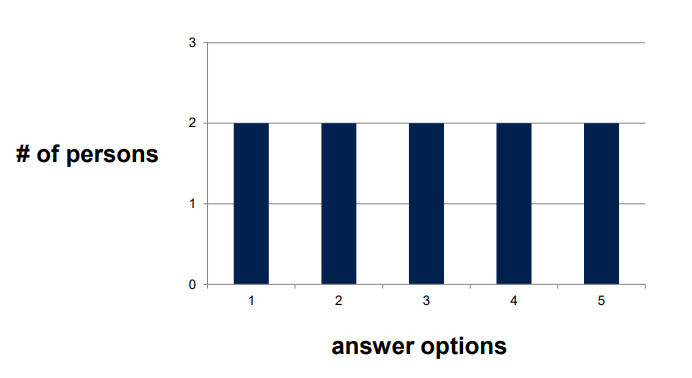
(You would need more than one item to assess usability – this is just for demo purposes!)

**People can choose from fiveoptions, and if you ask severalpeople, that can lead to different distributions of their answers.**

## Distribution A

|  |  |
| --- | --- |
| Participant ID Score | Participant ID Score |
| 1 | 1 |
| 2 | 2 |
| 3 | 2 |
| 4 | 2 |
| 5 | 3 |
| 6 | 3 |
| 7 | 4 |
| 8 | 4 |
| 9 | 5 |
| 10 | 5 |

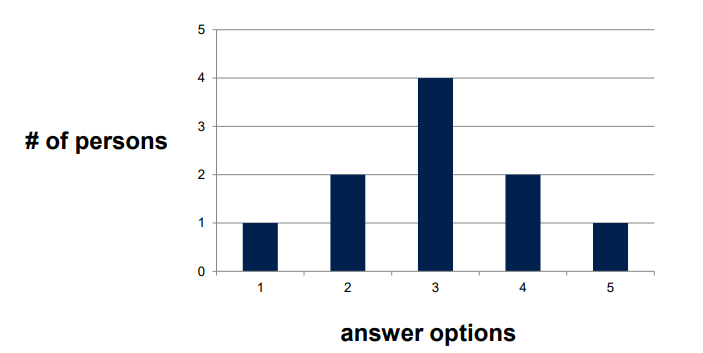
### Distribution: Uniform



## Distribution B

|  |  |
| --- | --- |
| Participant ID Score | Participant ID Score |
| 1 | 1 |
| 2 | 2 |
| 3 | 2 |
| 4 | 3 |
| 5 | 3 |
| 6 | 3 |
| 7 | 3 |
| 8 | 4 |
| 9 | 4 |
| 10 | 5 |

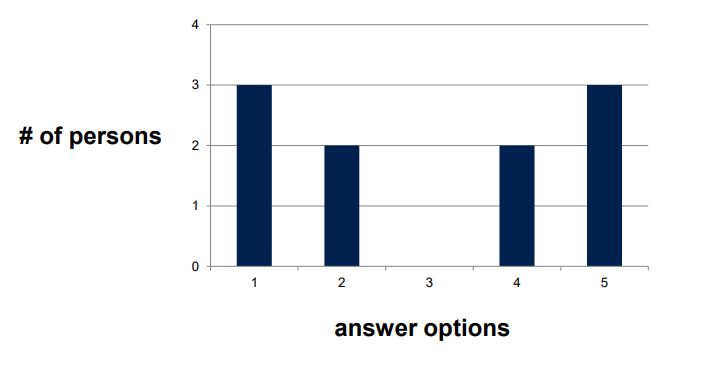
### Distribution: Normal



## Distribution C

|  |  |
| --- | --- |
| Participant ID Score | Participant ID Score |
| 1 | 1 |
| 2 | 1 |
| 3 | 1 |
| 4 | 2 |
| 5 | 2 |
| 6 | 4 |
| 7 | 4 |
| 8 | 5 |
| 9 | 5 |
| 10 | 5 |

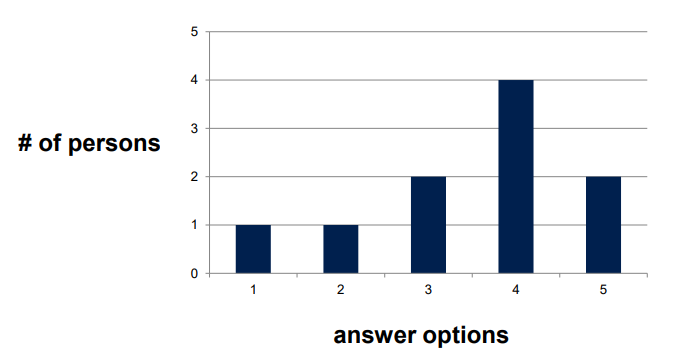
## Distribution: Bimodal

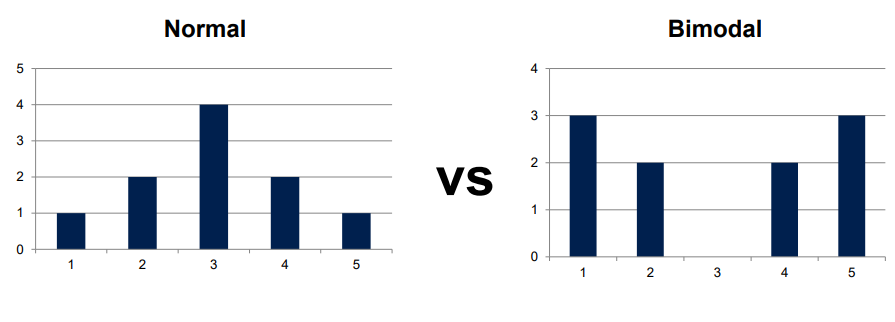


## Distribution D

|  |  |
| --- | --- |
| Participant ID Score | Participant ID Score |
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |
| 4 | 3 |
| 5 | 4 |
| 6 | 4 |
| 7 | 4 |
| 8 | 4 |
| 9 | 5 |
| 10 | 5 |

### Distribution: Skewed





**The average of both distributions is 3.**

The question we asked was “I enjoyed using the system”

(1 = strongly agree, 5 = strongly disagree).

Normal == some people enjoyed it, some didn’t, and most thought it was average.

Bimodal == half of the people really liked it, the other half really didn’t.

**What are the implications for developers?**

**Think game vs. government web service**

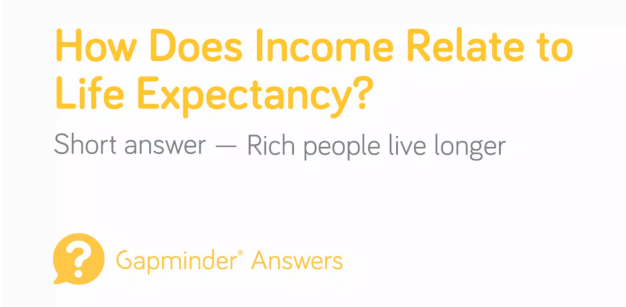
## Descriptive Stats: Distributions

• Distribution of data can give insights & will tell youwhich mathematical operations and statistical tests you can apply

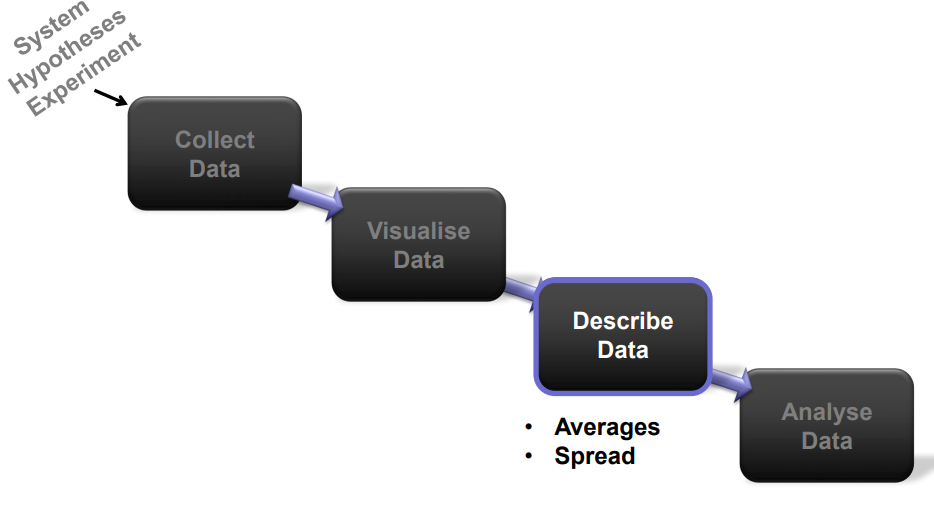
• Uniform, bimodal, normal / Gaussian, and skewed distributions

• Many statistical tests require normal distribution of data – even things like the arithmetic mean!

## An aside: on Visualisation



## System Hypothesis: Describe Data



## Descriptive Stats: Averages

“A value that describes the entire distribution.”

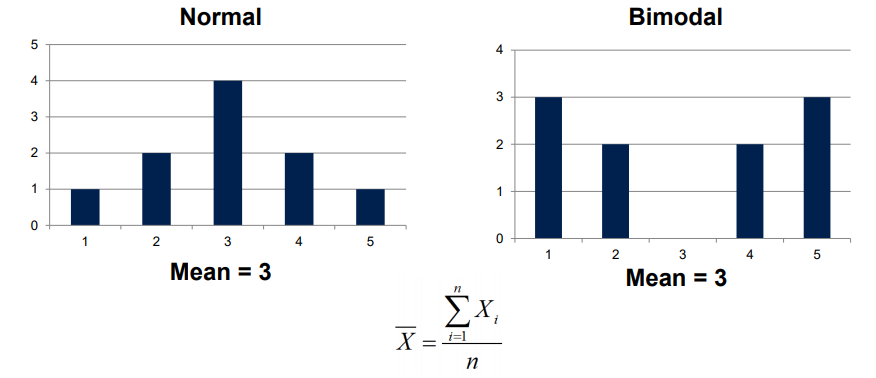
• Mode

• Median

• Arithmetic mean (“average”)

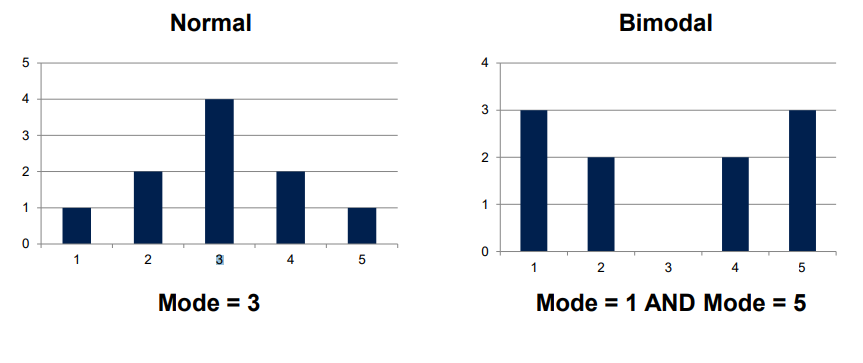
### Averages: Arithmetic Mean

Sum of all measurements divided by number of observations that were made (usually participants)



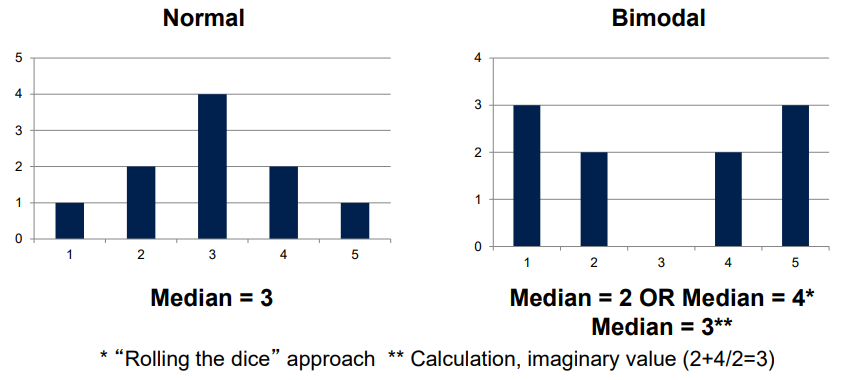
### Averages: Mode

Most frequent value in dataset



### Averages: Median

Value that splits dataset at 50%



## Descriptive Stats: Averages

• “A value that describes the entire distribution.”

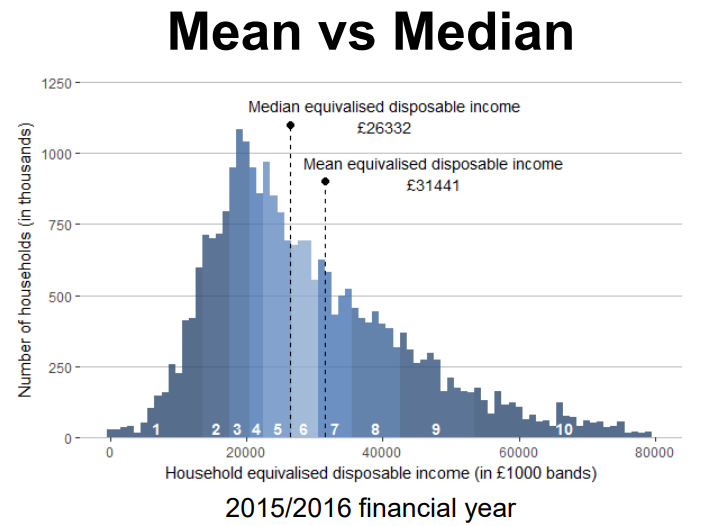
• Mode

• Median

• Arithmetic mean (“average”)

• Depending on distribution, calculation method leads

to different results and is more or less appropriate



### Mode - Non-ordinal data

What month of the year were you born?



Average though is not enough.

Need some way of quantifying this spread so that can improve our model

## Descriptive Stats: Spread

• Reporting the spread of data allows you to give additional information, e.g., outline extremes, and

demonstrate whether your model of choice is a good fit for sample (e.g., whether arithmetic mean

accurately describes data)

• Range

• Deviation: Total Error, Sum of Squared Errors,

Variance, and Standard Deviation

### Spread: Range

Highest and lowest value in a set of observations

|  |  |  |
| --- | --- | --- |
| PID Time Errors | PID Time Errors | PID Time Errors |
| 1 | 33 | 2 |
| 2 | 17 | 4 |
| 3 | 44 | 2 |
| 4 | 37 | 3 |
| 5 | 14 | 3 |
| 6 | 42 | 2 |
| 7 | 13 | 4 |
| 8 | 15 | 2 |

RangeT = 13 to 44

RangeE = 2 to 4

Useful when reporting data where extremes are of interest,

e.g., age, …

### Spread: Deviation

• Is the model (arithmetic mean) a good representation of our data?

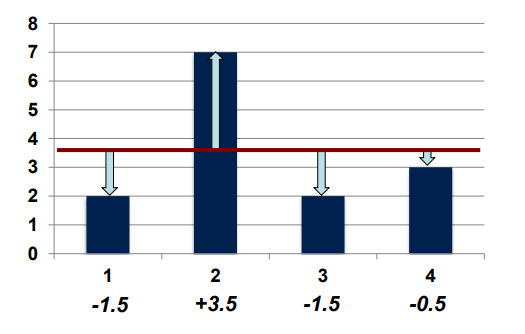
• What is the deviation between model and observed data?

|  |  |  |
| --- | --- | --- |
| PID Time Errors | PID Time Errors | PID Time Errors |
| 1 | 33 | 2 |
| 2 | 17 | 7 |
| 3 | 44 | 2 |
| 4 | 37 | 3 |

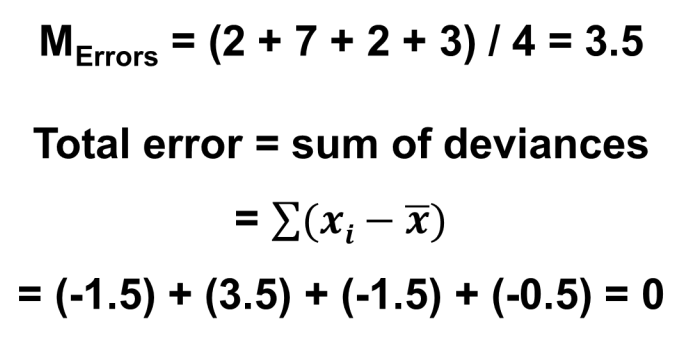
Average # of Errors:

(2 + 7 + 2 + 3) / 4 = 3.5

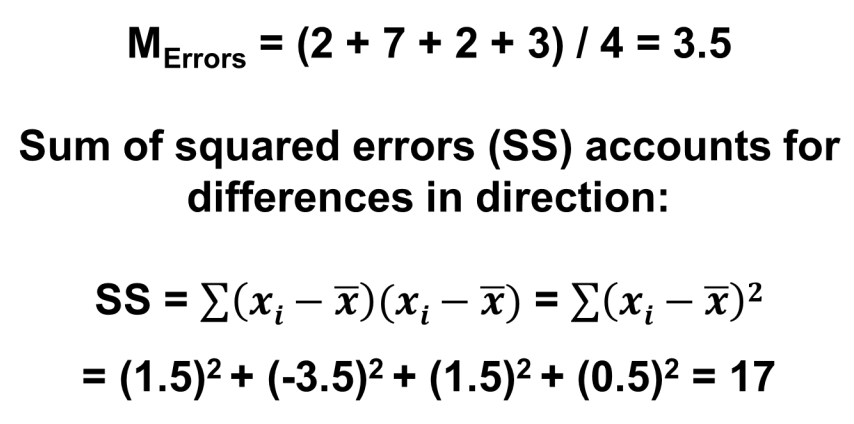
MErrors = (2 + 7 + 2 + 3) / 4 = 3.5



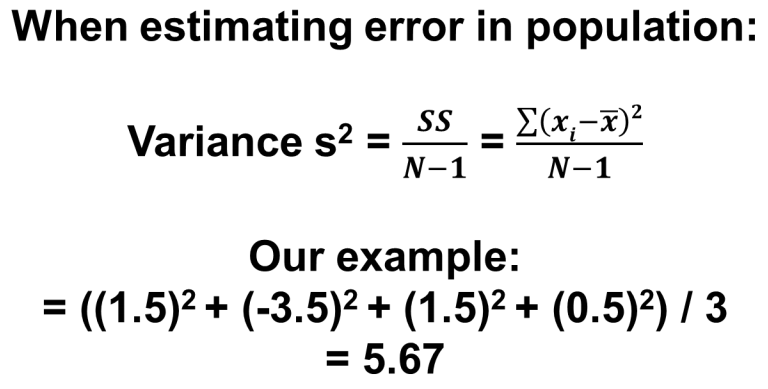
### Spread: Total Error



### Spread: Sum of Squared Errors



### Spread: Variance



When computing variance, divide by:

• N-1 for sample variance (estimate of population…) – the minus one is a correction factor

• N for the population variance (all individuals in the entire population)

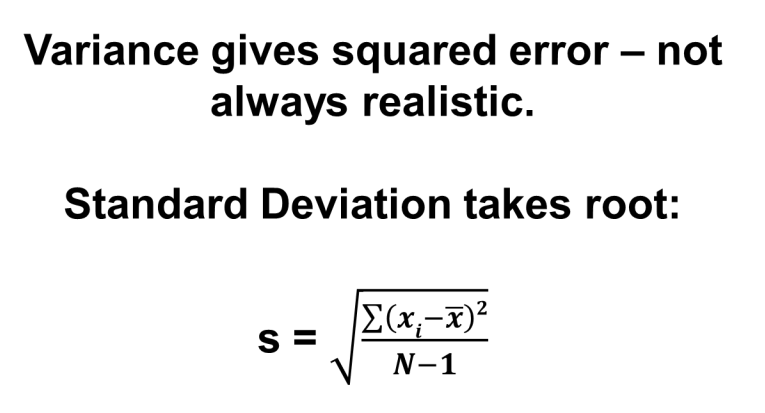
|  |  |  |
| --- | --- | --- |
| PID Time Errors | PID Time Errors | PID Time Errors |
| 1 | 33 | 2 |
| 2 | 17 | 7 |
| 3 | 44 | 2 |
| 4 | 37 | 3 |

Average # of Errors:

(2 + 7 + 2 + 3) / 4 = 3.5

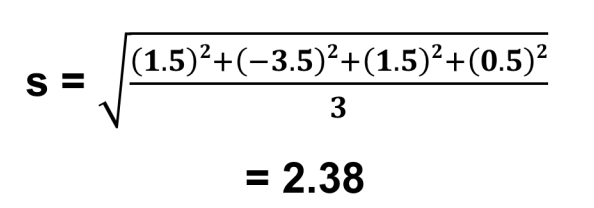
s 2 = ((1.5)2 + (-3.5)2 + (1.5)2 + (0.5)2 ) / 3 = 5.67

### Spread: Standard Deviation



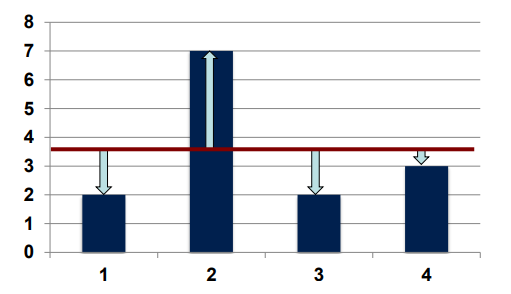
|  |  |  |
| --- | --- | --- |
| PID Time Errors | PID Time Errors | PID Time Errors |
| 1 | 33 | 2 |
| 2 | 17 | 7 |
| 3 | 44 | 2 |
| 4 | 37 | 3 |

Average # of Errors: (2 + 7 + 2 + 3) / 4 = 3.5

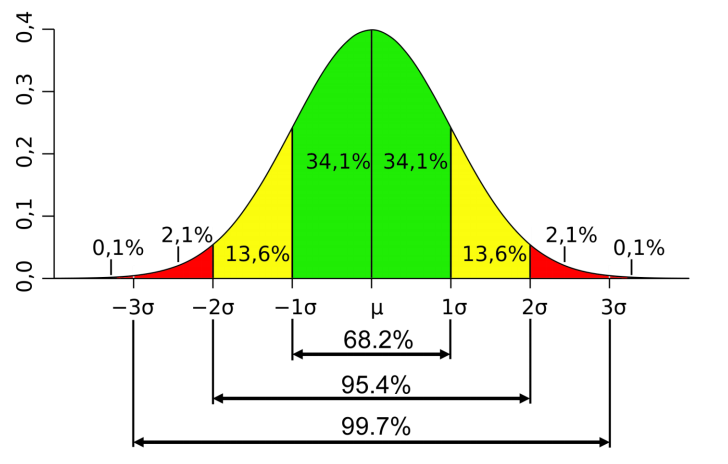


MErrors = (2 + 7 + 2 + 3) / 4 = 3.5

SD = 2.38



### The Normal Distribution



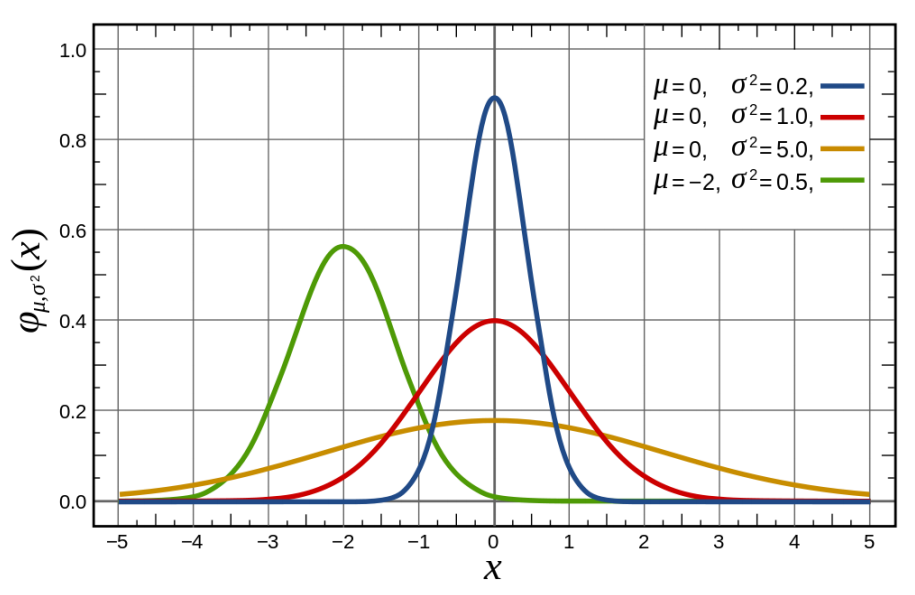
## Descriptive Stats: Spread

• Reporting the spread of data allows you to give additional information

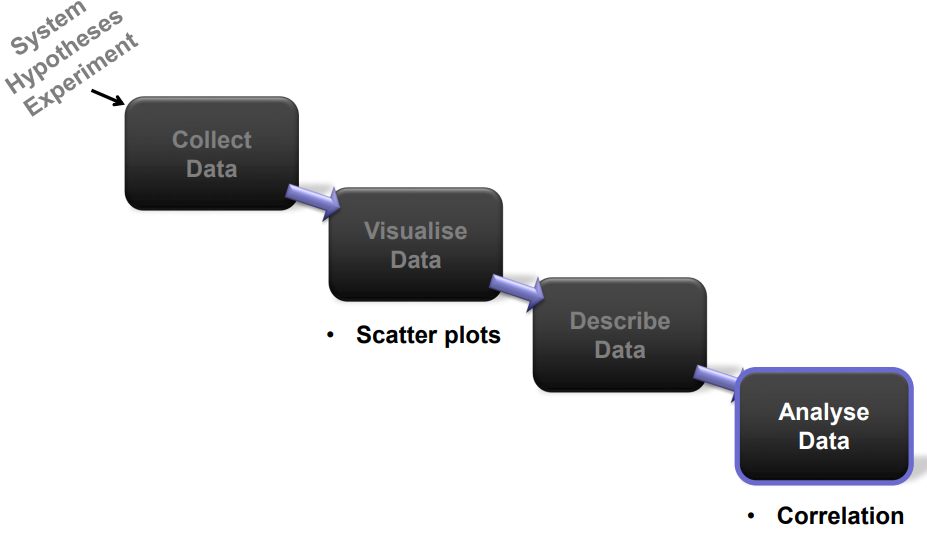
• Range

• Variance

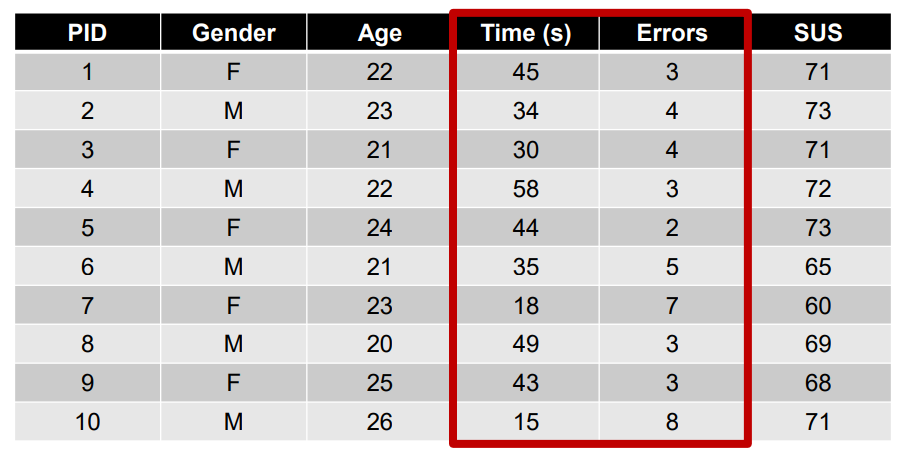
• Standard Deviation



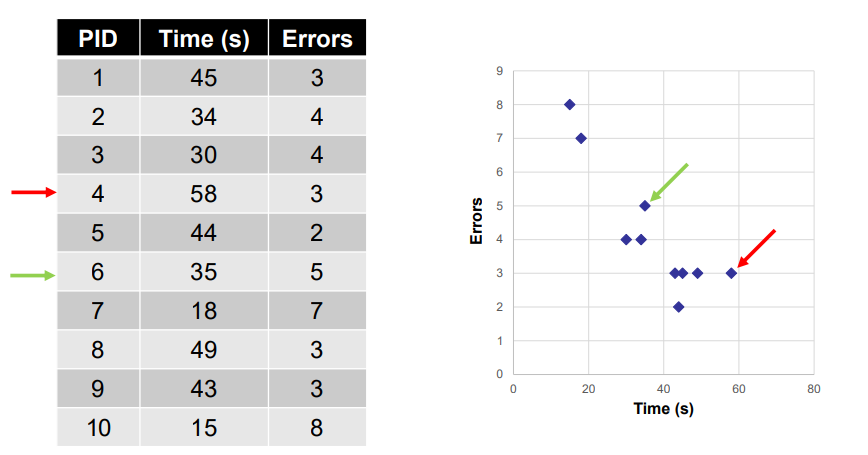
## System Hypotheses: Analyse Data



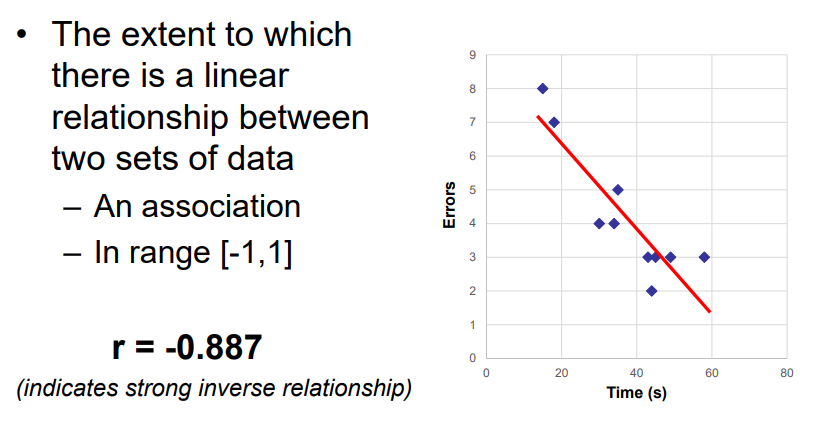
### Sample Data



### Scatterplots



### Correlation



This seems to suggest a strong relationship between time taken and number of errors… …but what if there isn’t one?

This seems to suggest a strong relationship between time taken and number of errors… …but what if there isn’t one?

### Warning: correlation != causation

Just because a (strong) association exists, does not

mean that one variable causes the other

Just because a (strong) association exists, does not mean that one variable causes the other

• Possible alternatives:

– The reverse relationship

– Common third factor causing both

– Bi-directional relationship

– Coincidence!

**The goal of statistical data analysis is to describe data and explore relationships within it.**

**It is important to choose right description and analysis approaches to accurately represent and interpret results.**

## Mixed Methods Approach

• Quantitative analysis ➔ explore, find patterns and generate high-level descriptions of the data.

• Quantitative analysis ➔ helps interpret results, explains why data is the way it is

**Next week:**

Data Analysis II Further Quantitative Data Analysis, Hypothesis Testing, etc…

# Week 11 – Data Analysis 2

## Exam

Essential for the exam:

– Weeks: 1,2,3,9,10,11,12.

– Core concepts:

• Cognition

• Accessibility

• Ethics

• Data Analysis

– Both lectures and workshops will be useful.

## Last Week

**Data Analysis, part I**

– Descriptive statistics

• Distributions

– …with an aside on visualisation…

• Central tendency (averages)

• Spread (standard deviation, CI)

– Discovering relationships

• Scatterplots

• Correlations

## This Week

**Data Analysis, part II**

– Recap: Descriptive stats

– Confidence Intervals

– Hypothesis Testing

• Effect size and significance

• Parametric tests (examples)

• Non-Parametric tests

– Linear Regression

## Sampling

• Simple random sampling

– People selected at randomfrom a known population.

e.g. Telephone polling.

• Convenience sampling

– The sample is selectedbecause they are convenient.

– Downside: Probably not representative of the whole population.

## Descriptive Stats: Central Tendency

• Mode

– Most frequently occurring value

– works on all types of data

• Median

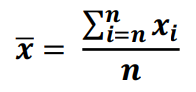
– The middle value (or the ‘virtual’ middle…)

– ordinal data and up

• Arithmetic mean

– The commonly used “average”

– continuous data

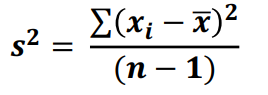


## Descriptive Stats: Spread

• Range

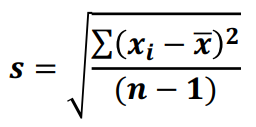
– From min to max

– Ordinal data and up

• Variance

– s2 for sample variance

– Continuous data

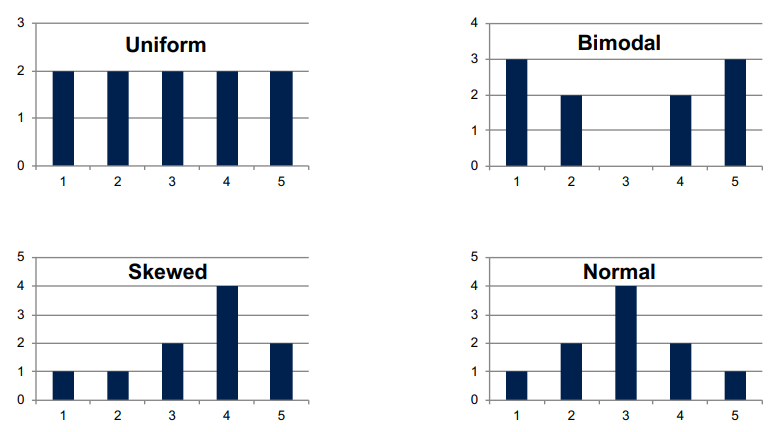
• Standard Deviation

– Has units that are the same

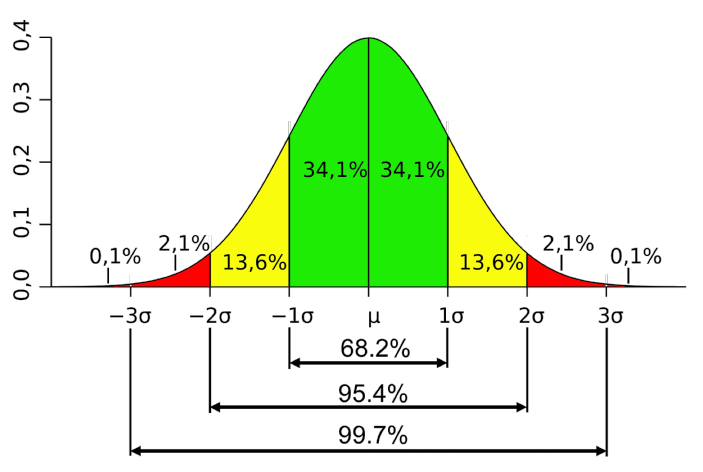
as the measurement

– Continuous data

### Distributions



### Normal Distribution



## An experiment and some data

Let’s say we’ve done an experiment

– E.g. assessing the usability of a new system GUI

– Asking participants a series of questions on a scale of

0-10, and getting and overall mean score (in range [0,10])

– Comparing with a previous version:

• Condition A: the previous version (control)

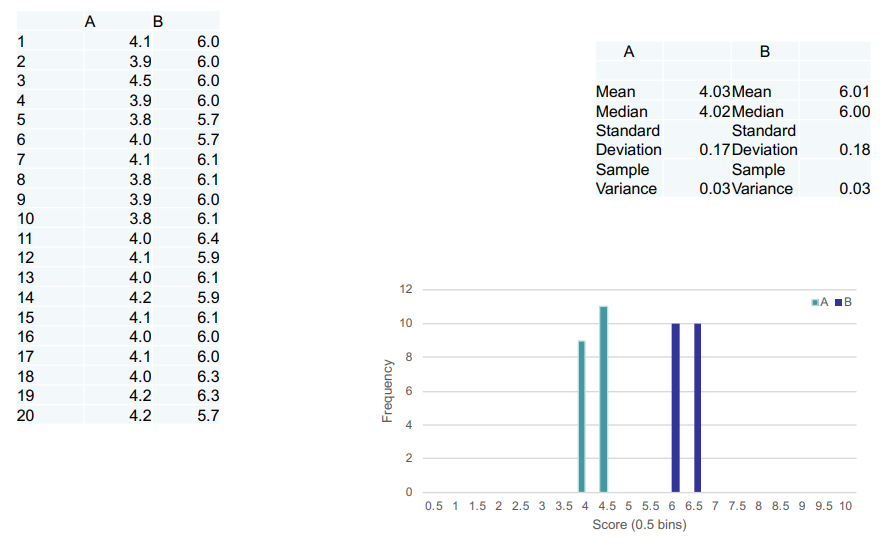
• Condition B: the new version

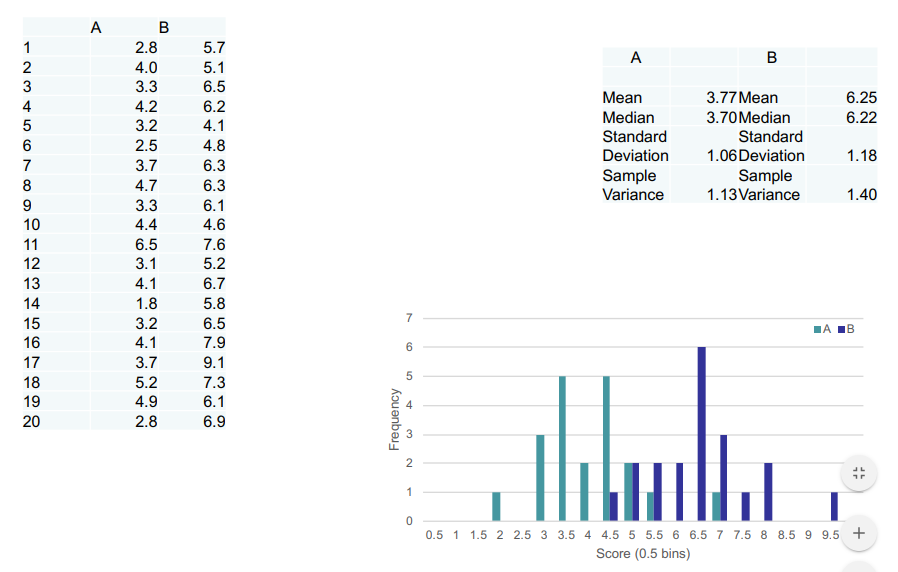
– Say 20 participants per condition

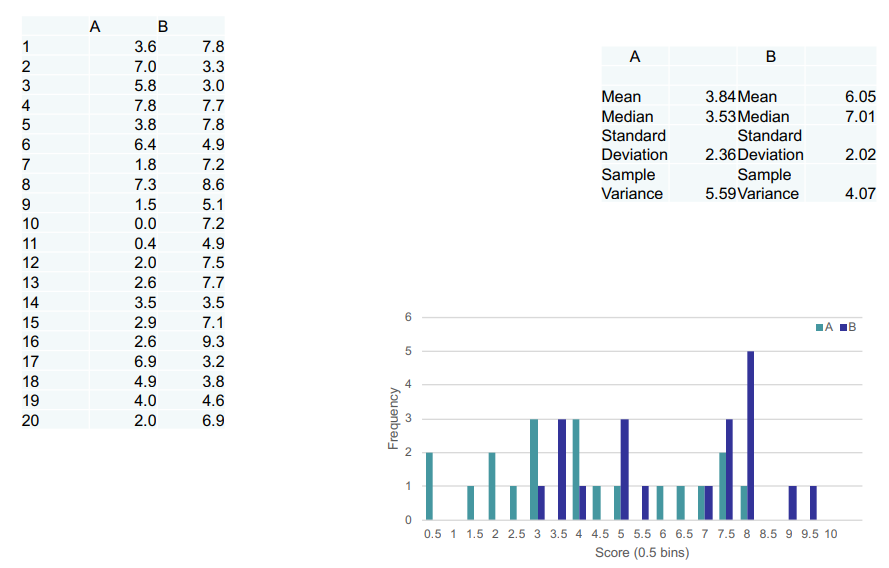
• Randomly sampled from the population

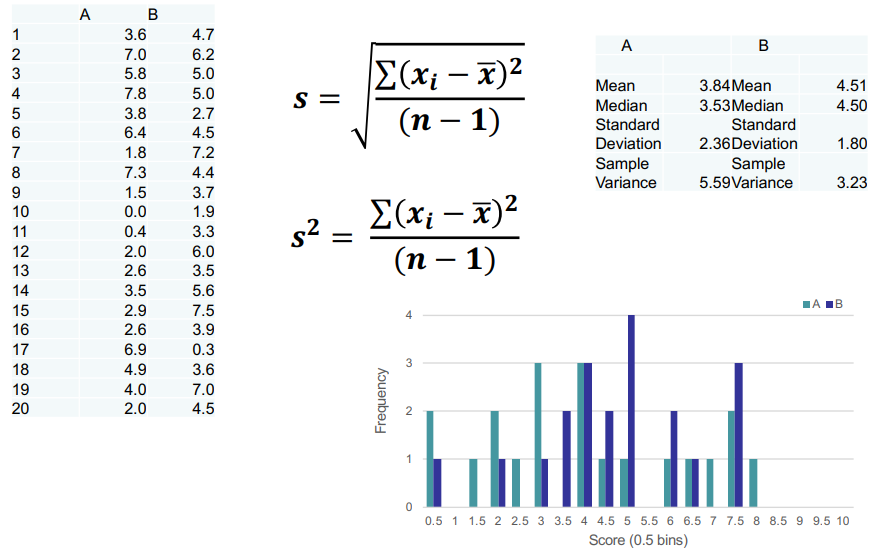
– Assume (for the time being) that our data are

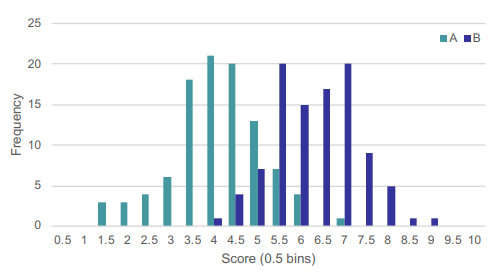
normally distributed…











• Example with 100 subjects per condition

– Means are roughly 4 and 6

– SDs are roughly 1.0

• The normal bell-shaped curve is more apparent

## Effect Size

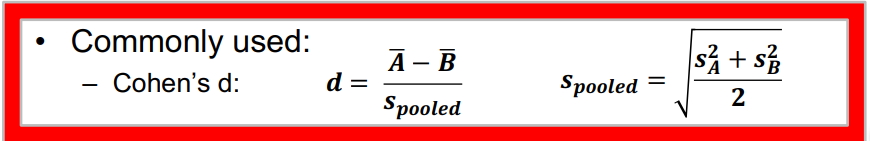
• When comparing two sets of data, want to know how large the effect is

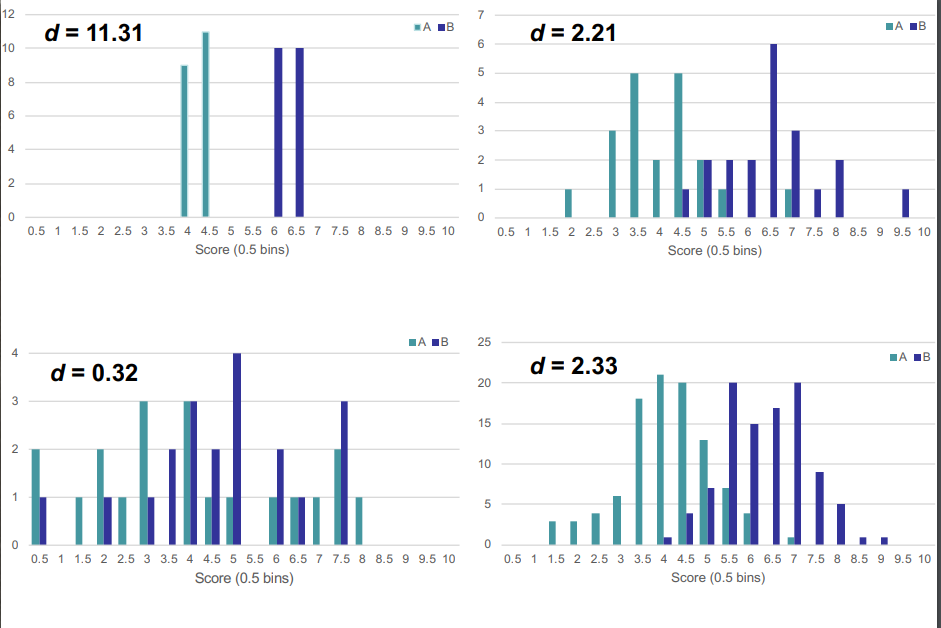
– What is the size of the difference between my two sets of data?

– What impact is this result likely to have?

• Simple:

– The difference:





## Confidence Intervals

• Describe the level of uncertainty in a sample parameter

(estimate of the population parameter)

– E.g. the mean of a sample

• Format CI:

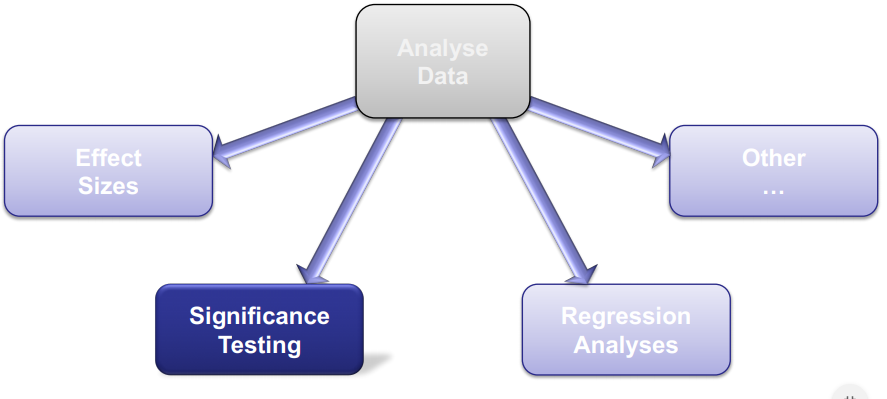
– [min, max]

– Contains the mean

• What affects CI?

– Variation of the population

– The sample size



To describe the data isnot enough…

Want to use our observations to help make decisions about and in the world (or at least our system…)

## Null-Hypothesis

• E.g. Example study assessing the usability of a new

system GUI

– Condition A: Existing interface

– Condition B: Newly developed interface

• Study Hypothesis H1: we hypothesize that Condition B is

better than Condition A.

• Null-Hypothesis H0

– “There is no real difference between our conditions”

– The “null hypothesis”

### Null-Hypothesis Significance Testing

If B is indeed better than A, we say that:

”We reject the null hypothesis H0 and accept H1”

If here is no real difference between our conditions we say

that:

”We reject our hypothesis H1 and accept the null hypothesis

H0”.

### p-value

What tells us if there is a significant difference

between the two conditions?

This is something called the “p-value”

• The p-value tells us how likely it is that the Null

Hypothesis (H0) is true.

• If p<0.05 that means it is likely that H1 is true.

• This chance is the “p-value”

– If p is small enough, then this indicates that the null hypothesis is

not true… and what we have hypothesised is true! B is better

than A

## Choosing the right significance test

• Study design

– Between participants

• Group 1 tested interface A

• Group 2 tested interface B

– Within participants

• All participants tested both A and B

• Study design

– Independent Variables (IV) are the controlled inputs.

• In our study example the IV = 1 because we only

control the interface condition A or B.

• If we also want to see how gender impacts the

results of our study, therefore split participants in

male vs female, then our IV = 2.

– Number of levels of IV

• We compare between two interfaces, A and B

therefore we say we have 2 levels of IV

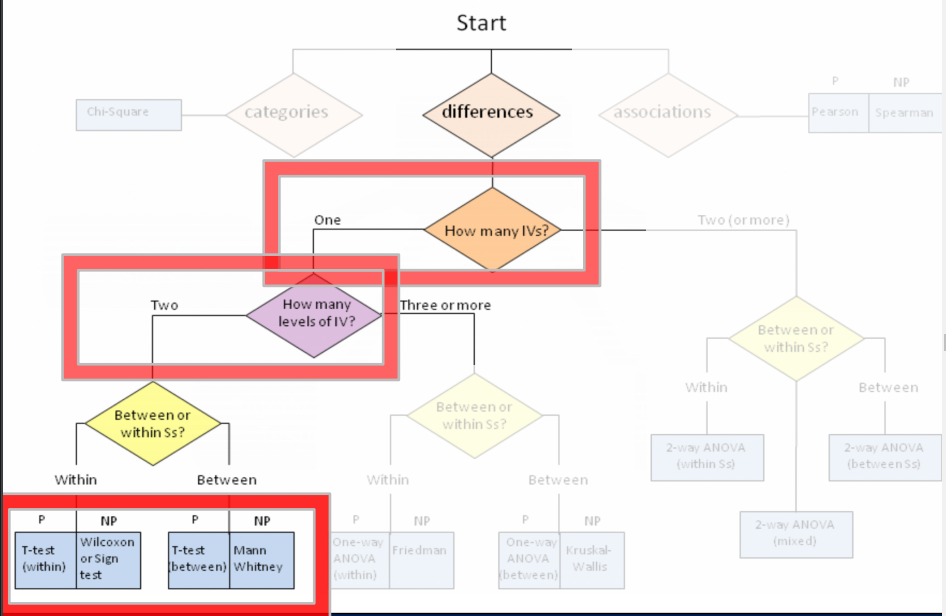
• Type of data type

– Parametric

• Assume data normally distributed

• Continuous data typically is parametric

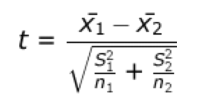
– Non-parametric (do not rely on any distribution)



## The t-test

• Assesses whether the means of two sets of data are

significantly different

– Given the previous assumptions…

– The most common test used during A/B testing

– Important to know about!

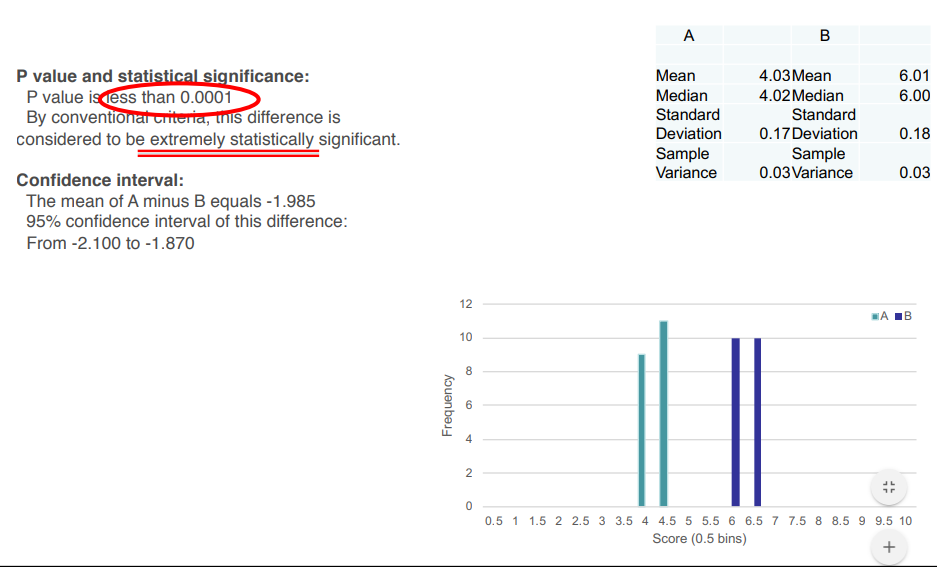
• Calculation of a t-statistic

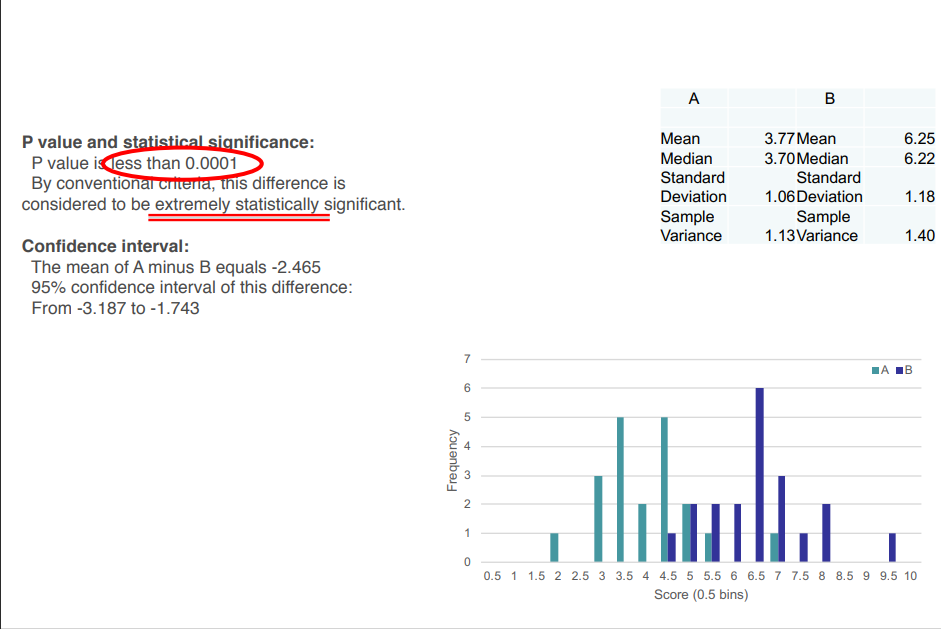
– A ratio: the difference in means over

the variability of the groups.

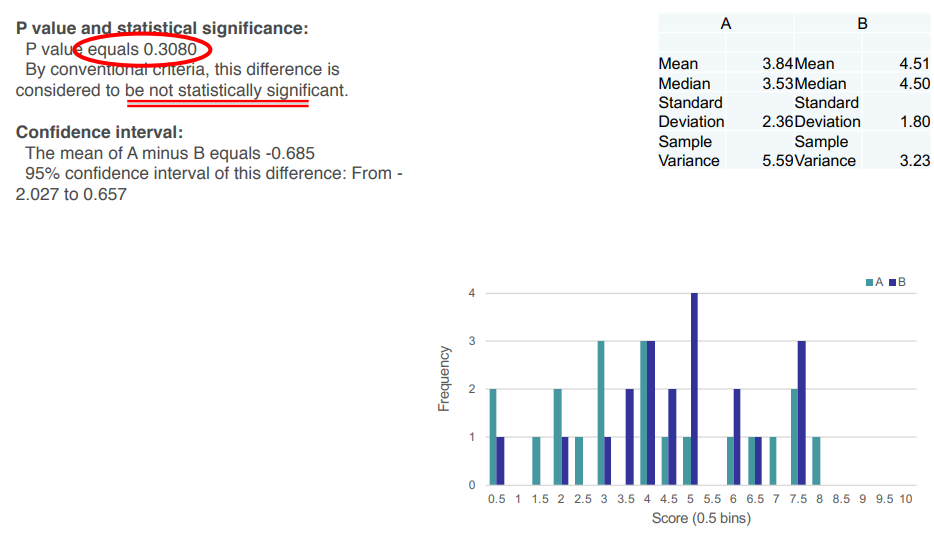
– Will output the actual t score which will tell us the p value.

**Let’s test the significance of the different data distributions we saw earlier…**









**But what if your data violates the assumptions for parametric testing? …non-parametric tests!**

## Assumptions

• Typically aim to use a parametric test, unless the

assumptions are specifically violated

– First look at parametric, then non-parametric

– More statistical power (greater sensitivity to significant

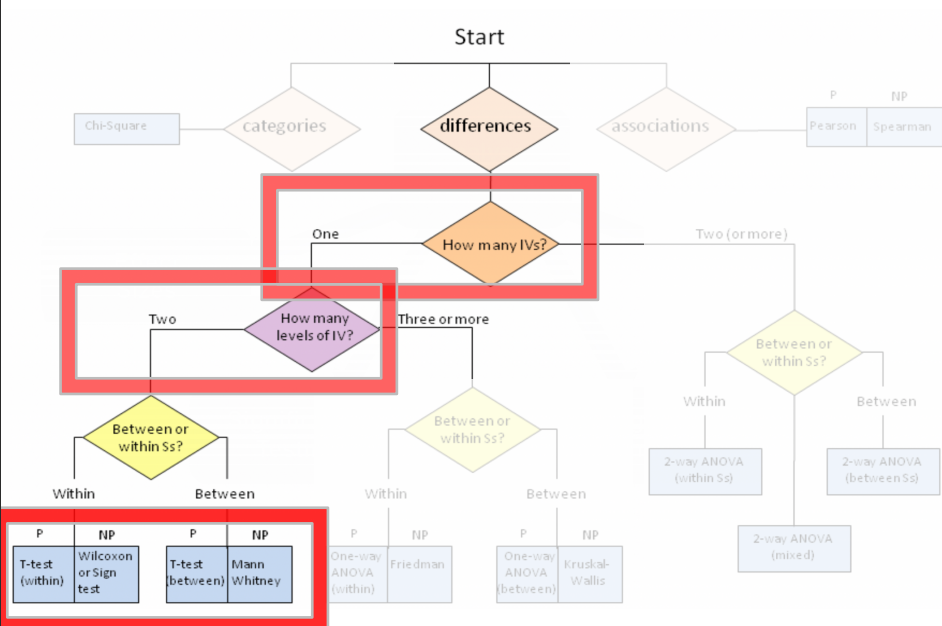
effects)

• Non-parametric tests

– Don’t assume that data distributed a particular way

(normality…)

– If you have a very small sample size…



**Whether parametric or nonparametric, you will get a p-value out that you can interpret as previously discussed…**

### Limitations of p-values

• The p-value says nothing about the magnitude of the

effect

– A p-value of 0.001 does not imply a stronger effect than a pvalue of 0.05…

– As a result: significant != important

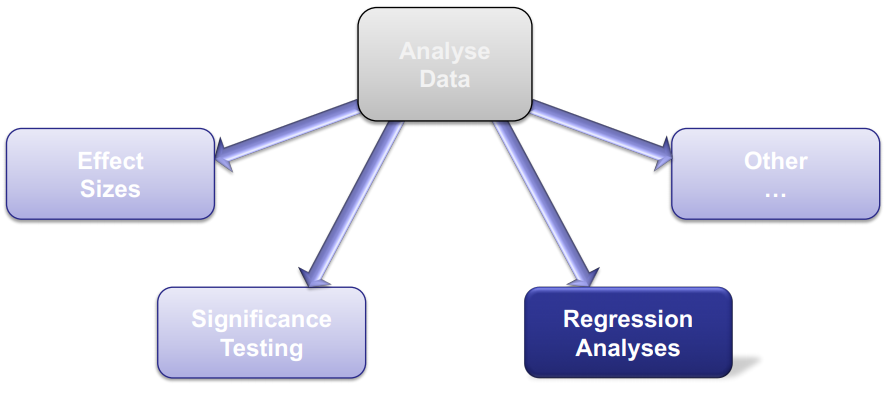
– Need separate indication of this – covered previously!

• P-values are not reliable indicators of replicability

– Confidence Intervals suggested as a way of providing better

information about replication

– See e.g. <http://pps.sagepub.com/content/3/4/286.short>



## Recap: Correlation

• From last week…

• The extent to which there is a linear relationship between

two sets of data

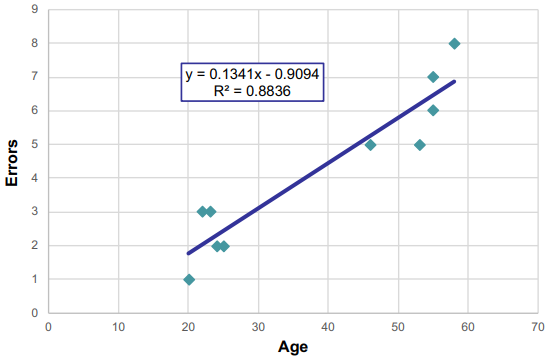
– An association

– In range [-1,1]

• “r” value indicates the amount of variability that is

accounted for by one set of data in the other

– “r2” is the percentage of related variance (the goodness of fit)



• Some Data

– Age (x) vs Errors (y)

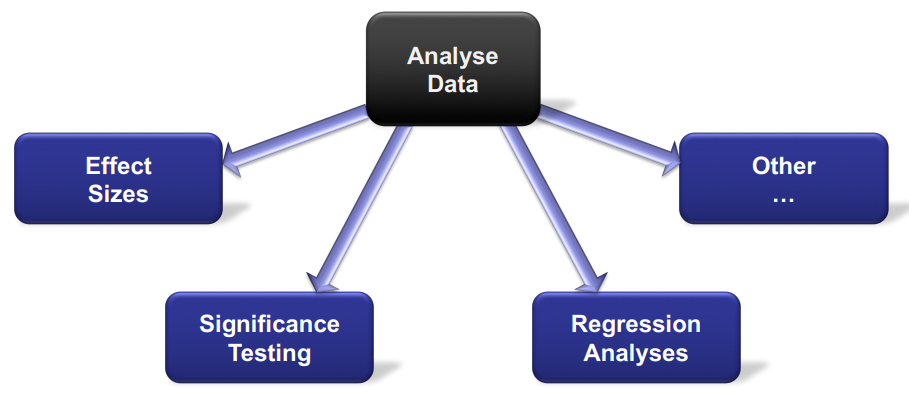
• Using linear model to predict errors from age (not other way around)

– Prediction of values not directly observed

– Does not imply causal link… (same as correlation)

• Can test significance of the line slope (null hypothesis: slope is zero)

– Can use a t-test for this



## Stats Tools

• Statistics software packages that automate many tests, descriptive

stats, and visualisations

• SPSS

– Comprehensive (http://www.ibm.com/analytics/us/en/technology/spss/)

– Or the free version PSPP (https://www.gnu.org/software/pspp/)

• R

– Free and comprehensive (https://www.r-project.org/)

• Built in “Data Analysis” extension of MS Excel

– Covers the basics, but more limited

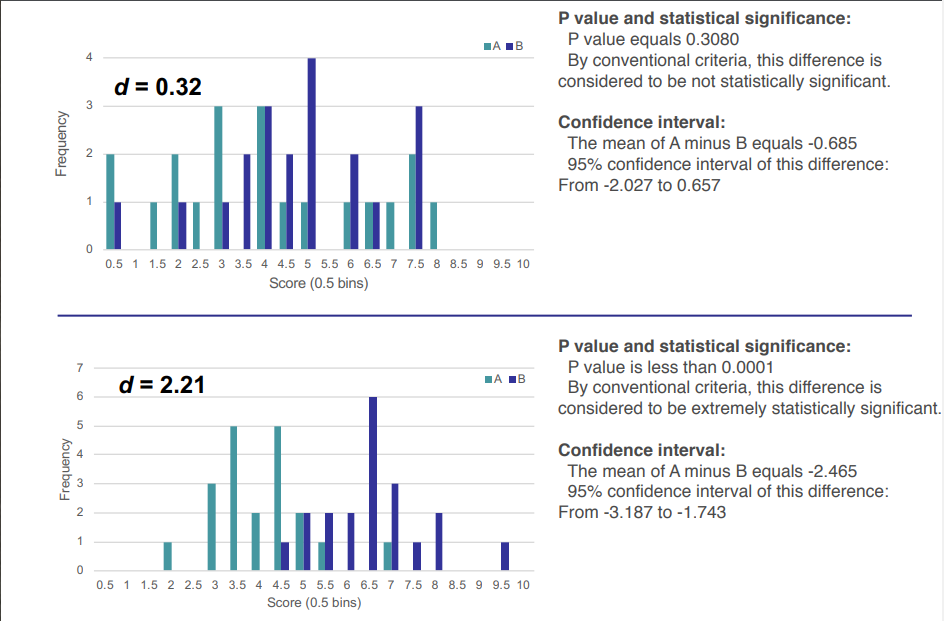
• Website-based tools

– E.g. <http://www.graphpad.com/quickcalcs/>

**The goal of statistical data analysis is to describe data and explore relationships within it**

**It is important to choose the right description and analysis approaches to accurately represent and interpret results**

**In this way, you can best use your results with respect to the initial experiment/hypotheses, and to the system you were evaluating… …so that you can better improve this system**



# Week 12 - Ethics of Human Research

## Topic

• Research Ethics in HCI:

– History and Background

– Principles

– UK legislation

• Ethics polling

## Research Ethics Today

### Principles of Research Ethics

1. Informed and Voluntary Consent

2. Favourable Risk-Benefit Ratio

3. Confidentiality and Data Protection

4. Independent Review Process for Research

5. Justice and Inclusiveness

### When do they apply?

• You are conducting research

An undertaking intended to extend knowledge

(establish facts, reach new conclusions) through

disciplined inquiry or systematic investigation.

• Any activity that isn’t “research” would not be subject

to research ethics or an ethical review.

• Research ethics only apply to studies that involve

any one or more of:

– human participants in an active capacity

– the recording of personal data

– the collection of biological tissue

– animal participants

Most HCI Research

### Principles of Research Ethics

**1. Informed and Voluntary Consent**

2. Favourable Risk-Benefit Ratio

3. Confidentiality and Data Protection

4. Independent Review Process for Research

5. Justice and Inclusiveness

#### Informed and Voluntary Consent

Participant is given all the information needed to make a decision on whether or not to take part in the research.

Researcher's identity - who is doing this research? Contact

details.

Purpose - why is this research taking place?

Method - what does participation in this study entail? What

are the potential risks, benefits and consequences of

taking part?

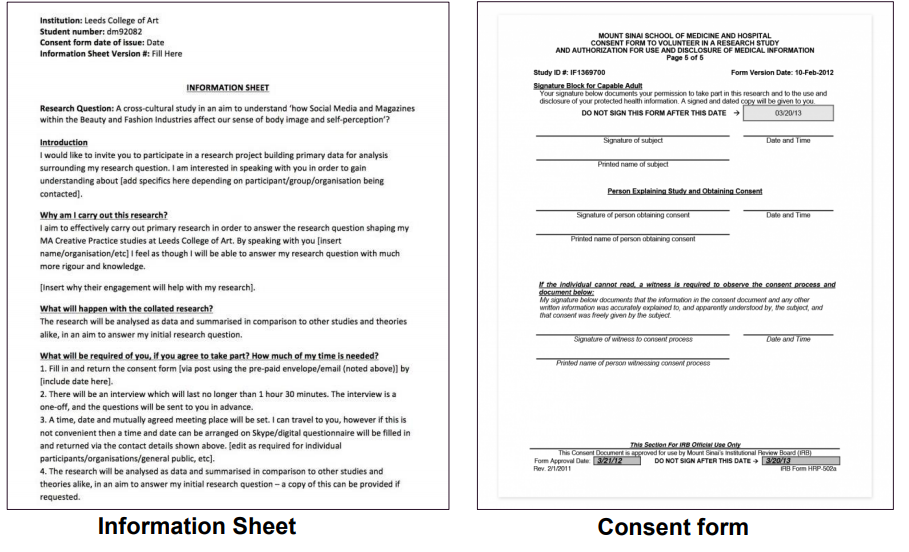
Data use – what data is being collected? How is the data going to be used? Who will have access to it? Will it be published? How and where?



Informed also means understood.

Short, simple and clear.

Ensure participant has read them



##### Voluntary Participation:

• Participants chooses to take part willingly.

• Participant knows they are free to participate.

**Right to withdraw:**

• No penalty (material or otherwise) for withdrawal.

• If participant chooses to withdraw, the study stops

immediately and all data collected is deleted.

• This can happen AFTER the study terminates.

– That’s why contact details are important!

• Participant knows they can withdraw consent at any

time.

##### Coercion: Exerting undue influence in order to incentivize

participation.

• Physical: threat of violence or harm (unlikely)

• Social: peer pressure, social pressure e.g. calling in a

favour

• Institutional: teacher/student, boss/employee

• Financial: offering so much money that participants take

part in contradiction to their values or self-interest

• Indirect Coercion: allowing another person to exert

influence on your behalf.

##### Exception: Gaining explicit consent is impractical

• E.g. Twitter Sentiment

Analysis study.

• Over 200,000 tweets

analysed.

• Impossible to get consent

from individual users.

##### Exception: Covert Research

When the researcher conceals, in full or in part, the nature of the research

activity.

e.g.: lie about who you are, what you are doing, why you are doing it, what you are recording… Or simply not tell the participant anything.

Only acceptable when:

• Overt research is not possible

• The study is of scientificsignificance.

NOT when knowing the true nature of

the research would stop people from

taking part!

You should, if possible, reveal to the

participant afterwards, and uphold

their right to withdraw.

Covert research may be undertaken when it may provide unique forms of evidence or where overt observation might alter the phenomenon being studied. The broad principle should be that covert research must not be undertaken lightly or routinely. It is only justified if important issues are being addressed and if matters of social significance which cannot be uncovered in other ways are likely to be discovered.’

ESRC Framework for Research Ethics, 2010

##### Exceptions should be made with care

Particular care must be taken

during this type of study.

If gaining consent is

impossible it does not

automatically mean the study

is ethical without consent.

Highlights need for ethical

approval procedures.

#### Favourable Risk-Benefit Ratio

Do no harm vs.

Minimize risk - Maximize benefit

##### Maximize Benefit

• The aim of the study should be beneficial.

• The benefit must reasonably follow from prior

scientific knowledge.

A research study is unethical if:

• It does not have a solid scientific foundation

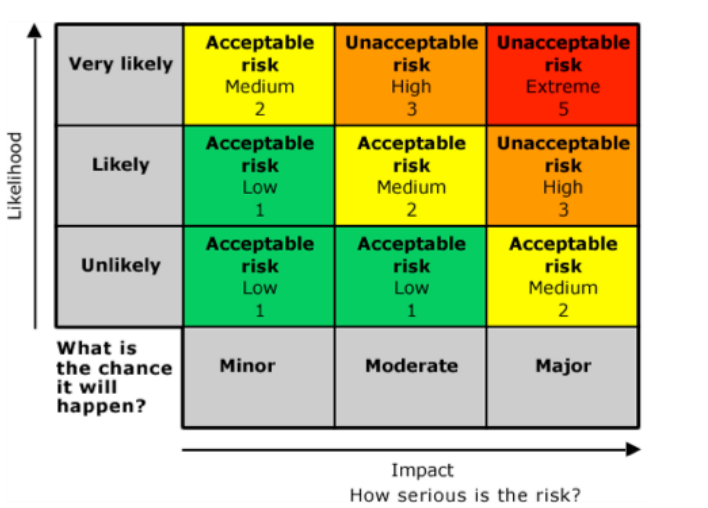
• Is not run by a trained competent professional

• It does not follow well-defined procedures to ensure

success.

… regardless of the level of risk it creates.

##### Minimize Risk



**What are potential risks in an HCI study?**

Physical?

Emotional/Psychological?

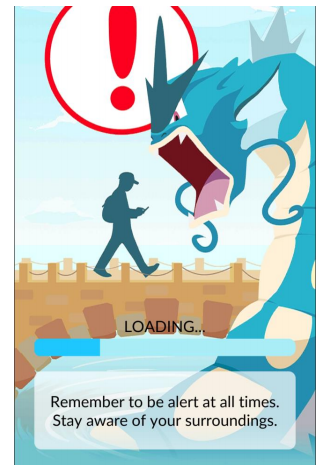
Social?

Financial?

Professional?

Not just to participants… To researchers as well.

**Location-based services/games create particular risks for testing.**



**How would you run a study to test the usability features of a SatNav?**

****

**The most prevalent HCI-related risk: Privacy!**



#### Confidentiality and Data Protection

• Participants in research study should be guaranteed

confidentiality.

– Particularly important for studies involving sensitive or

stigmatized topics

– Even the fact that someone took part might be

harmful to them if made public

• Personal data should be kept private and safe.

**Personal Data**

‘personal data’ means any information relating to an identified or identifiable natural person; an identifiable natural person is one who can be identified, directly or indirectly

-- General Data Protection Regulation (GDPR), May 2018

**directly:** the person can be identified from the information. indirectly: the person can be identified from that information in combination with other information.

**indirectly:** the person can be identified from that information in combination with other information.

Examples of Personal Data

• Name, photograph, video, audio, email address,

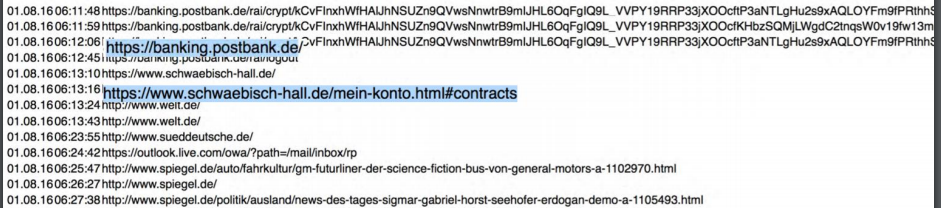
phone number, fingerprint

But also…

• [Age, Gender, Postcode, Occupation] together can

be enough to identify a person.

##### De-Anonymization/Re-identification



Other examples include:

• Identifying people from their Netflix movie ratings

Narayanan, Arvind, and Vitaly Shmatikov. "How to break anonymity of the netflix prize dataset." arXiv preprint cs/0610105 (2006).

• Identifying someone’s entire location history based on 4 known locations

Montjoye, Yves-Alexandre de, César A. Hidalgo, Michel Verleysen and Vincent D. Blondel. “Unique in the Crowd: The privacy bounds of human mobility.” Scientific reports(2013).

• Identifying someone’s surname from their DNAsequences

Gymrek, M., McGuire, A. L., Golan, D., Halperin, E. & Erlich, Y. “Identifying personal genomes by surname inference.” Science 339, 321–324 (2013).

##### Current Legal Approaches

• Difficult to enforce

• Stifles security research

• Leaves vulnerabilities undiscovered



##### Sensitive Personal Data

Especially high level of care should be given to data related to a person’s:

• Racial or ethnic origin

• Political opinions

• Religious beliefs

• Trade union activities

• Physical or mental health

• Sexual life

• Information about criminal offences

##### Best Practice: Data Protection

• Minimise personal data by keeping any necessary identifying information separate from the rest of your data.

• Use numeric index to link identifying with nonidentifying data.

But…

• You must not fully anonymize your data

• You must still ensure your participants’ right to

withdraw their data from your study

##### Best Practice: Data Retention

• Keep data only as long as it is required

– Funding bodies have specific requirements

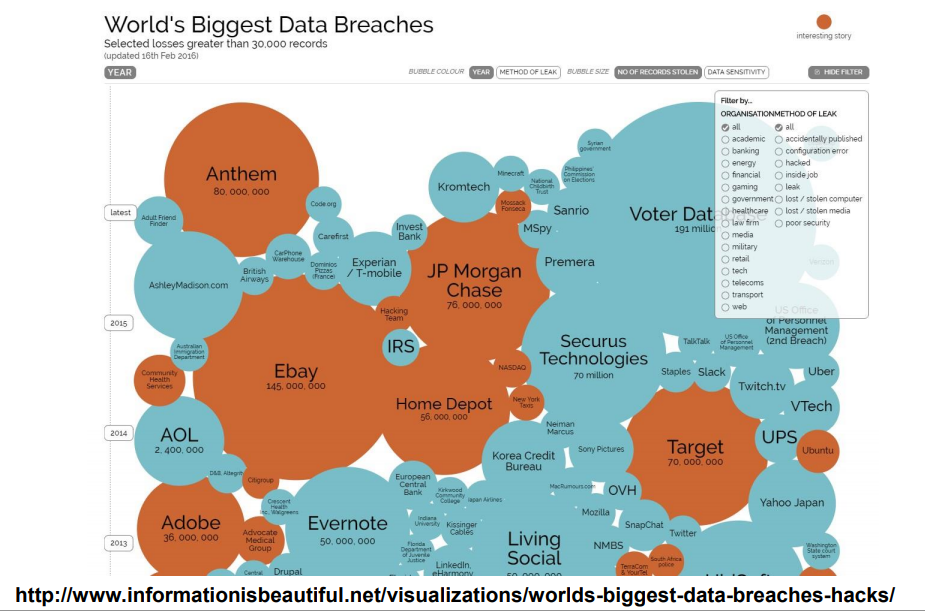
• Keep personal data secure:

– Digital data should be on a password protected server

– Physical data should be in a locked container

– Sensitive personal data should NEVER be left on a laptop/mobile device.







##### Best Practice: Data Collection

• Only collect data you need.

• Avoid boilerplate questionnaires.

• Never ask for sensitive personal data unless it is vital to your research.

e.g. If your study is about how people from different parts of the city experience Pokemon Go, you can ask for their postcode.

If your study is about how much time they spend on Facebook, postcode is probably not relevant.

##### Exceptions: Voluntary

• Participants may wish to be fully identified

– Expert interviews

– Participatory design

• Participants may partly waiver their confidentiality

– For publishing photos/video excerpts

– Explicit and separate consent must be sought

##### Exceptions: Statutory

• You are legally required to report:

– Evidence of crime/intent to commit crime

– Evidence of risk to vulnerable persons

Additionally, research data is not protected from court

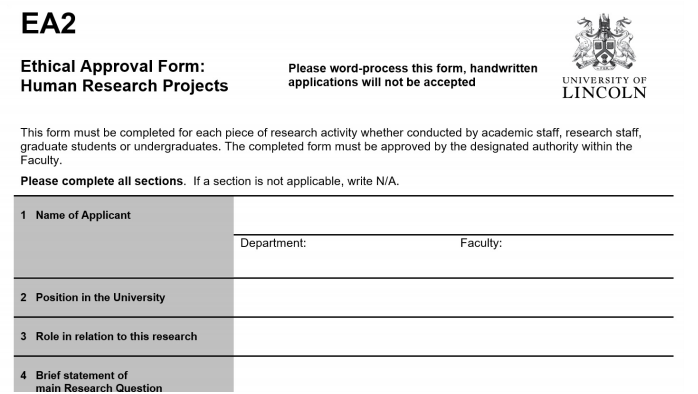
orders.

In cases where this may be relevant, participants should

be aware of this.

#### Independent Review Process for Research

Decision on whether or not research is ethical should NOT be left to the researcher to decide.



##### Research Ethics Committee

In the UK, a Research Ethics Committee (REC) is required in any institution that takes government

funding.

Similar requirements in other countries.

e.g. Institutional Review Board (IRB) in the US.

#### 5. Justice and Inclusiveness

• Justice: The benefits and the burdens of the

research must be similarly distributed.

• Inclusiveness: No group of people should be,

without good reason, excluded from research

activity that affects them.

##### Justice

• If your study will likely benefit a certain portion of the population, this should be reflected in the participants you use.

– e.g. The Tuskegee Syphilis experiment

• Participants were exclusively impoverished African-Americans, despite the benefits being for the general population.

E.g.



##### Inclusiveness

• You need to have a good reason to exclude someone from participating in your study.

– e.g. you can exclude young participants in a study on a new support systems for the elderly

– e.g. you cannot exclude the elderly in a study on a new fitness training monitor

• Pregnant women, elderly, children are often excluded from a study without good reason.

### Overview

**Principles of Research Ethics**

**1. Informed and Voluntary Consent**

**2. Favourable Risk-Benefit Ratio**

**3. Confidentiality and Data Protection**

**4. Independent Review Process for Research**

**5. Justice and Inclusiveness**