SEMINARS IN ARTIFICIAL INTELLIGENCE

HUMAN ROBOT INTERACTION INTRODUCTION

Marc Hanheide

with material from Feil-Seifer, Mataric, Goodrich, Schultz, Breazeal, Kathrin Gerling and many other colleagues

HRI: A SUB-DISCIPLINE OF HCI

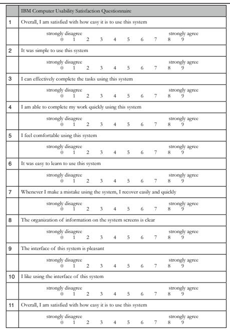
Disclaimer: Prepared with input from an actual HCI researcher: Dr Kathrin Gerling

DISCLAIMER II: HRI == HCI ???

- HRI is **strongly related** to Human-Computer Interaction (HCI) and Human-Machine Interaction (HMI).
- HRI, however, **differs** from both HCI and HMI because it concerns systems (robots) which have **complex**, **dynamic control systems**, which exibit **autonomy** and **cognition** and which operate in changing, **real-world environments**.

HUMAN FACTORS / ERGONOMICS

- Understanding the human element in design of technology
  - Designing equipment and devices that fit the human body and its cognitive abilities
- NASA and IBM pioneered this work
- A lot of the concepts and methods in HCI are borrowed from engineering



	IBM Computer Usability Satisfaction Questionnaire
1	Overall, I am satisfied with how easy it is to use this system
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
2	It was simple to use this system
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
3	I can effectively complete the tasks using this system
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
4	I am able to complete my work quickly using this system
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
5	I feel comfortable using this system
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
6	It was easy to learn to use this system
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
7	Whenever I make a mistake using the system, I recover easily and quickly
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
8	The organization of information on the system screens is clear
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
9	The interface of this system is pleasant
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
10	I like using the interface of this system
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
11	Overall, I am satisfied with how easy it is to use this system
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree

TORS / MICS

design of technology

fit the human body and

HCI are borrowed from

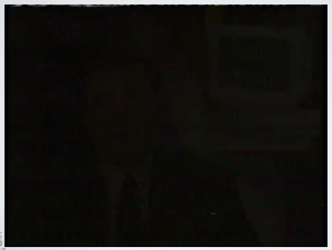
USABILITY & ACCESSIBILITY

- Started with the era of desktop computing, suddenly "usability" becomes a billion dollar industry
- The design of interfaces that allow people to do their work without being massively frustrated
- Based on cognitive psychology – understanding what people are capable of and comfortable with in terms of perception & memory

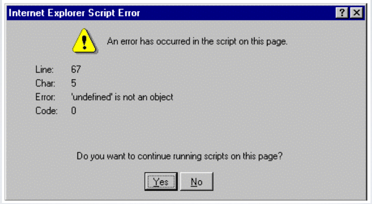
USER EXPERIENCE

- Putting the user at the centre of everything
  - Requirements, design, prototyping, dev & evaluation
  - Cognitive abilities, subjective experience, narratives, cultural impact
- Dialogue is the key - constant, constructive dialogue between designers, users and communities

USABILITY (FROM KATHRIN)



WHAT IS USABILITY?



WHAT IS USABILITY?

- "The effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments." ISO 9241

WHAT IS USABILITY?

- Lots of different ways of considering / measuring this
- It depends on what the software is supposed to do
- Efficiency
  - Time taken to do task
  - Errors
- Effectiveness
  - Number of tasks completed in a given time
- Satisfaction
  - Questionnaires

	IBM Computer Usability Satisfaction Questionnaire
1	Overall, I am satisfied with how easy it is to use this system
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
2	It was simple to use this system
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
3	I can effectively complete the tasks using this system
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
4	I am able to complete my work quickly using this system
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
5	I feel comfortable using this system
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
6	It was easy to learn to use this system
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
7	Whenever I make a mistake using the system, I recover easily and quickly
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
8	The organization of information on the system screens is clear
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
9	The interface of this system is pleasant
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
10	I like using the interface of this system
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree
11	Overall, I am satisfied with how easy it is to use this system
	strongly disagree 0 1 2 3 4 5 6 7 8 9 strongly agree

BILITY?

uring this

d to do

USABILITY

- Nielsen (1993): Usability Engineering
- There are vast amounts of usability research – progress in instruments, procedures, specialised for different user groups / devices
- It is inherent in the design of most software / devices

WHAT ARE USABILITY HEURISTICS?

- Essentially rules-of-thumb that "experts" use to ensure that their software follows established usability principles
- They are used primarily because they have been found useful.
- Applied before software is exposed to real people.
- There are a few different versions
  - Nielsens (1993) 10 usability heuristics
  - Schneidermans (1998) 8 golden rules

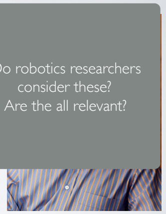
NIELSEN'S 10 USABILITY HEURISTICS

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help recognise, diagnose, and recover from errors
10. Help and documentation

http://www.useit.com/papers/heuristic/heuristic_list.html

NIELSEN'S 10 USABILITY HEURISTICS

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help recognise, diagnose, and recover from errors
10. Help and documentation



Do robotics researchers consider these?
Are the all relevant?

http://www.useit.com/papers/heuristic/heuristic_list.html

SHNEIDERMAN'S 8 GOLDEN RULES

1. Strive for consistency.
2. Enable frequent users to use shortcuts.
3. Offer informative feedback.
4. Design dialog to yield closure.
5. Offer simple error handling.
6. Permit easy reversal of actions.
7. Support internal locus of control.
8. Reduce short-term memory load.

http://lncn.eu/dtu

SHNEIDERMAN'S 8 GOLDEN RULES

1. Strive for consistency.
2. Enable frequent users to use shortcuts.
3. Offer informative feedback.
4. Design dialog to yield closure.
5. Offer simple error handling.
6. Permit easy reversal of actions.
7. Support internal locus of control.
8. Reduce short-term memory load.

Use these as tick boxes when designing HRI!

http://lncn.eu/dtu

WHAT IS GOOD DESIGN?

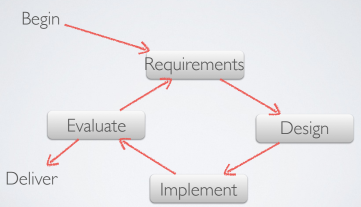
What makes digital products appeal to people?

What are the requirements of different user groups?

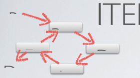
OVERVIEW

- **Iterative and experience-centred** design process
- Introduction to methods for **requirement establishment** in HCI
- Focus on approaches that can help us understand users:
  - **Requirements analysis**: Focus groups and interviews
  - **Design phase**: Participatory design and personas

ITERATIVE DESIGN



ITERATIVE DESIGN



- **Cyclic** (software development) process that starts with requirements establishment, leads over to the design and implementation of solutions which are then evaluated
- After **evaluation**, either exit, or enter new iteration in which requirements are adapted and design and implementation are updated
- System **evolves** throughout development

EXPERIENCE-CENTRED DESIGN

- Experience-centred design = **commitment to dialogue** with users and communities of interest
- At each stage of the iterative process there are different methodologies that have been found useful for informing design
- Lots of overlap – use the most appropriate one to find out the information you need to improve your design at each stage of the process

EXPERIENCE-CENTRED DESIGN

- Requirements establishment
  - **Ethnography**
  - Probes (cultural, experiential, evocative)
  - **Diary** studies
  - **Focus** groups
  - Card sorting tasks
  - **Interviews**
  - **Drama** / Role Play
  - Fictional Inquiry

EXPERIENCE-CENTRED DESIGN

- Design
  - Participatory design (focus groups)
  - Personas
  - Scenarios
  - Paper prototyping

EXPERIENCE-CENTRED DESIGN

- Evaluation
  - Early stages
    - Paper prototyping
    - Experience prototyping (Wizard of Oz)
- Later stages
  - Working prototype – lab usability study
  - Diary study
  - Analysis of data log files
  - Interviews

EXPERIENCE-CENTRED DESIGN

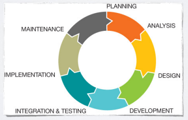
- The point of all of these is to "*invite participants to creatively express something about themselves, their values, their relationships, and the ways they make sense of experience*" (Wright & McCarthy, 2010, p.84).
- "*…to elicit personally meaningful reflections… to understand how people experience and make sense of situations, relationships and life events.*" (Wright & McCarthy, 2010, p.84).

EXPERIENCE-CENTRED DESIGN

- …Because understanding that, is what will lead you to design useful, interesting, successful technology
- …..and (more engineeringly) it's the best way to avoid costly errors in the development process
- A poor requirements phase can cost 100 times more to fix later in a project (Boehm & Basili, 2001).

REQUIREMENTS ANALYSIS

- Functional & Non-functional requirements
- These are software engineering practices
  - HCI is part of the SDLC, not an extra thing



FUNCTIONAL REQUIREMENTS

- Functional requirements are "a statement about an intended product that specifies what it should do" (Sharp, Rogers, & Preece, 2010)
- Put simply: Tangible system features

NON-FUNCTIONAL REQUIREMENTS

- Describe product or system characteristics with respect to performance – not just efficiency, but also softer qualities such as user experience
- Describe how a system is supposed to act

where does this stem from in HRI?

- Ideally need to be measurable so project progress and success can be assessed

REQUIREMENTS: QUICK GUIDE

- Identify what is **important** to stakeholders
  - This will allow you understand what is not acceptable (e.g., getting fired from my job because I need to play game during office hours)
- Involve the **stakeholder**
  - And more than one of them
- Be **prepared** – use props, examples, prototypes to support discussion

REQUIREMENTS: SUMMARY

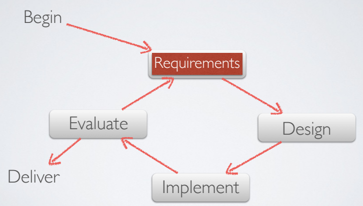
- **Functional and non-functional requirements go hand in hand** – define system features and execution
- Revisit throughout development process, reassess, ensure that nothing went wrong along the way

most users are novices when it comes to robots

- Understand that users may have different takes on requirements, show behaviours that were not anticipated, …

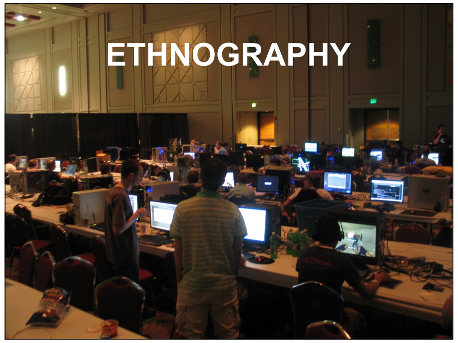
USER-CENTRED DESIGN

ITERATIVE DESIGN



REQUIREMENTS ESTABLISHMENT

- Ethnography
- Focus groups
- Interviews
- There are many other approaches – this is just a selection to give you tools to use for this class!



ETHNOGRAPHY

many lessons learned in care home!

- Everything is viewed naively – viewed as strange
- No theoretical framework in advance
- Gives a detailed and nuanced understanding of peoples lives
  - Other methods may be difficult - people may not answer questionnaires or interviews 100% honestly

Ethnography =
Direct observations of how people actually live, work etc., allows the designer to gain understanding of how technology will fit intuitively into their lives.

FOCUS GROUPS

- Discussing ideas with target audience to elicit their feedback on developers'/designers' thoughts – not just observing, but also interacting!
- These are very commonly used
  - Designers meet stakeholders
  - Social, personal interaction
  - Public discussion
  - Highlights agreement, disagreement
  - People express their understanding

FOCUS GROUPS

- The designer has a plan…
  - They have a rough idea of the information they want to elicit
    - Functional requirements
  - They provide props, materials, some guidance on discussion
  - They record participant responses

INTERVIEWS

- Interviews are a common method for establishing requirements
- Different types
  - Structured
  - Unstructured
  - Semi Structured

STRUCTURED INTERVIEWS

- The interviewer asks pre-determined questions
  - Similar to a questionnaire
- Exactly same questions used with every participant
- Closed questions
  - Require an answer from a pre-determined set of alternatives
  - Works if the range of possible answers is known
  - And people are in a rush
- Only useful if goals are very well understood

UNSTRUCTURED INTERVIEWS

- Exploratory conversations around a particular topic
- Questions are open
  - No expectation of the format or content of answers
  - Useful when you want to explore the range of possible opinions
- But you still need a plan – you need to be aware of what information you need to find out.
- Results in rich, unstructured data

SEMI-STRUCTURED INTERVIEWS

- Use both open and closed questions
- Basic script so that the same topics are covered
- Begin with pre-planned questions, then probe participant to expand upon that point
- Don't lead people to conclusions
  - "as an older person you probably like tea…"
- Probe – "Is there anything else you'd like to tell me"

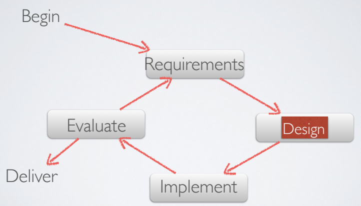
REQUIREMENTS ESTABLISHMENT

A. Ethnography – good to observe persons in natural environment without interfering with their activities, but no insights into personal opinions and ideas

B. Focus groups – good to get input from groups of people and their discussion of your ideas, will lead to interesting insights, but are hard to manage

C. Interviews – a way of obtaining individual feedback, but need to make sure you are not leading participants on

ITERATIVE DESIGN



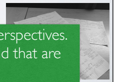
DESIGN

- Participatory Design
- Personas

- There are many other approaches – this is just a selection to give you tools to use for this class!

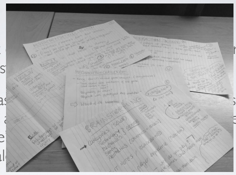
PARTICIPATORY DESIGN

- The involvement of end-users in the development process; not just in the testing phase, but as actual designers
- Goal: "[…] increase the public's engagement with research, facilitate learning and change, ensure that technologies are aligned to people's needs and remove designer subjectivity." (Khaled & Vasalou, 2014)



Allows you to gain new insights into the user's perspectives. They may have insights that you do not have, and that are impossible to imagine!

PARTICIPATORY DESIGN



- The involvement … rocess; not just in the tes…
- Goal: "[…] increas… search, facilitate learning … es are aligned to people… subjectivity." (Khal…

Allows you to gain new insights into the user's perspectives. They may have insights that you do not have, and that are impossible to imagine!

PERSONAS

- Alternative to directly involving end-users in design
- Fictional characters that are created by designers to represent target audience
- Still requires insights into needs and wishes of target audience, but does not require direct contact
- Can be based off literature, own experiences, group discussions, observations (ethnography!)
- Still subjective, so a bit tricky – keep that in mind when playing through scenarios

PERSONAS: EXAMPLE

- Sarah, 70 years old, widow, lives alone
- Has Arthritis but otherwise healthy and active
- Enjoys reading, going for walks, spending time with her children and grandchildren
- Little experience using computers, children bought her a tablet so they can Skype her
- Arthritis sometimes makes it hard to use tech, but also other things in daily life

PERSONAS: EXAMPLE

- Marcus, 21 years old, girlfriend lives in same town
- Student
- In his second year and a bit worried about finishing uni soon – thinking about what to do next
- Likes going out, seeing friends
- Wishes he could go home to see family more
- Uses phone all the time to text people

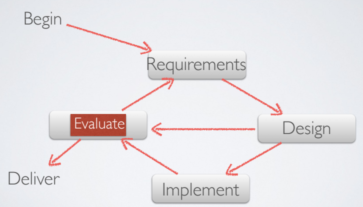
PERSONAS: QUICK GUIDE

- Include 'hard facts' about person
- Talk about strengths and weaknesses
- Try to create a 'full' image – include hobbies, interests, whatever makes the person unique
- Don't focus too much on the software you're trying to design – this will follow later
- When in doubt, read up about topics
- Sometimes you'll need to create several personas

REMEMBER: QUICK GUIDE

- Identify what is important to stakeholders
  - This will allow you understand what is not acceptable (e.g., getting fired from my job because I need to play video game during office hours)
- Involve the stakeholder
  - And more than one of them
- Be prepared – use props, examples, prototypes to support discussion

ITERATIVE DESIGN



DESIGN! EVALUATION

- Wizard-of-Oz
- (Paper) Prototype
- There are many other approaches – this is just a selection to give you tools to use for this class!

WIZARD-OF-OZ

- "The phrase Wizard of Oz has come into common usage to describe a testing or iterative design methodology wherein an experimenter (the "wizard"), [...] simulates the behavior of a theoretical intelligent computer application."
- Long version: http://deepblue.lib.umich.edu/bitstream/handle/2027.42/174/71952.0001.001.pdf

REMEMBER

- The point of user-centred and participatory design is to "invite participants to creatively express something about themselves, their values, their relationships, and the ways they make sense of experience" (Wright & McCarthy, 2010, p.84).
- Not all people are the same!
- Make no assumptions, or at least try the best to stop your brain from getting ahead of you!

The bottom line:

sit down with your target audience and find out about their preferences.

Test your stuff.

Test it a lot.

HUMAN-CENTRED DESIGN APPLIED TO SOCIAL ROBOTS

- Human mind is exquisitely suited to make sense of the world and people
- Use Natural Cues
  - Indicate what parts to operate and how
  - The mapping between intended and actual operations is intuitive
  - The effects of the operations are apparent
- Just the right things need be visible to avoid gadget overload

2-3 Weiss, A. et al., 2009. The USUS evaluation framework for human-robot interaction. In Adaptive and Emergent Behaviour and Complex Systems - Proceedings of the 23rd Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour, AISB 2009. pp. 158–165.

Harold Agudelo & Andrea Gigli

2-1 Bethel, C.L. & Murphy, R.R., 2010. Review of human studies methods in HRI and recommendations. International Journal of Social Robotics, 2(4), pp.347–359.

Matteo Bordoni & Sina Baharlou

MEASURING HRI - BASICS OF QUANTITATIVE RESEARCH

AIMS AND OBJECTIVES

- understand what quantitative research actually is
  - and what it's useful for
- facilitate critical thinking about quantitative results
- help design experiments
- critically appraise quantitative research (statistical) methods
- providing useful sources and references to dig deeper

REFLECTION ON LEARNING

- complement session on qualitative research methods
- providing a theoretical foundation
- hands-on exercises

QUANTITATIVE RESEARCH METHODS

Positivism

- The generation and testing of models, theories and hypotheses
- The development of instruments and methods for measurement
- Experimental control and manipulation of variables
- Collection of empirical data
- Modeling and analysis of data

Statistics

POSITIVISM

- a "philosophy of science"
- Etymology: re-imported from French "*positivisme*", derived from positif => 'imposed on the mind by experience'.
- assumes that there are natural laws that can be empirically measured or experienced
- only information measured and mathematically treated is a valid source of knowledge

"Foreknowledge is power."

Auguste Comte

WHY QUANTITATIVE METHODS?

WHY QUANTITATIVE METHODS?

WHY QUANTITATIVE METHODS?

- the world follows "laws"
- hypothesis testing
- "proving" a theorem
- good to defend
- separating effects
- reproducibility
- the truth is in the data
- parameterised model
- generalisation
- falsifiability
- measuring deviation
- identifying outliers

QUANTITATIVE VS QUALITATIVE

Qualitative research	Quantitative research
Inductive	Deductive
Subjective	Objective
Impressionistic	Conclusive
Holistic, interdependent system	Independent and dependent variables
Purposeful, key informants	Random, probabalistic sample
Not focused on generalization	Focused on generalization
Aims at understanding, new perspectives	Aims at truth, scientific acceptance
Case studies, content and pattern analysis	Statistical analysis
Focus on words	Focus on numbers
Probing	Counting

Source: Based on Patton (1990) and Chisnall (2001)

QUALITATIVE VS QUANTITATIVE

Hypothesis generation

derive a model

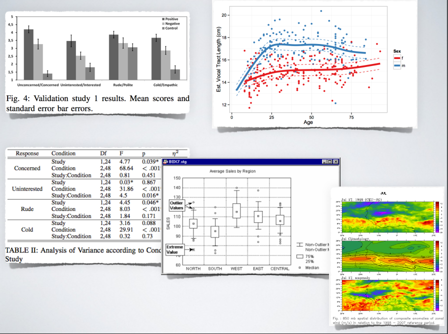
(but also quantitative can do that)

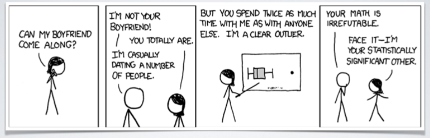
Hypothesis testing

verify your hypotheses (empirically)

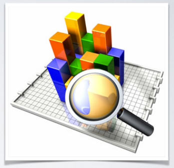
They go hand in hand!

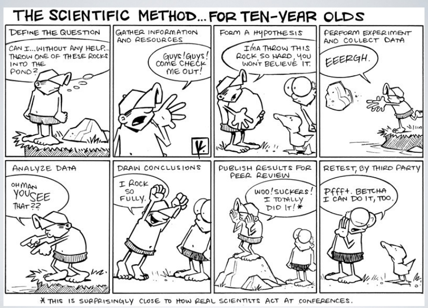
QUANTITATIVE RESEARCH METHODS ARE VISUAL



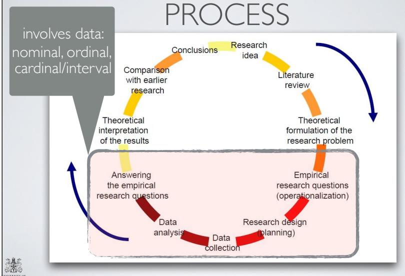


QUANTITATIVE METHODS IN YOUR RESEARCH





THE EMPIRICAL RESEARCH PROCESS



TYPES OF QUANTITATIVE DATA

- Nominal or categorical data – categories with counts
  - e.g., no. of men versus no. of women
- Ordinal data – ranked data
  - e.g., grades A, B, C, D, E
- Cardinal data – continuous measurements
  - e.g., velocity, height, etc

DISCLAIMER:

DOES NOT APPLY TO ALL SORTS OF RESEARCH!

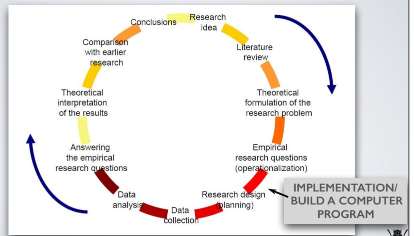
SCIENTIFIC METHODS IN COMPUTER SCIENCE?!

- "Science is generally taken as meaning either (a) the exact sciences, such as chemistry, physics, etc., or (b) a method of thought which obtains verifiable results by reasoning logically from observed fact." (Orwell, 1945)
- "Computer Science ... is the science of information processes and their interactions with the world" (Denning, 2005)
  - "The fundamental question underlying all of computing is, 'What can be (efficiently) automated?'" (Denning, 1989)
- "Computer Science integrates theory and practice, abstraction (general) and design (specific)" (Dodig-Crnkovic, 2002)

SCIENTIFIC METHODS IN COMPUTER SCIENCE?!

- "Scientific research can be seen as the usage of a rigorous method or methodology for achieving new knowledge." (Freitas, 2009)
- "Like in all science the first step will consist of modelling the problem or phenomenon to be studied – abstraction"
- "This simplifies the subject of research taking in consideration only the relevant aspects. After modelling the problem we must endorse a methodology: Theoretical, Experimental, and Simulation"

THE RESEARCH PROCESS



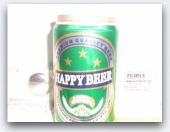
STAGES OF A TYPICAL C.S. RESEARCH PROJECT

Lemon and Nehmzow (1998)

- Specification of the problem or questions … within a body of background literature and common practice
- Synthesis – build a system (e.g. program) that is expected to succeed in the specified task – prediction!
- Experiment
- Analysis – interpretation of the experimental results
- Conclusion – judgement regarding success or failure of the synthesis, and its possible modification

EVALUATING THE ARTEFACT

- "Does exactly what it says on the tin."



- But does it really? How can we tell …? How to show we improved?

PERFORMANCE METRICS

- Count
  - Of how many times an event occurs
- Duration
  - Of a time interval
- Size
  - Of some parameter
- A value derived from these fundamental measurements

PERFORMANCE METRICS

- "Rate" metrics
  - Transactions per second
  - Bytes per second
  - (Number of events) ÷ (time interval over which events occurred)
- Error rates
  - Or success rates – e.g. percentage accuracy of a pattern recognition algorithm

EXAMPLE – METRICS FOR PATTERN RECOGNITION

- Pattern recognition problems, e.g. face detection
- We are often interested in the types of errors:
  - "false positives" – items incorrectly labelled as positive cases
  - "false negatives" – items incorrectly labelled as negative cases
- Common metrics: recall, precision, (overall) accuracy
  - Precision can be seen as a measure of exactness or fidelity, whereas recall is a measure of completeness

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

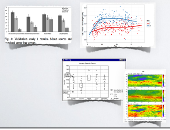
$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

TYPES OF QUANTITATIVE METHODS

TWO TYPES OF QUANTITATIVE

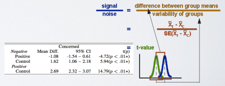
descriptive

summarise your observations



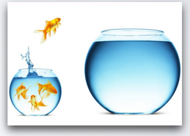
inferential (tests)

test your hypotheses



BIG CHALLENGES FOR QUANTITATIVE RESEARCH

- requires
  - quantification
  - control (of variables)
  - models or hypothesis
  - (laborious) repetition
  - access to the measurement



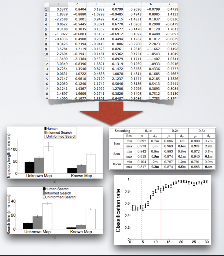
QUANTITATIVE =STATISTICS?

- well, not only…
  - …but mostly

DESCRIBING DATA

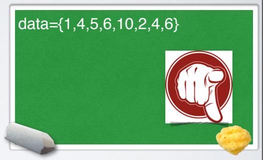
GENERATION OF KNOWLEDGE

- Research requires abstraction and generalisation
  - to generate models
  - to form hypotheses
- quantitative methods summarise observations



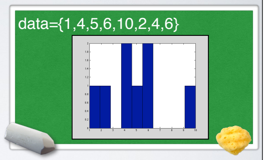
DESCRIBING DATA

- summarise a sample!

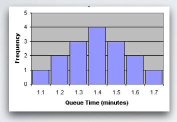


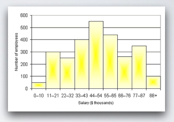
DESCRIBING DATA

- summarise a sample!
  - graphically with a histogram



HISTOGRAMS





- A histogram is a graphical version of a table showing the proportion of cases falling into a number of categories (bins).
- There is no "best" number of bins, and different bin sizes can reveal different aspects of the data.
- Also can be used to represent probability distributions (if the frequency values are normalized to sum to 1)

DESCRIBING DATA

- summarise a sample!
  - central tendency
  - measures of dispersion

data={1,4,5,6,10,2,4,6}

DESCRIBING DATA

- summarise a sample!
  - **central tendency**
  - measures of dispersion

data={1,4,5,6,10,2,4,6}

DESCRIBING DATA

- summarise a sample!
  - **central tendency**
  - measures of dispersion

data={1,4,5,6,10,2,4,6}
mean=4.75
median=4.5

How good is your estimate?

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

DESCRIBING DATA

- summarise a sample!
  - central tendency
  - **measures of dispersion**

data={1,4,5,6,10,2,4,6}

DESCRIBING DATA

- summarise a sample!
  - central tendency
  - **measures of dispersion**

data={1,4,5,6,10,2,4,6}
variance=7.6429
std dev=2.7646

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

DESCRIBING DATA

- summarise a sample!
  - confidence of mean estimate?

data={1,4,5,6,10,2,4,6}
mean m=4.75
std dev s=2.7646
sample size n=8

"magic number" for p=95%

$$ci = m \pm 1.96 \cdot \frac{s}{\sqrt{n}}$$

std error

http://graphpad.com/quickcalcs/CImean1/

DESCRIBING DATA

- summarise a sample!
  - confidence of mean estimate?

data={1,4,5,6,10,2,4,6}
mean m=4.75
std dev s=2.7646
sample size n=8
ci=[2.44, 7.06]

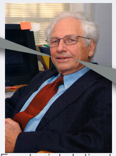
bad estimate?!

http://graphpad.com/quickcalcs/CImean1/

$$ci = m \pm 1.96 \cdot \frac{s}{\sqrt{n}}$$

EMPIRICS

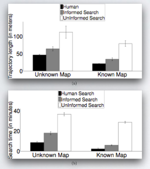
"There is no better data than more data!"

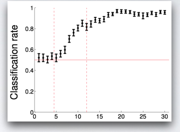


Frederick Jelinek

"Every time I fire a linguist, the performance of the speech recogniser goes up"

STD ERROR AND CONFIDENCE INTERVALS





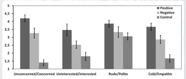
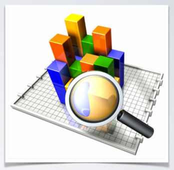


Fig. 4: Validation study 1 results. Mean scores and standard error bar errors.

TAKE HOME MESSAGE

- repeat any experiment to get "good" estimates
- report how good your estimates are
- consider in your experiment design how you could get enough data
- There are many more methods in "descriptive statistics"
  e.g. [Perry R. Hinton: Statistics explained, 2004, ebook]

TESTING HYPOTHESES



STATISTICAL ANALYSIS OF DATA

- When analysing data, your goal is simple: to make the strongest possible conclusion from limited amounts of data. To do this, you need to overcome two problems:
  - Important differences can be obscured by natural variation and experimental imprecision (noise). This makes it difficult to distinguish real differences from random variability. Is the observed effect genuine or merely due to chance?
  - The human brain excels at finding patterns, even from random data. Our natural inclination (especially with our own data) is to conclude that differences are real, and to minimise the contribution of random variability. Statistical rigour helps to prevent you from jumping to false conclusions.

CREATING A TESTABLE HYPOTHESIS

- e.g. "Does our system really work?" – does it lead to some measurable difference in performance?
- Often need to specify:
  - Independent variable (the aspect of the environment that we are interested in – "what do I change?")
  - Dependent variable (the behaviour that we are interested in – "what do I observe?")
  - Controlled variables ("what do I keep the same?")

(variable = something that changes, takes different values, that we can alter or measure)

"ARE MY RESULTS STATISTICALLY SIGNIFICANT?"

- Basic idea: to predict the probability of events occurring given our existing state of knowledge or by making certain assumptions
- If we observe something unlikely, we can either
  - Accept it as something new or different or significant
  - Reject it as being due to chance
- Typical confidence level: 95%, i.e. the probability of this event happening by chance is around 1 in 20 (p=0.05)

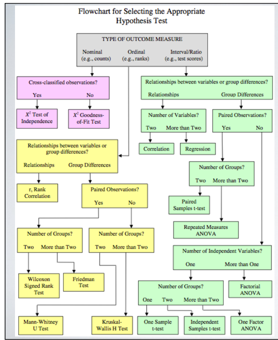
THE NULL HYPOTHESIS

- Statisticians argue by a process of falsification
- Experimental hypothesis - there is a difference between two populations
- Null hypothesis – there is no difference between the two populations
- However, we cannot "prove" the absolute truth or falsehood of the hypothesis. Rather we use statistical evidence to either "tend to accept" or "tend to reject" the null hypothesis at a given confidence level.

WHICH STATISTICAL TEST TO CHOOSE?

- Wide variety of different tests, each with its own underlying assumptions. E.g. some common tests:
  - Pearson correlation – quantify association between two variables
  - Chi-squared test – compare two groups of nominal data
  - Student's t-test for the difference between means (unpaired or paired) – compare two groups of cardinal data
  - ANOVA – compare three or more groups of cardinal data
- If in doubt, consult a statistics expert!

http://statpages.org
http://www.graphpad.com/www/Book/Choose.htm

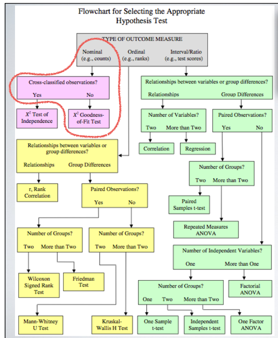


A LITTLE STUDY

- Example: a random sample of 200 men and 200 women. Are the observed differences between male and female occupations genuine or due to chance?
- Nominal? Ordinal? Cardinal/Interval?

Occupation	Men	Women	Total
Professional	40	20	60
Skilled	100	80	180
Unskilled	60	100	160
Total	200	200	400

Which test to choose?



Occupation	Men	Women	Total
Professional	40	20	60
Skilled	100	80	180
Unskilled	60	100	160
Total	200	200	400

Which test to choose?

CHI-SQUARED TEST

- Example: a random sample of 200 men and 200 women. Are the observed differences between male and female occupations genuine or due to chance?
- Nominal? Ordinal? Cardinal/Interval?

Occupation	Men	Women	Total
Professional	40	20	60
Skilled	100	80	180
Unskilled	60	100	160
Total	200	200	400

CHI-SQUARED TEST

Occupation	Men	Women	Total
Professional	40	20	60
Skilled	100	80	180
Unskilled	60	100	160
Total	200	200	400

Observed values

Occupation	Men	Women	Total
Professional	30	30	60
Skilled	90	90	180
Unskilled	80	80	160
Total	200	200	400

Expected values

- Could this pattern have been produced by chance?
- Calculate χ^2 value and use tables to look up the probability p of this value occurring by chance
- Use p-value to accept or reject the null hypothesis

CHI-SQUARED TEST

Occupation	Men	Women	Total
Professional	40	20	60
Skilled	100	80	180
Unskilled	60	100	160
Total	200	200	400

Observed values

Occupation	Men	Women	Total
Professional	30	30	60
Skilled	90	90	180
Unskilled	80	80	160
Total	200	200	400

Expected values

- Null hypothesis: the observed value can be explained by the expected distribution
- use http://graphpad.com/quickcalcs/chisquared1/
- for "Professional":

The value of the test-statistic is

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

where

χ^2 = Pearson's cumulative test statistic, which asymptotically approaches a χ^2 distribution.
O_i = an observed frequency;
E_i = an expected (theoretical) frequency, asserted by the null hypothesis;
n = the number of cells in the table.

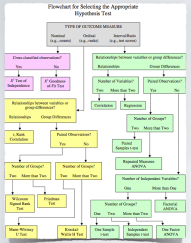
$$\chi^2 = \frac{(40-30)^2}{30} + \frac{(20-30)^2}{30}$$

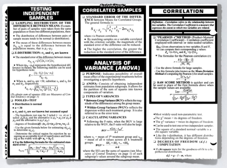
$$\chi^2 = 6.67$$

$$p(\chi^2 = 6.67) = 0.0098$$

null hypothesis rejected by "convention"

MANY MORE TESTS EXIST…



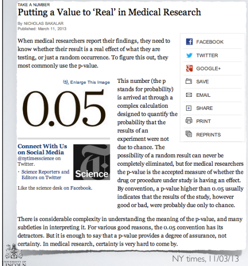


all end up with a p-value for the null hypothesis

CRITICAL REFLECTIONS

traps to avoid

TO P OR NOT TO P


TAKE A NUMBER

Putting a Value to 'Real' in Medical Research

By NICHOLAS BAKALAR
Published: March 11, 2013

When medical researchers report their findings, they need to know whether their result is a real effect of what they are testing, or just a random occurrence. To figure this out, they most commonly use the p-value.

FACEBOOK
TWITTER
GOOGLE+
SAVE
EMAIL
SHARE
PRINT
REPRINTS

Enlarge This Image

0.05

This number (the p stands for probability) is arrived at through a complex calculation designed to quantify the probability that the results of an experiment were not due to chance. The possibility of a random result can never be completely eliminated, but for medical researchers the p-value is the accepted measure of whether the drug or procedure under study is having an effect. By convention, a p-value higher than 0.05 usually indicates that the results of the study, however good or bad, were probably due only to chance.

Connect With Us on Social Media
@nytimesscience on Twitter.
· Science Reporters and Editors on Twitter

Science

Like the science desk on Facebook.

There is considerable complexity in understanding the meaning of the p-value, and many subtleties in interpreting it. For various good reasons, the 0.05 convention has its detractors. But it is enough to say that a p-value provides a degree of assurance, not certainty. In medical research, certainty is very hard to come by.

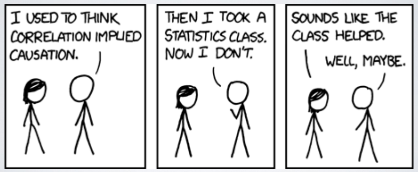
NY times, 11/03/13

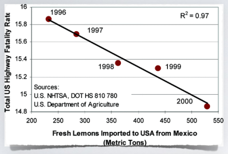
- http://hilaryparker.com/2013/03/12/about-that-pvalue-article/
- obtaining a **very small p-value implies that our result is very unlikely under the null hypothesis**
- "If the data is unlikely under the null hypothesis, then either we observed a low probability event, or it must be that the null hypothesis is not true."
- The p-value tells us something about the likelihood of the data, in a world where the null hypothesis is true.
- The p-value is never a statement about the probability of hypotheses, but rather is a statement about data under hypothetical assumptions. Since the p-value is a statement about data when the null is true, it cannot be a statement about the data when the null is not true.
- "A lack of evidence for something is not a stack of evidence against it."

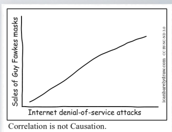
BE CAREFUL WITH YOUR WORDS

- still: the smaller the p-value the more significant your result of rejecting the null-hypothesis
- by convention:
  - p<5%, statistically significant, (denoted as *)
  - p<1%, highly statistically significant, (denoted as **)
  - p<0.1%, extremely statistically significant (denoted as ***)

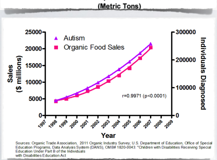
CORRELATION DOES NOT IMPLY CAUSATION







Correlation is not Causation.



CORRELATION DOES NOT IMPLY CAUSATION

- Correlation can have various causes:
  - A causes B
  - B causes A
  - Both A and B are consequences of a common cause
  - or some combination
  - A and B aren't related and the correlation is a coincidence.
- However, often you see arguments like:
- A occurs in correlation with B
  - Therefore, A causes B

No! Solution: Experiments where you vary A and check for a statistical effect on B

CHALLENGES REMAIN

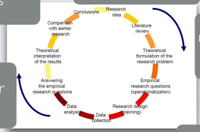
too many free variables

too little data

measuring the right data

interpreting your findings

choosing the right description/tests



2-2 Riek, L.D., 2012. Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines. Journal of Human-Robot Interaction, 1(1).

Daniele Evangelista & Maurizio Maisto

HRI IN RESEARCH

6.1 Human-Robot Interaction-Specific Conferences

- **Conference on Human Robot Interaction (HRI):** This conference, created in 2006, is focused specifically on HRI research. Attendees and submissions to this conference are mostly from engineering (electrical engineering and computer science) with contributions from allied fields, such as psychology, anthropology, and ethics.
- **International Workshop on Robot and Human Interactive Communication (RO-MAN):** RO-MAN provides a forum for an interdisciplinary exchange for researchers dedicated to advancing knowledge in the field of human-robot interaction and communication. Importantly, RO-MAN has traditionally adopted a broad perspective encompassing research issues of human-machine interaction and communication in networked media as well as virtual and augmented tele-presence environments. RO-MAN is somewhat longer-standing than HRI.
- **International Conference on Development and Learning (ICDL):** This conference brings together the research community at the convergence of artificial intelligence, developmental psychology, cognitive science, neuroscience, and robotics, aimed at identifying common computational principles of development and learning in artificial and natural systems. The goal of the conference is to present state-of-the-art research on autonomous development in humans, animals and robots, and to continue to identify new interdisciplinary research directions for the future of the field.
- **Computer/Human Interaction (CHI) Conference:** CHI is an established conference in Human-Computer Interaction (HCI). Every year, it is a venue for 2000 HCI professionals, academics, and students to discuss HCI issues and research and make lasting connections in the HCI community. HRI representation in this meeting is small, but the two fields (HRI and HCI) have much to learn and gain from each other.

HRI IN RESEARCH

6.2 General Robotics and AI Conferences

- **Association for the Advancement of Artificial Intelligence (AAAI):** AAAI's annual conference affords participants a setting where they can share ideas and learn from each other's artificial intelligence (AI) research. Topics for the symposia change each year, and the limited seating capacity and relaxed atmosphere allow for workshoplike interaction.
- **AAAI Spring and Fall Symposia:** These annual symposia cover a broad range of focused topics. With the rapid growth of HRI, the topic and related areas (e.g., service robotics, socially assistive robotics, etc.) symposia are held in each session.
- **Epigenetic Robotics (EpiRob):** The Epigenetic Robotics annual workshop has established itself as an opportunity for original research combining developmental sciences, neuroscience, biology, and cognitive robotics and artificial intelligence is being presented.
- **International Conference on Robotics and Automation (ICRA):** This is one of two most major robotics conferences, covering all areas of robotics and automation. In recent years, the themes of the conference have included many areas of HRI research, such as "Humanitarian Robotics," "Ubiquitous Robotics," and "Human-Centered Robotics", reflecting the rapid growth in the field.
- **International Conference on Intelligent Robots and Systems (IROS):** This is the other major international robotics conference, featuring a very large number of papers, with a growing representation of HRI. Tutorials and workshops, as well as organized/special sessions in HRI are featured regularly.
- **International Symposium on Experimental Robotics (ISER):** ISER is a single-track symposium featuring around 50 presentations on experimental research in robotics. The goal of these symposia is to provide a forum dedicated to experimental robotics research with principled foundations. HRI topics have become a regular part of this venue.