1 Introduction

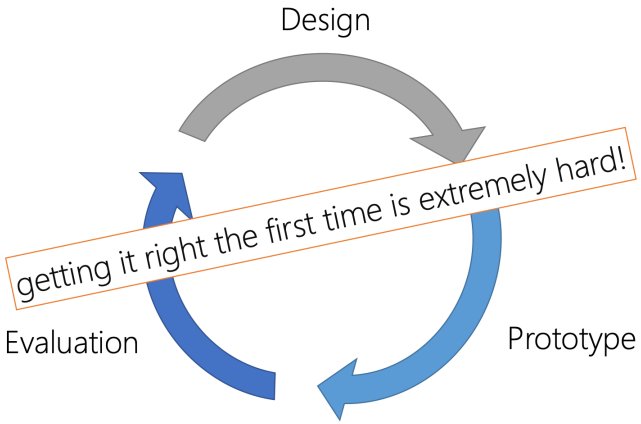
Goal of course: Understand principles of user-centered design and able to apply these to practice. Learn about the basic notions of Computational Design in HCI context.

Moore's Law Computational power grow exponentionally. Transistor count doubles every two years. Also with RAM and pixel densities. However: Human capabilities stay stable.

Good System design Accounts for human capabilities, human error and exceptional circumstances.

Human Computer Interaction Concerned with design, evaluation and implementation of interactive computing systems for human use.

Process in HCI



Formative : understand problem and user to inform our design. Evaluative: understand how well design works. Also detects mistakes in design.

Time to move to element i

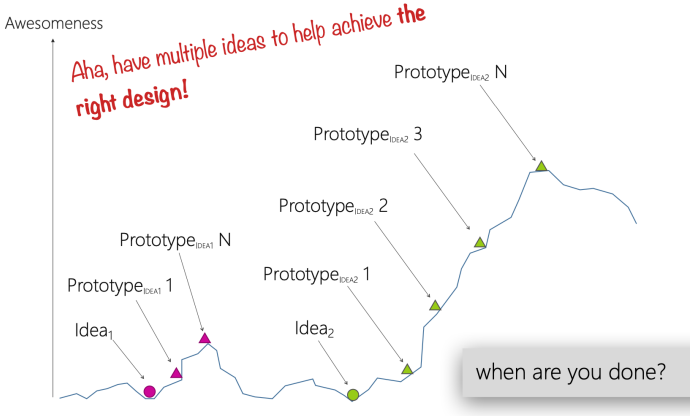
$$t_i = a + b \log_2 \left(\frac{A_i}{W_i} + 1 \right)$$

Average time to operate menu

$$T = \frac{1}{n} \sum_i p_i t_i$$

2 User-centered Design

Design intention vs User Needs
Prototyping as an interative process



Does the design work properly in the context of use? If not fix the problems and carry out more tests.

Early focus on users and tasks: Cognitive, behavioral, anthropomorphic AND attitudinal characteristics.

Empirical Measurement: Observe user's reactions and performance in scenarios, manuals simulations and prototypes, record and analyze.

Root-Cause Analysis Problems need to be discovered (find the right problem to solve, not any problem to solve) and find the right solution to it.

Need finding
Users rarely know what they want. Cannot imagine what is possible. Instead look at tasks, context:

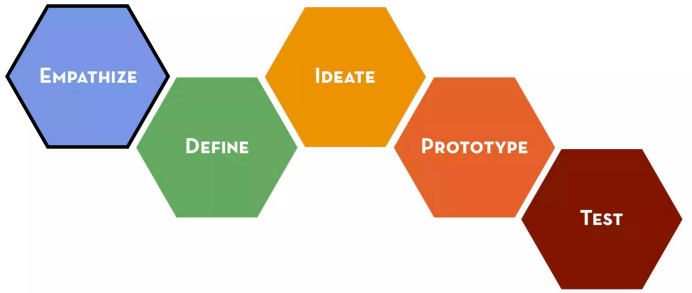
- What information needed?
- Identify collaborators
- Why is task achieved the way it is?
- Identify tasks in existing behavior
- Identify tasks in future scenarios

We ourselves are not representative of the typical user. To learn about customers conduct interviews, self-reports and logging/analytics. Also observe users performing tasks and understand their cognition.

Understanding the User Active observation is not knowing yet what you are looking for.

- Immerse
- Observer
- Engage

Design Thinking Process



Goals of Need finding

- Distill useful and actionable insights
- Make meaning from needfinding data
- reframe problem to guide solution search

We start with closed ended questions and move to open ended questions: "What's and why's of feelings". Engage people in their environment. The goal is to find inspiring users, that surprise us and bring us to gamge-changing ideas.

Needs vs. features vs. requirements
Requirements are goals that the system needs to accomplish. Solutions fulfill requirements. What does the user want to accomplish and how is he doing it? What would they like to be doing? What are they currently disliking? For what is the system usable and what tasks will it support? Answering these questions will make the system more usable.

There are tons of methods to needfinding such as:

- Task Analysis
- Interviews
- Affinity Diagrams
- Cognitive Walkthrough
- Questionnnaires
- Focus Groups
- Diary Studies
- "Speed dating"
- etc.

Interview
Interviewee speaks 90 percent of time and stays on topic. We choose participants to be representative target users, either

current or potential future users. We like both experts and typical users. Try to provide and explanation into how users make sense to themselves.

Common pitfalls in interviews Suggesting answers. Hypthetical questions.

Diary Study Ask people to record events as they happen. User diary studies for rare events, easily forgotton events and events where the actual frequncy is important.
Problems with diary studies is that the simple tracking of their behaviors will change their behaviors.

Retrospective Survey Ask about things happened in the past. Use this for critical events (well remembered), recent and memorable events, rare events that had a big impact an are memorable. DO not use them for hard to remember events.

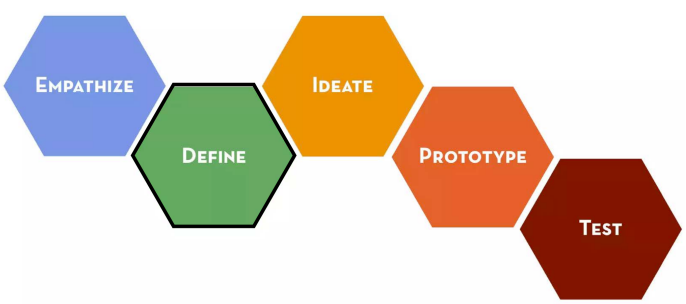
Artifact Analysis Look at things people leave around to understand a problem they might have. Use this for physical spaces (physical artifacts from workflows), tasks involving artifacts and interactions generating artifacs (emails, social media posts etc.). Only use if there are in fact artifacts and there is no faster way to learn information.

Contextual Inquiry Ethnographics or participatory design, combining aspects of other methods. Interviewing, observing in the context of work. Goal is to discover real requirements of the work. Interview people while they are working and gather real artifacts. User decides the tasks, but you decide the focus.

Key Differences

interviews, surveys, focus groups	Contextual Inquiry
summary data & abstraction	in-situ experience & data
what customers say	what users do
subjective	objective
limited reliability of humans	spontaneous, as it happens
what users/customers *think* they want	what users actually need

Result of Need-Finding We know what works and what does not yet exist. Problems and incomplete parts in process. So we have a long list of problems.

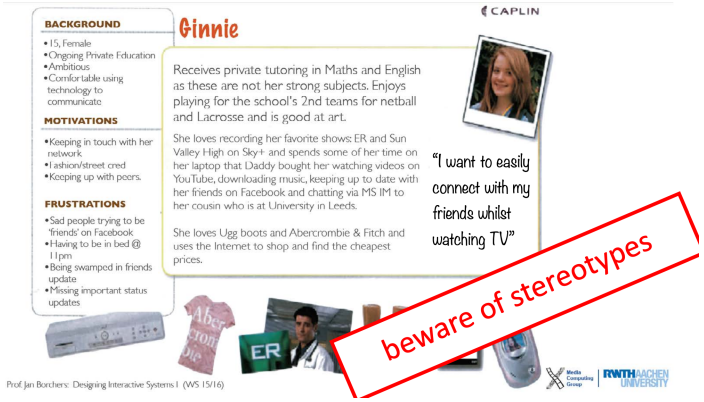


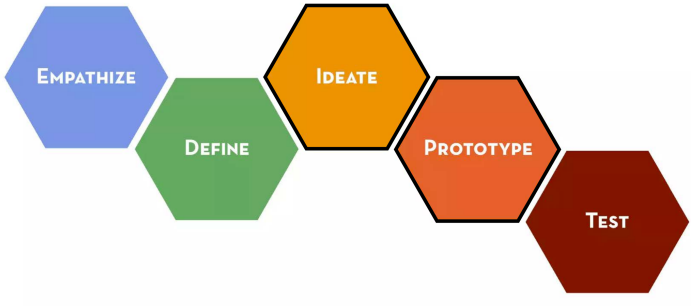
Define This part is more a focussing part and not flaring. Figure out what is important from collected data. Group info and find relations.

Affinity Diagrams
Data with affinity to each other are grouped together to form category. Groups are given labels, can be one or more categories in the end. Identify user, need and insight. Combined to create point of view.
Good point of view insires the team, frames the problem in a focussed way. Empowers to make decisions and fiels brainstorming by suggesting "how might we" statements.

The elastic user The elastic user can mean everyone and also noone. Vague and unfocused, lack of specifics makes it easy to rationalize any design.

Personas Personas are precisely described. Act as stand-in for real users. Guide design decision. Fictitious but based on knowledge of real users. Informed from observations. Personas are not elastic, don't make them fit the prototype.





Ideate Flaring here, not focussing anymore.

Ideation techniques

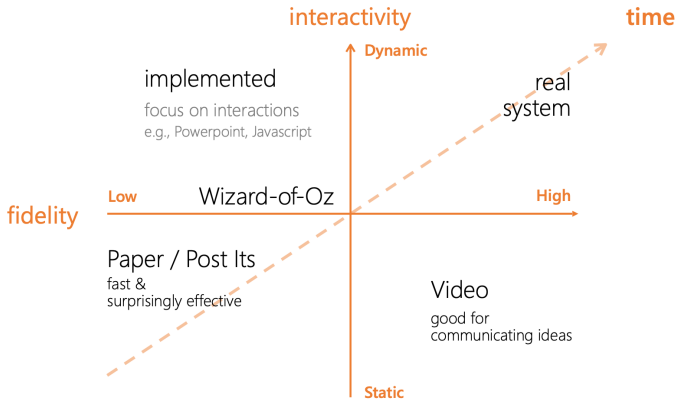
- brainstorming
- mind-mapping
- storyboarding
- sketching
- low-fi prototyping

3 Prototyping

Prototypes develop from sketches over time and are more defined in their criteria weights. Make multiple protopytes to evaluate different approaches and check for failure/success. Prototypes help us uderstand requirements and specifications of the idea. They answer a specific question.

Vertical vs. horizontal Vertical provides critical path of one or few features (real feature on that path is completed). Horizontal provides only overview with little to no functionality.

Fidelity and Interactivity

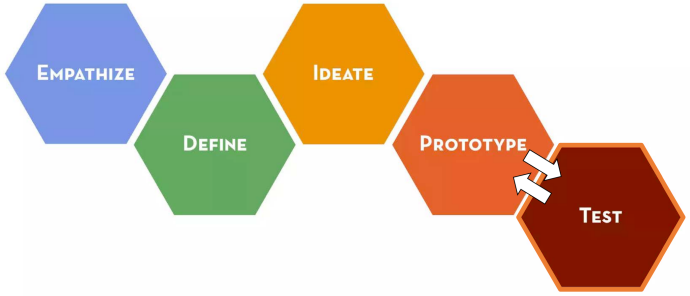


Paper prototypes Are rapid and cheap.

Wizard of Oz Interprets user input and simulated a system response. Allowd rapid testing of complex feature before implementing.

MidFi-Prototypes
Physical (paper, cardboard, lego etc.) to software.

- Powerpoint, Keynote
- AdobeXD, Figma



Anylytical vs. Empirical

Analytical

Look at inherent attributes of the design, rather then the desing in use. Intrinsic characteristics of the design. Examples are usability/UX inspection methods, design walkthroughs, heuristic evaluations.

Empirical

Based on how well a design or design change pays off in terms of real observable usage. Includes quantitative and qualitative data. Examples are usability/UX scores, controlled experiments and case studies.

Formative vs. Sumative

Formative

Helps form the design. Part of iterative process. Evaluations done during testing. Mainly collects qualitative data but also quantitative. Focusses on what is not working.

Summative

Helps sum up the design. Evaluates the success of the finished product, and compares with competitors. Collects quantitative data, and focussed mainly on what is working.

Evaluation criteria for UI design

Usability

Extent to which product can be used by specific users to achieve goals with effectiveness, efficiency and satisfaction. Five quality components of Usability:

- learnability
- efficiency
- memorability
- errors
- satisfaction

User Experience

Totality of the effect(s) felt by a user as a result of interaction with the usage context of a system. device or product.
It includes:

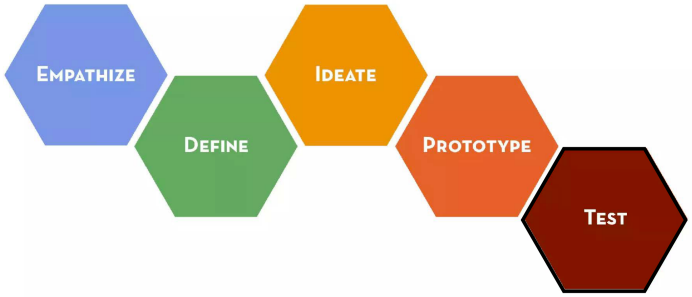
- Usability
- usefulness
- emotional impact
- savoring memory after interaction

It embraces seeing, touching, thinking about system or product and admiring it and its presentation. Focusses on holistic experience of user.

Affordances Actions that the design of an obkect suggest to the user. Provide strong clues to how objects are to be used without labels, explanations or manuals.

Works for both physical objects and software. Up to a certain degree of complexity.

4 Analytical Investigation



Is performed by usability experts and domain experts. They use their knowledge of users and technology to assess the usability and user experience. Result can be formal or informal reports.
Two types of analytical investigation:

1. Usability and UX inspection (Design, cognitive walkthroughs, heuristic evaluation)
2. Predictive user performance models (GOMS, KLM)

1. Usability and UX inspection

Cognitive Walkthroughs

Evaluate design by experts, with the goal of exploring the design on behalf of the users. Difference to UX inspection: Ux inspection only one aspect of a design presented to experts. Cognitive walkthrough: More focussed on ease-of-learning.

Heuristic Evaluation

Heuristics are design guidelines. Examine the interface, judge is compliance with recognized usability principles (heuristics). Is cheap, fast and easy to use. Is developed for inexperiences practicioners, experts can be limited through heuristics. Can be done on paper-only prototypes.

1. Briefing to tell evaluators what to do
2. Each evaluator inspects interface alone (at least twice, get feel for flow of interaction and scope of system, also focus on specific interface elements)
3. evaluators aggregate findings
4. debriefing session, discussion of possible redesigns for major UX problems, look as positive aspects

Optimally between 3 and 5 evaluators. Limited because it does not encourage to take a rich and comprehensive view of interaction. Its only a rough outline, and expert find problems withs inspection not heuristics. Danger of overestimating heuristics and use for any evaluation.

Nielsen's Heuristics

1. Visibility of system status

System should always keep users informed about what is going on, through approp. feedback in reasonable time.

2. Match between system and the real world

System should speak the users' language, wih words, phrases and concepts familiar to the user, rathen thans system-oriented terms. Follow real world conventions, make info appear in natural and logical order.

3. User control and freedom

Users need a clearly marked emergency exit from unwanted state, if chosen system functions by mistake.

4. Consistency and standards

User should not have to wonder whether different words, situations or actions mean the same thing.

5. Error prevention

Good error messages, but better is design that prevents a proble from occurring in the first place. Eliminate error-prone conditions or check for them and give users a confirmation option before commiting to the action.

6. Recognition rather than recall

Minimize the user's memory load by making opjects, options and actions visible. Instructions for use of the system should be visible or easily retrievable, whenever appropriate.

7. Flexibility and efficiency of use

Accelerators may often speed up the interaction for the expert user, such that the system can support both inexperienced and experienced users.

8. Asthetic ans minimalist design

Dialogs should not contain irrelevant or rarely needed information. Extra infos compete with the relevant units of information.

9. Help users recognize, diagnose and recover from errors

Error messages should be expressed in plain language, precisely indicate the problem and constructively suggest a solution.

10. Help and documentation

Even it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Should be easy to reach, list concrete steps and should not be too extensive.

2. Predictive User performance models

Way of evaluating roducts or design without directly involving users. Estimated of efficiency of systems for different tasks.

We use GOMS to model knowledge about the system and cognitive provesses involved when users interact with systems.

We use KLM to provide numerical predictions to performance and estimate chains of operations.

GOMS model

Goals

The state the user wants to achieve.

Operators

Cognitive processes and pyhsical actions needed to attain the coals (mouse click etc.)

Methods

Procedures for accomplishing th goals, drage mouse over search giel, type in term, press go etc ...

Selection Rules

Decide which method to select when there is more than one.

Goms example:

Goal: delete a word in a sentence

Method for goal of deleting a word using menu option:

- Step 1. Recall that word to be deleted has to be highlighted
- Step 2. Recall that command is 'cut'
- Step 3. Recall that command 'cut' is in edit menu
- Step 4. Accomplish goal of selecting and executing the 'cut' command
- Step 5. Return with goal accomplished

Keystroke Level model (KLM)

Mesarues and compares execution times.

Operator name	Description	Time (s)
K	Pressing a single key or button	0.35 (average)
	Skilled typist (55 wpm)	0.22
	Average typist (40 wpm)	0.28
	User unfamiliar with the keyboard	1.20
	Pressing shift or control key	0.08
P	Pointing with a mouse or other device to a target on a display	1.10
P_1	Clicking the mouse or similar device	0.20
H	Homing hands on the keyboard or other device	0.40
D	Draw a line using a mouse	Variable depending on the length of line
M	Mentally prepare to do something, e.g. make a decision	1.35
R(t)	System response time – counted only if it causes the user to wait when carrying out his/her task	*t*

Predictive models strengths and weaknesses

- Relatively to perform comaprative analysis for different interfaces and prototypes, specifications.
- Can only model high-level tasks, involving small set of high routine low level tasks
- Only valid for predictable/expert behavior (no multitasking, fatigue, learning effects etc)

5 Evaluations and Experimental Design

Formative early in the design process, sanity checks that we're building the right thing. *Summative* to check if we improved upon our last iteration, does it work better than other solutions?

Quantitative Evaluation Methods

Ensure certain level of quality, comparesolutions objectively, attain a scientific statement.

Primary Usability Metrics

A **usability metric** reveals something about the **interaction** between **the user** and **the thing:**

Effectiveness	Efficiency	Satisfaction
being able to complete a task	amount of effort required to complete the task	degree to which the user was happy with his/her experience while completing the task

these metrics can help answer these critical questions:

- Will users **like** the product?
- Is this new product **more efficient** than **past products**?
- **How does the usability** of this product/version **compare** to others?
- What are the **most significant** usability **problems** with this product?
- **Are improvements** being made from **one design iteration** to another?

Cause and Effect

We want to identify clear causal links. Cause precedes effect, they need to correlate and other explanations have to be ruled out. Isolate causality by controlled experiments. Alter design with suspected cause absent (control) and present (experimental condition). All other conditions should be identical.

Quasiexperimentell — Observational

We observe that independent variable and dependent variable are highly correlated, but did not control for anything (for instance participation in exercices and final exam grade).

Experimental — Controlled

We randomly assign students to exercise and no exercise condition, then we controlled for other variables and results implies causality.

characteristics of Emprical Methods

- Objectivity
- Reproducibility
- Validity (internally and externally)
- Relevance

For instance threat to external validity is over-use of specific participant groups (only psychology or cs students).

The experiment

Indepdendent variables affect the dependent (measured) variables through experiment.
Variables can be catagorical, ordinal (ordered discrete), or cardinal/interval (continuous) data.

Designing an empirical study

1. What is being compared? (which Indepdendent variables)
2. What are they bing comapred in? (dependent variables, metrics)
3. What lse is being varied? (extraneous variables to control/eliminate)
4. Relevance

Look at slide set 5 for various examples.

More complex comparisons

Different experimental designs possible: *Within subjects:* Everyone-does everything. *Between subjects:* Only on condition per group.

Latin Square Counterbalancing

full randomization can lead to huge experiments (e.g., 6! = 720)

Latin square design reduces number of experimental orderings

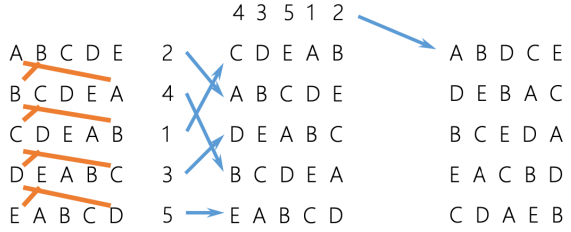
- total number of experimental conditions is the square of the number of treatments
- each treatment appears once and only once in each row and column

A	B	C	D	E
D	E	B	A	C
B	C	E	D	A
E	A	C	B	D
C	D	A	E	B

Latin Square Example for 5

first row in alphabetical order ⇒ A B C D E

subsequent rows – shift letters one position



Then: randomize the order of the rows: i.e., 2 4 1 3 5
randomize the order of the columns: i.e., 4 3 5 1 2

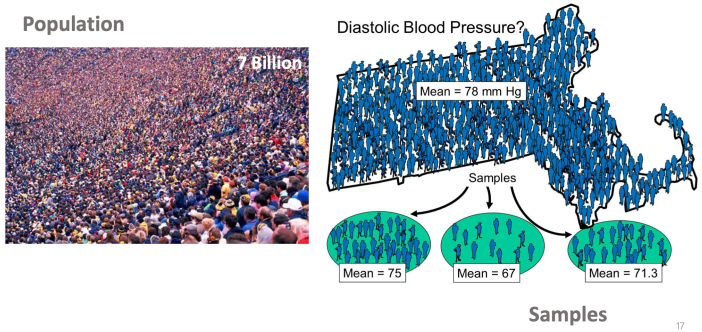
6 Statistical Analysis

Data Collection

Important:

- Choose representative sample
- Form hypthesis to make asssumptions testable
- collect data to test Hypothesis
- collect all avaiable data (better too much)

Population vs Sample



Generalizability

Results should be valid for all people. Participants should be representative of population. Look at relevant factors, such as Age, Gender, Occupation.

Hypothesis testing

Effect size is difference in the means of H_A and H_0. It is unknown a-priori and hence we don't know how to show the threshold for acceptance. Therefore instead of showing H_A is true, we show that the obtained data is incompatible with H_0.

Descriptive statistics and validation

How should data be validated?

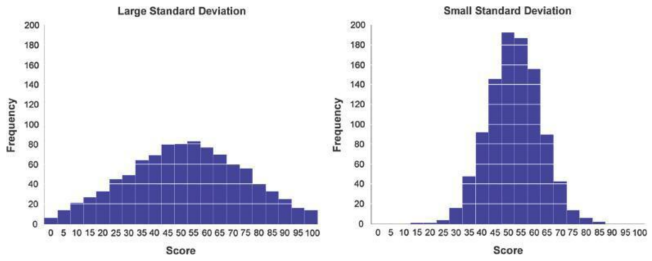
Mean

Least distance to all other data points. Good representation of data points piled together. Bad representation if some data points are extreme values (outliers). Doesn't make sense if we divide through categorical or ordinal data.

Median

is robust against outliers. The number at the middle of the ordered data points. Natural choice to represent ordinal data.

Distribution



Confidence Interval (CI)

Interval in which we are very sure that our true values lie in. We mostly choose 95 percent of values to lie within this interval if often replicated. Confidence interval of mean difference (two samples)

Analysis

Bayesian quantitative approach no covered in this course, also not qualitative analysis methods.

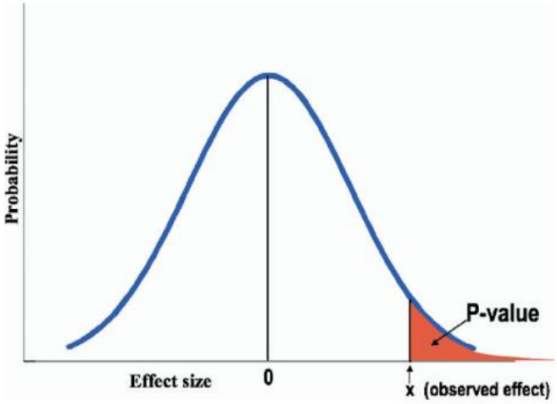
Frequentist Approaches

Hypothesis testing

We assume H_0 to be true. The lower our p-value the less likely that H_0 is true and H_A is true. P-value indicated how compatible the data to which hypothesis is.

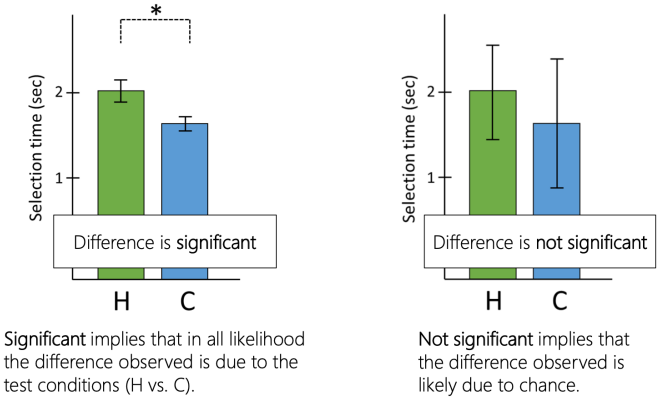
P-Value

P-value is probability of observed data if H_0 were true.



Alpha-Level

Threshold to determine if p-value is lowe enough. Usually $\alpha = 0.05$. In medicine even lower. If p-vlue is larger than α this does not mean that H_0 is true!



Significant implies that in all likelihood the difference observed is due to the test conditions (H vs. C).

Not significant implies that the difference observed is likely due to chance.

Errors

Type I error: Effect was found, but no effect in reality (False-Positive).
Type II error: No effect was found, but effect exists in reality (False-Negative).

Degrees of Freedom

Number of values that are free to cary. Number of observations (n) minus the number of parameter estimates. For one-sample t-test:

$$v = n - 1$$

.

Independent vs dependent samples

Independent: One subject only exposed to one condition (use different subjects for differen conditions). Also referred to as Independent measures or means.
Dependent: Same subject exposed to all conditions (Within subject design). Also referred to as matched pairs or paired samples.

Parametrix vs. non-parametric tests

Non-parametric do not assume specific distribution. Assume equal spread of group samples. Less statistical power. Type II error more likely.

Examples: Chi-Square, Mann Whitnes, Wilcoxons signed rank test, Kruskal Wallis, Friedman

Parametric tests assume gaussian distribution and homoscedasticity (equal variances). More power!
Examples: one-sample t-test, two-sample t-test, paired-sample t-test, one-way/factorial ANOVA, repeated measured ANOVA.

A/B Testing

Common example as in our case: Change one categorical independent variable with two levels (A and B) and measure one interval dependent variable. In our case task execution time.

Independent t-test

Checks if two means are reliably different from each other. $t =$(variance between groups) /(variance within groups).
Large t means different groups (H_0 refuted).

From t-value to significance

T-values lead us to our p-value over degrees of freedom in standardized tables. T-distribution depends on sample size (degrees of freedom). Its a distribution of t-values of a population where the null hypothesis is true.

ANOVA analysis of Variance

Use this if independent variable /factor has three or more levels. One-way ANOVA is used for data with one factor and multiple levels. Factorial ANOVA is used for data with multiple factors and levels. Does not tell us which levels are different.

Effect size

Statistical significance does not mean that the measured effect is meaningful. So we need a standardized effect size. We use Cohen's d, Pearson's correlation, odds ratio.
Cohens' d is a standardized mean difference between the samples. Depends on the field.

$$\text{Cohen's } d = \frac{\text{mean}(A) - \text{mean}(B)}{\text{mean}([\text{std}(A), \text{std}(B)])}$$

Power analysis

Compute min. number of participants to achieve desired effect. Can be calculated from

- prob. of finding an effect that is not there ($\alpha = 0.05$)
- prob. of finding an effect that is there ($1 - \beta = 0.8$)

- the desired effect size (HCI d = 0.8)

Software for statistical analysis

- SPSS
- Python
- R
- etc.

Reporting

Writing up the results

To compare the effect of the independent variable on the dependent variable, we conducted a statistical test as data [not] meets assumptions.

Which variables?
Which statistical test?
Which assumptions?

With condition A, ... (M = XX, SD = X). With condition B, (M = XX, SD = XX).

How do the samples look like?

The mean difference between the two groups was statistically [not] significant; DOF of test, p-value, etc.

How do differences look like?

These results indicate that condition A was ... than condition B.

What do differences mean?