4th February 2020

CSCM39/CSDM001: Human Computer Interaction

Dr Siyuan Liu

Department of Computer Science

siyuan.liu@swansea.ac.uk

Office Hour: Thursday 2-4pm

PURPOSE OF

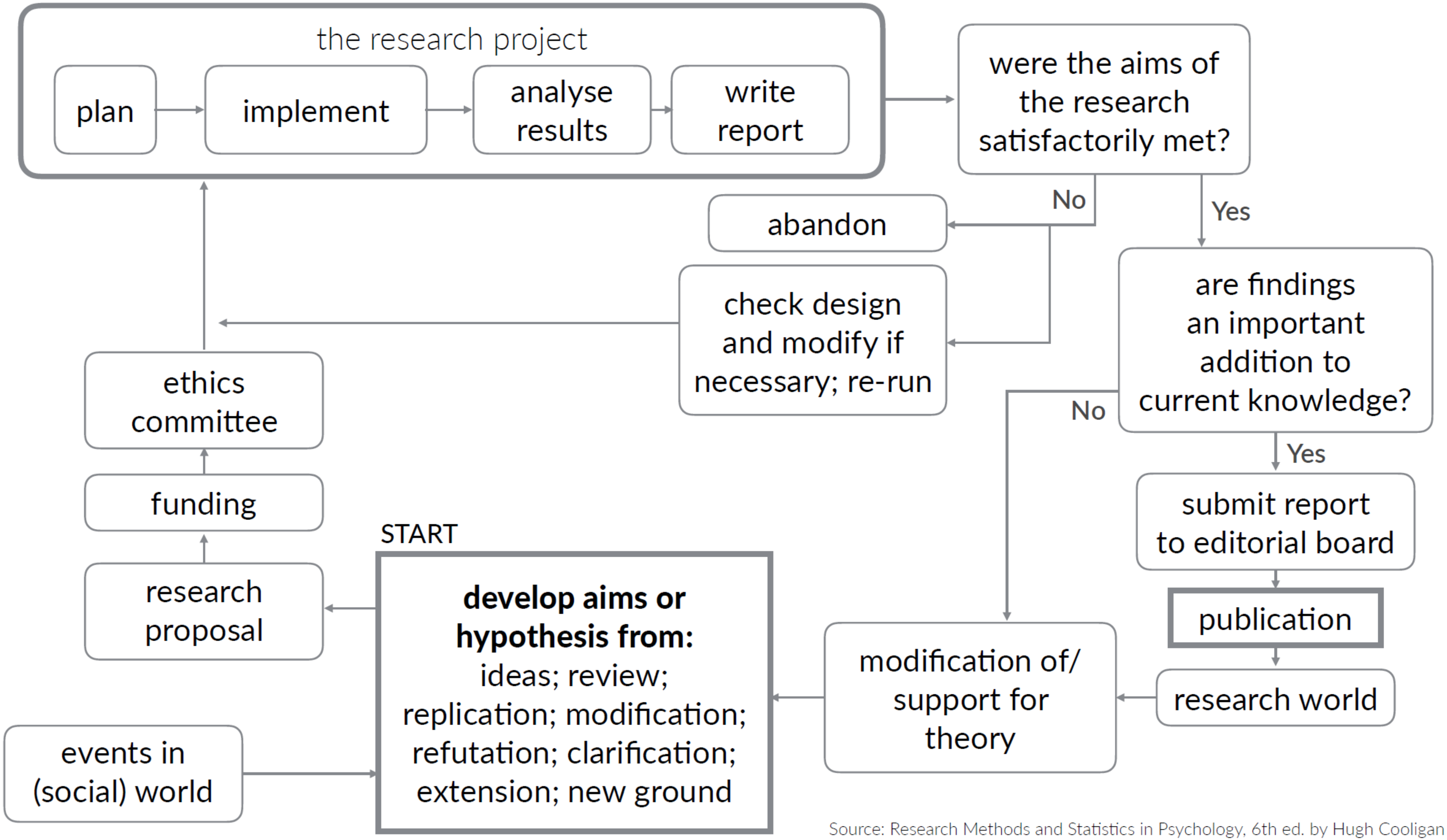
QUANTITATIVE RESEARCH METHODS

Explaining phenomena by collecting numerical data that are analysed using mathematically based methods (in particular statistics), e.g., controlled experiment, surveys.

QUALITATIVE RESEARCH METHODS

Discovering why and how people behave in the way that they do to provide in-depth information about human behaviour, e.g., observations, field studies, focus groups, interviews.

THE RESEARCH LIFE CYCLE

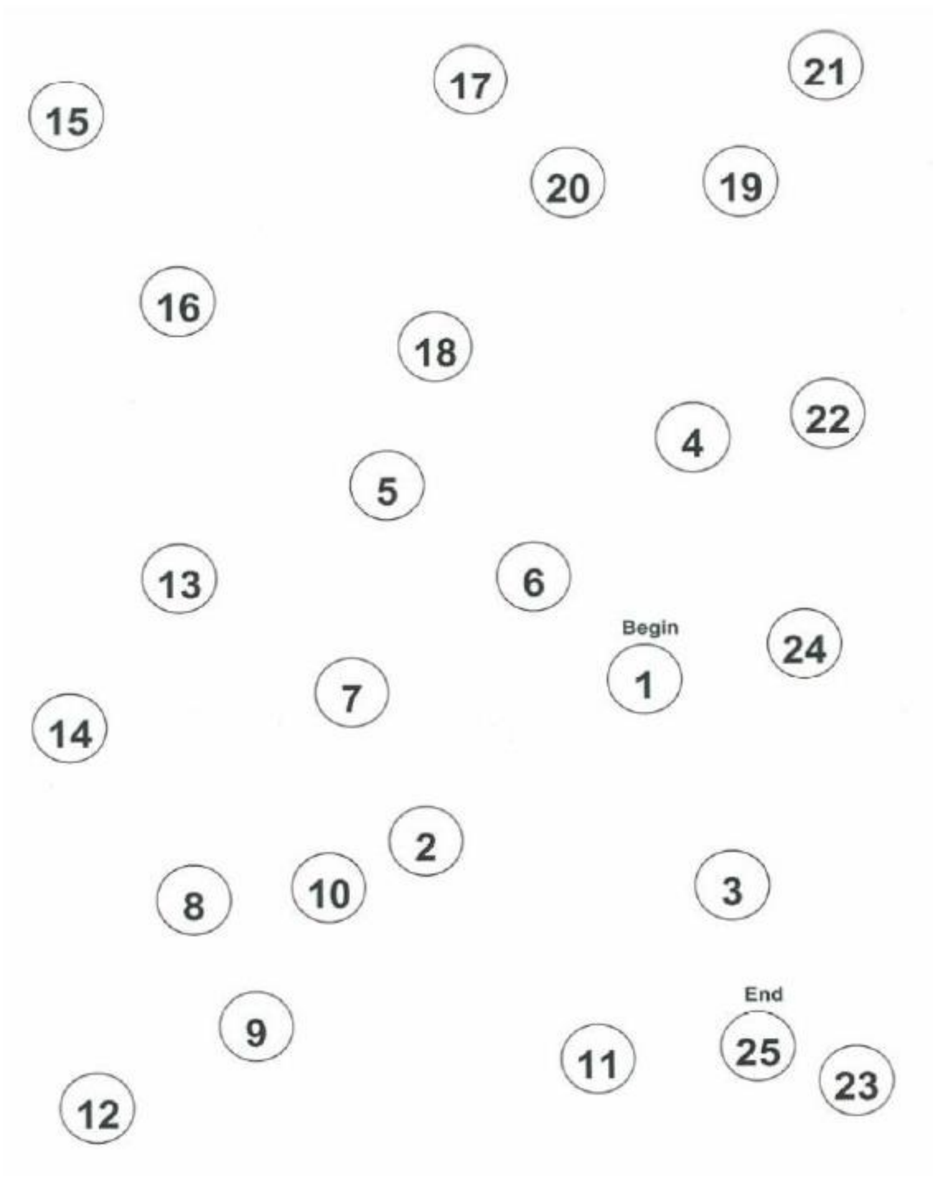


Source: Research Methods and Statistics in Psychology, 6th ed. by Hugh Cooligan

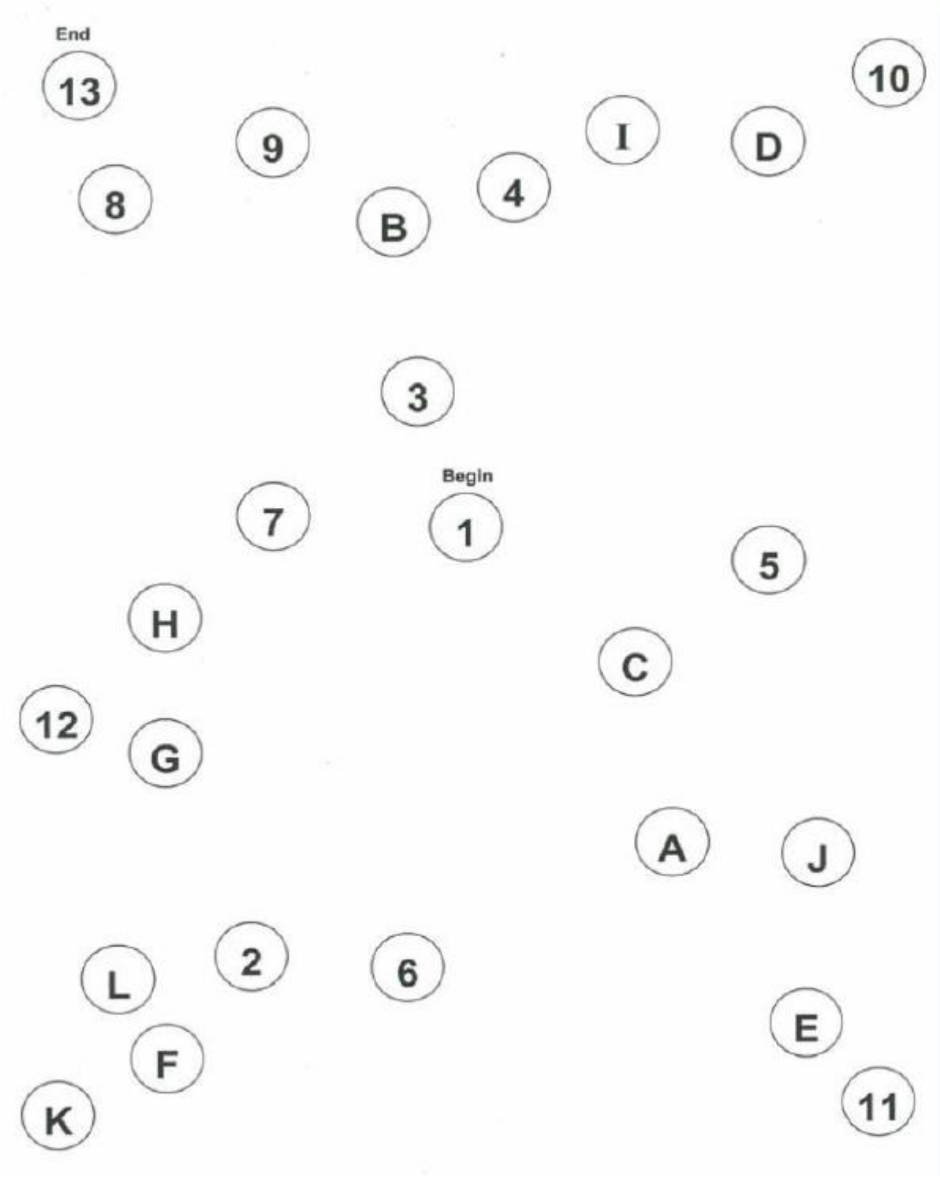
Standard procedure for running an experimental study

1. Identify a hypothesis and measures for that hypothesis
2. Specify a study design
3. Build/make any forms/docs/materials/instruments
4. Submit the study design forms to ethics procedures
5. Run a pilot
6. Recruit participants
7. Run the study
8. Analyse the data
9. Write a paper (or dissertation/etc) acknowledging the limitations and biases of the work

Trail making test part A



Trail making test part B



1. Research hypotheses

- An experiment normally starts with a research hypothesis
- A hypothesis is a precise problem statement that can be directly tested through an empirical investigation
- Compared with a theory, a hypothesis is a smaller, more focused statement that can be examined by a single experiment

Types of hypotheses

- Null hypothesis: typically states that there is no difference between experimental treatments
- Alternative hypothesis: a statement that is mutually exclusive with the null hypothesis
- A hypothesis should specify the independent variables and dependent variables

Hypotheses examples

- H0: There is no difference between the pull-down menu and the pop-up menu in the time spent locating pages
- H1: There is a difference between the pull-down menu and the pop-up menu in the time spent locating pages

- H0: There is no difference in user satisfaction rating between the pull-down menu and the pop-up menu
- H1: There is a difference in user satisfaction rating between the pull-down menu and the pop-up menu.

Variables

- Am I retaining my website users?
  - Measure the time on the site (satisfaction)
- Is this making my product faster to use
  - Measure time to complete tasks (efficiency)
- Do my users make fewer mistakes
  - Number of times that a user has to undo (error recovery)
- Can my users find features the first time they need them
  - Count the times they use help or back button (learnability)
- Do my users learn to use the interface
  - Repeat users are better at tasks than new users (memorability)

Variables

- Independent variables (IV)
  - The factors of interest or the possible “cause” of the change in the dependent variable
  - Is independent of a participant’s behavior
  - Is usually the treatments or conditions that the researchers can control
- Dependent variables (DV)
  - The outcome or effect of interest
  - Is dependent on a participant’s behavior or the changes in the IVs
  - Is usually the outcomes that the researchers need to measure

Variable examples

- H0: There is no difference between the pull-down menu and the pop-up menu in the time spent locating pages
- H1: There is a difference between the pull-down menu and the pop-up menu in the time spent locating pages

- Independent variable: the type of menu (pull-down or pop-up)
- Dependent variable: time spent in locating pages.

Typical independent variables in HCI field

- Relate to technology
  - Types of technology or device
  - Types of design
- Relate to users: age, gender, computer experience, professional domain, education, culture, motivation, mood, and disabilities
- Relate to context of use:
  - Physical status: environmental noise, lighting, temperature, vibration
  - User status: seated, walking or jogging
  - Social status: the number of people surrounding the user and their relation to the user

Typical dependent variables in HCI field

- Efficiency: how fast a task can be completed
  - e.g., task completion time, speed
- Accuracy: the states in which the system or the user makes errors
  - e.g., error rate
- Subjective satisfaction: the user's perceived satisfaction with the interaction experience.
  - e.g., Likert scale ratings
- Ease of learning and retention rate: how quickly and how easily an individual can learn to use a new application or complete a new task and how long they retain the learned skills
- Physical or cognitive demand than an application or a task exerts on an individual
  - e.g., NASA task load index

We are going to develop a pad-based TMT. Will it achieve the same purpose as the traditional one?

Hypotheses examples

- H0: There is no difference between the traditional and pad-bad TMT in the time spent completing task A.
- H1: There is a difference between the traditional and pad-bad TMT in the time spent completing task A.

Other hypotheses?

Standard procedure for running a study

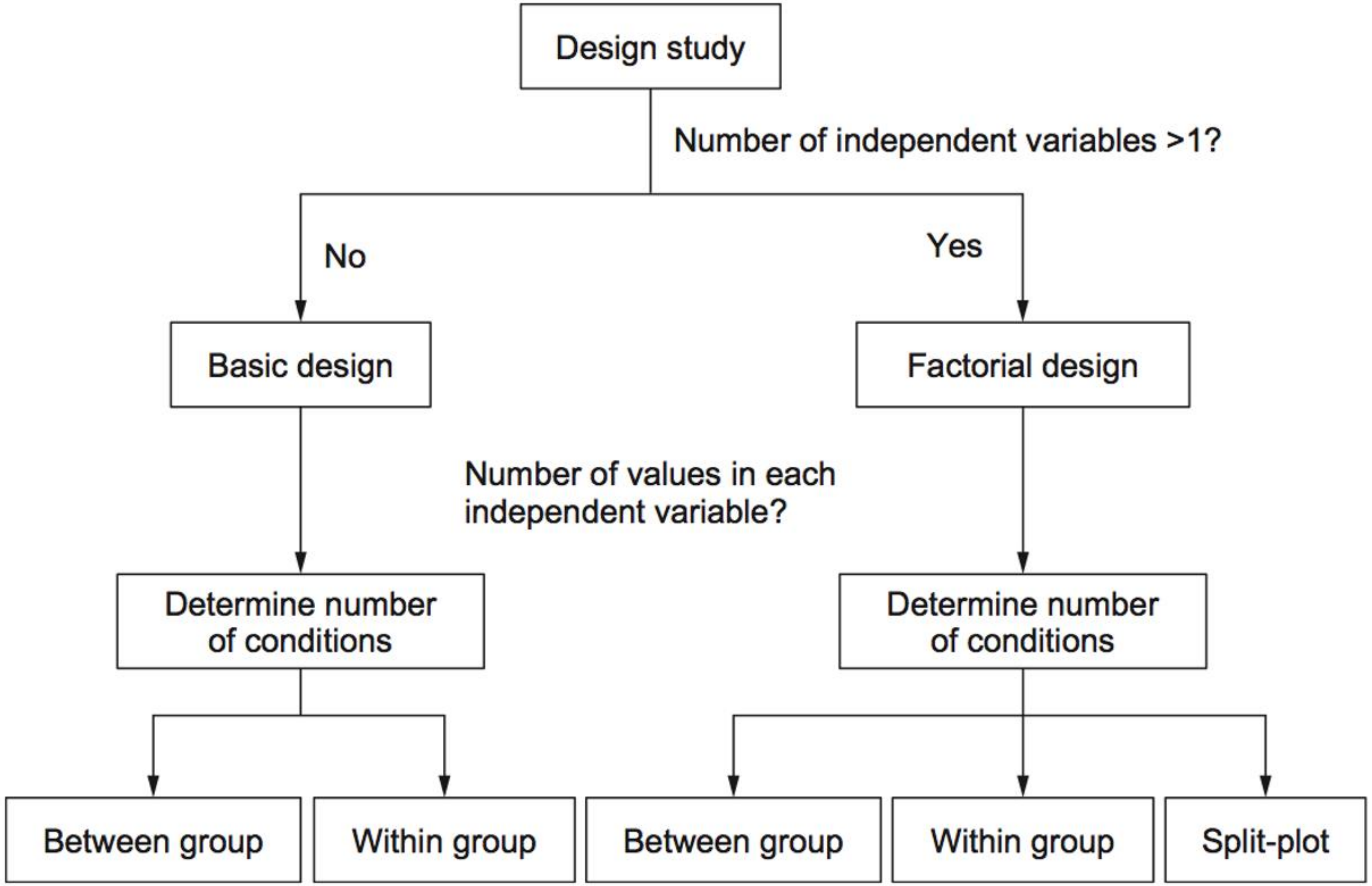
1. Identify a hypothesis and measures for that hypothesis
2. Specify a study design
3. Build/make any forms/docs/materials/instruments
4. Submit the study design forms to ethics procedures
5. Run a pilot
6. Recruit participants
7. Run the study
8. Analyse the data
9. Write a paper (or dissertation/etc) acknowledging the limitations and biases of the work

2. Experimental design

- Two basic questions:
  - How many independent variables do we want to investigate in the experiment?
  - How many different values does each independent variable have?
- The number and values of independent variables directly determine how many conditions the experiment has

There is no difference between the target selection speed when using a mouse, a joystick, or a trackball to select icons of different sizes (small, medium, and large)

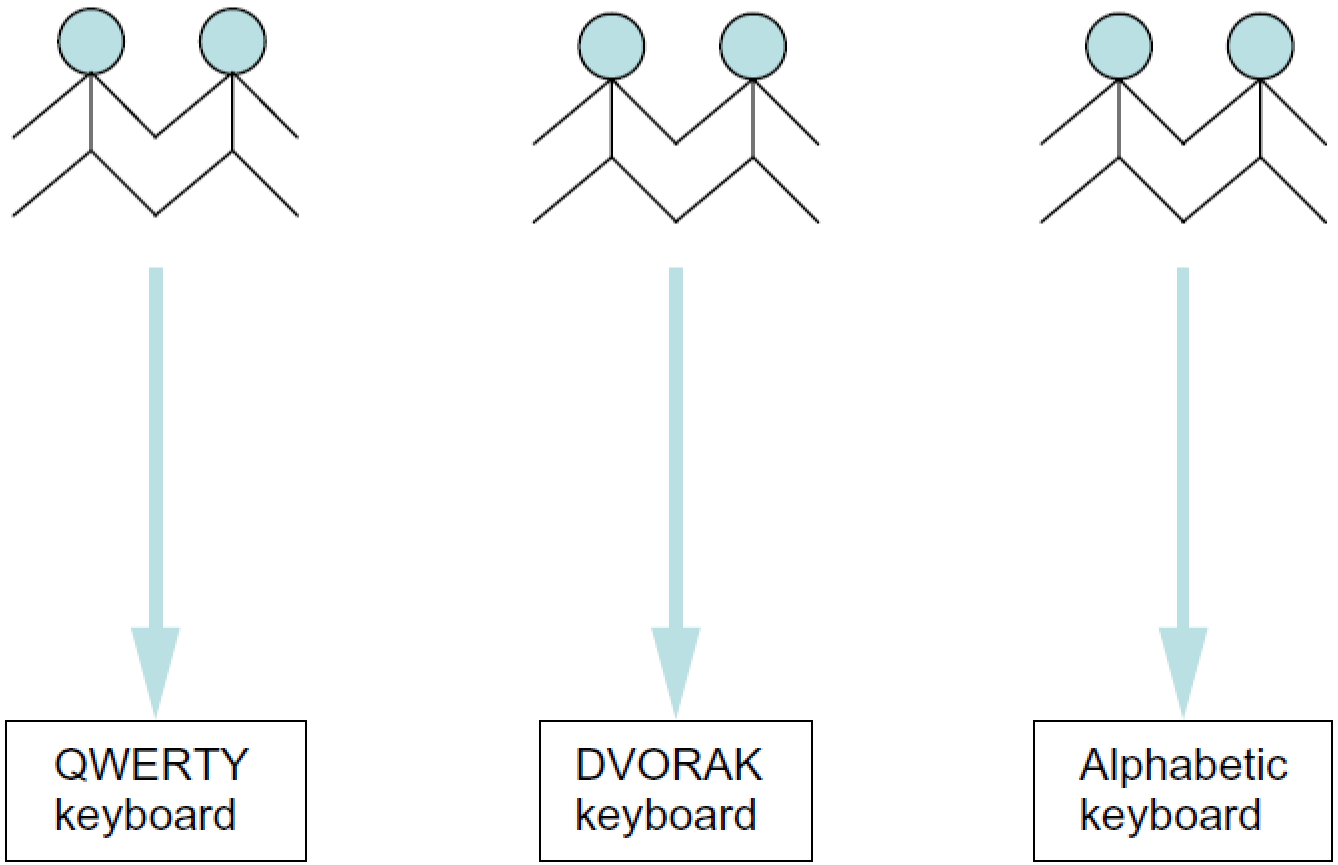
Basic design structure



Between group design

- Also called “between subject design”
- One participant only experience one condition

H0: There is no difference in typing speed when using a QWERTY keyboard, a DVORAK keyboard, or an alphabetically ordered keyboard.



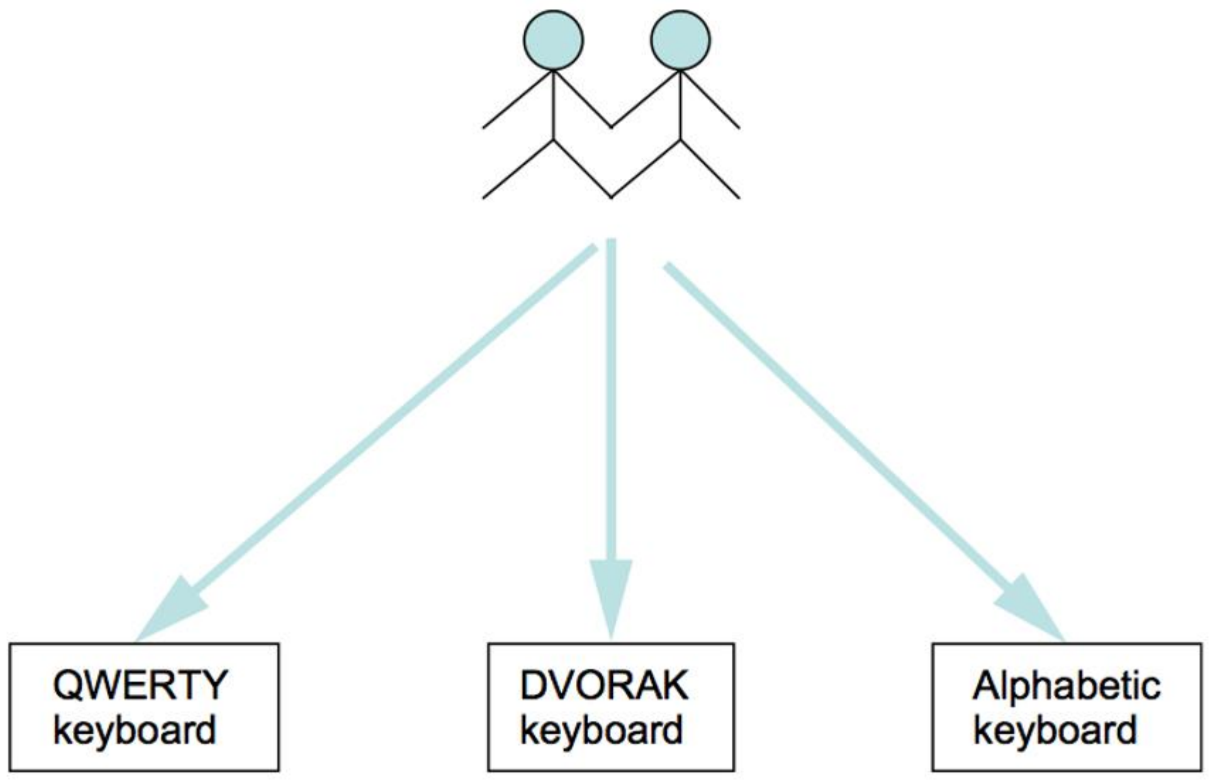
Between group design

- Advantages
  - Cleaner, better control of learning effect
  - Requires shorter time for participants
  - less impact of fatigue and frustration
- Disadvantages
  - Impact of individuals difference
  - Harder to detect difference between conditions
  - Require larger sample size

Within group design

- Also called ‘within subject design’
- One participant experience multiple conditions

H0: There is no difference in typing speed when using a QWERTY keyboard, a DVORAK keyboard, or an alphabetically ordered keyboard.



Within group design

- Advantages
  - Requires smaller sample size
  - Easy to detect difference between conditions
- Disadvantages
  - Learning effect
  - Takes longer time
  - Larger impact of fatigue and frustration

How to choose?

- Between-group design should be taken when:
  - Simple tasks
  - Learning effect has large impact
  - Within-group design is impossible, e.g., There is no difference in the time required to locate an item in an online store between novice users and experienced users.
- Within-group design should be taken when:
  - Tasks with large individual differences
  - Learning effect has small impact
  - Small participant pool

How to choose?

- H0: There is no difference between the traditional and pad-bad TMT in the time spent completing task A.

More than one independent variables

- Three options of factorial design
  - Between group design
  - Within group design
  - Split-plot design
- Split-plot design
  - Has both a between-group and a within-group component

	20–40 Years Old	41–60 Years Old	Above 60
Driving without GPS assistance	1	2	3
Driving with GPS assistance	4	5	6

Standard procedure for running a study

1. Identify a hypothesis and measures for that hypothesis
2. Specify a study design
3. Build/make any forms/docs/materials/instruments
4. Submit the study design forms to ethics procedures
5. Run a pilot
6. Recruit participants
7. Run the study
8. Analyse the data
9. Write a paper (or dissertation/etc) acknowledging the limitations and biases of the work

Materials, forms and equipment

- Consent forms
- Questionnaires (demographics, experience, etc.)
- Surveys
- Interview questions, etc.

- Computers
- Monitors
- Speakers
- Headset, etc.

- Furniture arrangement
- Lighting
- Temperature, etc.

Standard procedure for running a study

1. Identify a hypothesis and measures for that hypothesis
2. Specify a study design
3. Build/make any forms/docs/materials/instruments
4. Submit the study design forms to ethics procedures
5. Run a pilot
6. Recruit participants
7. Run the study
8. Analyse the data
9. Write a paper (or dissertation/etc) acknowledging the limitations and biases of the work

4. Plan in advance

- You need **ETHICAL APPROVAL!**
  - If you want your work to be submitted to a conference or a journal, or
  - You are collecting data (personal or sensitive)
- Consent form
  - Aware of what's going to happen (e.g., procedure)
  - Aware of any risks
  - Aware of their rights (e.g., can stop at any time)
  - How data will be stored (privately) and disposed of
  - How they can find out any results
  - Sign here to say your are happy to take part
- For more info:

  https://www.swansea.ac.uk/science/cosethics/

Standard procedure for running a study

1. Identify a hypothesis and measures for that hypothesis
2. Specify a study design
3. Build/make any forms/docs/materials/instruments
4. Submit the study design forms to ethics procedures
5. Run a pilot
6. Recruit participants
7. Run the study
8. Analyse the data
9. Write a paper (or dissertation/etc) acknowledging the limitations and biases of the work

5. Running a pilot study

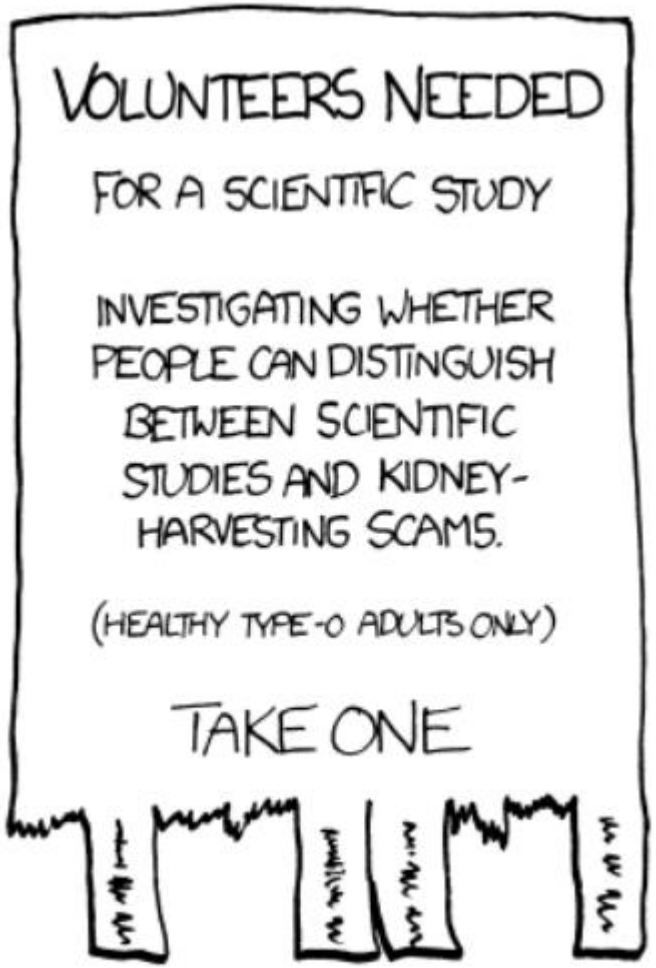
- A preliminary study/trial often carried out to predict snags and assess features of a main study to follow.
  - Permits preliminary testing of the hypotheses that lead to more refined hypotheses in the main study: modifying, dropping, and creating new hypotheses
  - Provides new idea, approaches, and clues that have been overlooked previously, increasing the chances of getting clearer findings in the main study.
  - Avoid basic errors and unforeseen cases in the main study

Standard procedure for running a study

1. Identify a hypothesis and measures for that hypothesis
2. Specify a study design
3. Build/make any forms/docs/materials/instruments
4. Submit the study design forms to ethics procedures
5. Run a pilot
6. Recruit participants
7. Run the study
8. Analyse the data
9. Write a paper (or dissertation/etc) acknowledging the limitations and biases of the work

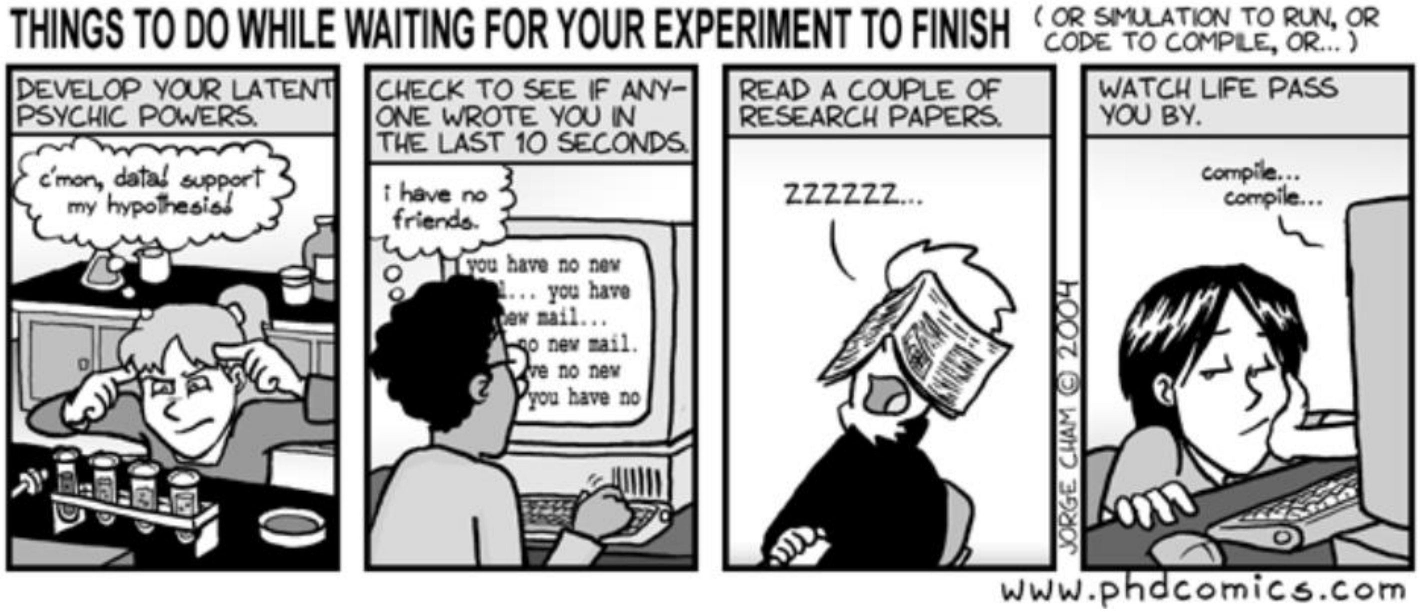
6. Recruiting participants

- Where do I find them?
  - Opportunity sampling
  - Social media channels
  - Participants database
  - Participant referral
  - Online advertisement
  - Flyers and posters
- How do I persuade them to take part
  - What is this study for (e.g., who you are, your objectives)
  - How long it will take (time, what is required)
  - What's in it for them (£ or perhaps something else?)



Standard procedure for running a study

1. Identify a hypothesis and measures for that hypothesis
2. Specify a study design
3. Build/make any forms/docs/materials/instruments
4. Submit the study design forms to ethics procedures
5. Run a pilot
6. Recruit participants
7. Run the study
8. Analyse the data
9. Write a paper (or dissertation/etc) acknowledging the limitations and biases of the work



7. Experiment session procedure

1. Preparation
2. Greet participants
3. Introduce the purpose of the study and the procedures
4. Get consent
5. Assign participants to a specific experiment condition
6. Training task(s)
7. Actual task(s)
8. Participants answer questionnaires (if any)
9. Debriefing session
10. Payment (if any)

Standard procedure for running a study

1. Identify a hypothesis and measures for that hypothesis
2. Specify a study design
3. Build/make any forms/docs/materials/instruments
4. Submit the study design forms to ethics procedures
5. Run a pilot
6. Recruit participants
7. Run the study
8. Analyse the data
9. Write a paper (or dissertation/etc) acknowledging the limitations and biases of the work

Preparing data for analysis

- Cleaning up data
  - Detect errors: age “223”
  - Formatting: “9”, “nine”, “nine and a half”, “He will turn nine in January”
- Coding
  - Types of data that need to be coded: “male”->”0”, “female”->”1”
  - Be consistent
- Organizing the data
  - Accommodate to the requirements of statistical software

Preparing data for analysis

	Age	Gender	Highest Degree	Previous Experience In Software A
Participant 1	34	Male	College	Yes
Participant 2	28	Female	Graduate	No
Participant 3	21	Female	High school	No

	Age	Gender	Highest Degree	Previous Experience In Software A
Participant 1	34	1	2	1
Participant 2	28	0	3	0
Participant 3	21	0	1	0

Descriptive statistics

- Measures of central tendency: where the bulk of the data is located
  - Mean
  - Median
  - Mode

Example: {15, 19, 22, 29, 33, 45, 50}
Mean: 30.4
Median: 29
Mode: 22

- Measures of spread: how much the data points deviate from the center of the data set
  - Range
  - Variance
  - Standard deviations

Example: {15, 19, 22, 29, 33, 45, 50}
Range: 35
Variance: 149.1
Standard deviation: 12.2

Significance tests

- H0: There is no difference between the traditional and pad-bad TMT in the time spent completing task A.
- Suppose we recruit 20 users, using within-group design.
- Can we directly compare the means?

Female: 176cm, 160cm, 180cm
Male: 165cm, 175cm, 170cm
Females are higher than males?

Significance tests

Experiment Design	Independent Variables (IV)	Conditions for each IV	Types of Test
Between-group	1	2	Independent-samples *t* test
	1	3 or more	One-way ANOVA
	2 or more	2 or more	Factorial ANOVA
Within-group	1	2	Paired-samples *t* test
	1	3 or more	Repeated measures ANOVA
	2 or more	2 or more	Repeated measures ANOVA
Between- and within-group	2 or more	2 or more	Split-plot ANOVA

- Commonly used significance tests for comparing means and their application context

Comparing 2 means: T tests

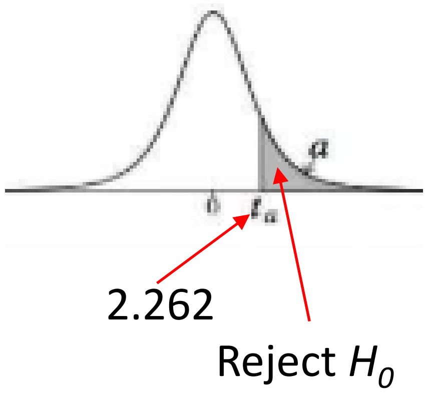
- The most widely adopted method for comparing 2 means
- The comparison will provide you with a statistic for evaluating whether the difference between two means is statistically significant.
- t test can be used either:
  - Independent-samples t test: between-group design
  - Paired-samples t test: within-group design

Sample mean and population mean

- H_0: sample mean = population mean
- $t = \frac{\bar{X} - \mu}{SE}$, where $SE = \frac{standard\ deviation}{\sqrt{N}}$
- Example: the following data represents haemoglobin values in gm/dl for 10 patients: 10.5, 9, 6.5, 8, 11, 7, 7.5, 8.5, 9.5, 12
- Degree of freedom = N-1
- Is the mean value for patients significantly differ from the mean value of general population (12gm/dl)?

Sample mean and population mean

- $t = \dfrac{8.95 - 12}{1.80201/\sqrt{10}}$ = -5.352
- Compare with tabulated value, for degree of freedom of 9, and at 95% confidence interval.
- abs(t) > tabulated value
- Reject H_0 and conclude that there is a statistically significant difference between the mean of sample and population mean, and this difference is unlikely due to chance (p<0.05).



Degrees of freedom	Significance level					
	20% (0.20)	10% (0.10)	5% (0.05)	2% (0.02)	1% (0.01)	0.1% (0.001)
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	$\boxed{2.262}$	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.043	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Two independent samples

- The following data represents weight in kg for 10 males and 12 females

Males	**80**	**75**	**95**	**55**	**60**	
	70	75	72	80	65	
Females	80	65	70	62	77	82
	60	70	50	85	45	60

H_0: there is no statistically significant difference between the mean weight of males and females

Two independent sample

- $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}(\frac{1}{n_1}+\frac{1}{n_2})}}$
- $\bar{X}_1 = 72.7, \bar{X}_2 = 67.17$
- $S_1 = 128.46, S_2 = 157.787$
- $t = 1.074$
- Degree of freedom = $n_1 + n_2 - 2$=20
- The tabulated value for significance level 1% is 2.845
- Accept H_0 and conclude that there is no significant difference, while this difference may be due to chance (p>0.01).

Degrees of freedom	Significance level 20% (0.20)	10% (0.10)	5% (0.05)	2% (0.02)	1% (0.01)	0.1% (0.001)
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.043	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.158	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Paired samples

- $t = \frac{\bar{d}}{sd/\sqrt{n}}$, where $sd = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}}$
- Blood pressure of 8 patients, before & after treatment

BP before	BP after	d	d²
180	140	40	1,600
200	145	55	3,025
230	150	80	6,400
249	155	85	7,225
170	120	50	2,500
190	130	60	3,600
200	140	60	3,600
165	130	35	1,225

Paired samples

- $t = 9.387$
- Tabulated value = 2.36 with significance level =5%, degree of freedom= 7
- Reject H_0 and conclude that there is significant difference between BP readings before and after treatment, this difference is unlikely due to chance (p<0.05).

Degrees of freedom	Significance level 20% (0.20)	10% (0.10)	5% (0.05)	2% (0.02)	1% (0.01)	0.1% (0.001)
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	$\boxed{2.365}$	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.043	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Comparing 2 or more means: analysis of variance (ANOVA)

- Also called F test
- One-way ANOVA: for empirical studies that adopt a between-group design and investigate only one independent variable with three or more conditions
- Example: *There is no significant difference in the task completion time between individuals who use the word-prediction software, using standard word-processing software, and those using speech-based dictation software.*

One-way ANOVA

Group	Participants	Task Completion Time
Standard	Participant 1	245
Standard	Participant 2	236
Standard	Participant 3	321
Standard	Participant 4	212
Standard	Participant 5	267
Standard	Participant 6	334
Standard	Participant 7	287
Standard	Participant 8	259
Prediction	Participant 9	246
Prediction	Participant 10	213
Prediction	Participant 11	265
Prediction	Participant 12	189
Prediction	Participant 13	201
Prediction	Participant 14	197
Prediction	Participant 15	289
Prediction	Participant 16	224
Speech-based dictation	Participant 17	178
Speech-based dictation	Participant 18	289
Speech-based dictation	Participant 19	222
Speech-based dictation	Participant 20	189
Speech-based dictation	Participant 21	245
Speech-based dictation	Participant 22	311
Speech-based dictation	Participant 23	267
Speech-based dictation	Participant 24	197

Source	Sum of Squares	*df*	Mean Square	*F*	Significance
Between-group	7842.250	2	3921.125	2.174	0.139
Within-group	37,880.375	21	1803.827	—	—

Denominator	Numerator DF									
DF	1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.220
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165
31	4.160	3.305	2.911	2.679	2.523	2.409	2.323	2.255	2.199	2.153
32	4.149	3.295	2.901	2.668	2.512	2.399	2.313	2.244	2.189	2.142
33	4.139	3.285	2.892	2.659	2.503	2.389	2.303	2.235	2.179	2.133
34	4.130	3.276	2.883	2.650	2.494	2.380	2.294	2.225	2.170	2.123
35	4.121	3.267	2.874	2.641	2.485	2.372	2.285	2.217	2.161	2.114
36	4.113	3.259	2.866	2.634	2.477	2.364	2.277	2.209	2.153	2.106
37	4.105	3.252	2.859	2.626	2.470	2.356	2.270	2.201	2.145	2.098
38	4.098	3.245	2.852	2.619	2.463	2.349	2.262	2.194	2.138	2.091
39	4.091	3.238	2.845	2.612	2.456	2.342	2.255	2.187	2.131	2.084
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077
41	4.079	3.226	2.833	2.600	2.443	2.330	2.243	2.174	2.118	2.071
42	4.073	3.220	2.827	2.594	2.438	2.324	2.237	2.168	2.112	2.065
43	4.067	3.214	2.822	2.589	2.432	2.318	2.232	2.163	2.106	2.059
44	4.062	3.209	2.816	2.584	2.427	2.313	2.226	2.157	2.101	2.054
45	4.057	3.204	2.812	2.579	2.422	2.308	2.221	2.152	2.096	2.049
46	4.052	3.200	2.807	2.574	2.417	2.304	2.216	2.147	2.091	2.044
47	4.047	3.195	2.802	2.570	2.413	2.299	2.212	2.143	2.086	2.039
48	4.043	3.191	2.798	2.565	2.409	2.295	2.207	2.138	2.082	2.035
49	4.038	3.187	2.794	2.561	2.404	2.290	2.203	2.134	2.077	2.030
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026

Interpretation:

- The returned *F* value of 2.174 is lower than the value at the 95% confidence interval, suggesting that there is no significant difference among the three conditions.

Factorial ANOVA

- For empirical studies that adopt a between-group design and investigate two or more independent variables
- Example: *You may also want to know whether different types of task, such as composition or transcription, have any impact on performance*

	Standard	Prediction	Speech
Transcription	Group 1	Group 2	Group 3
Composition	Group 4	Group 5	Group 6

Factorial ANOVA

Task type	Entry method	Participant Number	Task time
Transcription	Standard	Participant 1	245
Transcription	Standard	Participant 2	236
…	…	…	…
Transcription	Prediction	Participant 9	246
Transcription	Prediction	Participant 10	213
…	…	…	…
Transcription	Speech-based dictation	Participant 17	178
Transcription	Speech-based dictation	Participant 18	289
…	…	…	…
Composition	Standard	Participant 25	256
Composition	Standard	Participant 26	269
…	…	…	…
Composition	Prediction	Participant 33	265
Composition	Prediction	Participant 34	232
…	…	…	…
Composition	Speech-based dictation	Participant 41	189
Composition	Speech-based dictation	Participant 42	321
…	…	…	…
Composition	Speech-based dictation	Participant 48	202

Source	Sum of Square	*df*	Mean Square	*F*	Significance
Task type	2745.188	1	2745.188	1.410	0.242
Entry method	17,564.625	2	8782.313	4.512	0.017
Task*entry	114.875	2	57.437	0.030	0.971
Error	81,751.625	42	1946.467	—	—

Interpretation:

- There is no significant difference between participants who completed the transcription tasks and those who completed the composition tasks ($F(1, 42) = 1.41$, n.s.).
- There is significant difference among participants who used different text entry methods ($F(2, 42) = 4.51$, $p < 0.05$).

Repeated measure of ANOVA

- For within-group design
- Can investigate one or more variables
- A one-way, repeated measures ANOVA test can be used for within-group studies that investigate just one independent variable
- Example: *There is no significant difference in the task completion time between individuals who use the word-prediction software, using standard word-processing software, and those using speech-based dictation software.*

One way repeated measures ANOVA

	Standard	Prediction	Speech
Participant 1	245	246	178
Participant 2	236	213	289
Participant 3	321	265	222
Participant 4	212	189	189
Participant 5	267	201	245
Participant 6	334	197	311
Participant 7	287	289	267
Participant 8	259	224	197

Source	Sum of Square	*df*	Mean Square	*F*	Significance
Entry method	7842.25	2	3921.125	2.925	0.087
Error	18,767.083	14	1340.506	—	—

Interpretation:

The returned *F* value with degree of freedom (2, 14) is 2.925. It is below the 95% confidence interval, suggesting that there is no significant difference between the three text entry methods.

Two way repeated measures ANOVA

- Multiple-level, repeated measures ANOVA tests are needed for within-group studies that investigate two or more independent variables.
- Example: If you are interested in the impact of both the text entry method and the types of task, the study involves six conditions. A two-way, repeated measures ANOVA test can be used to analyze the data collected under this design.

	Standard	Prediction	Speech
Transcription	Group 1	Group 1	Group 1
Composition	Group 1	Group 1	Group 1

Two way repeated measures ANOVA

	Transcription			Composition		
	Standard	Prediction	Speech	Standard	Prediction	Speech
Participant 1	245	246	178	256	265	189
Participant 2	236	213	289	269	232	321
Participant 3	321	265	222	333	254	202
Participant 4	212	189	189	246	199	198
Participant 5	267	201	245	259	194	278
Participant 6	334	197	311	357	221	341
Participant 7	287	289	267	301	302	279
Participant 8	259	224	197	278	243	229

Source	Sum of Square	*df*	Mean Square	*F*	Significance
Task type	2745.187	1	2745.187	14.217	0.007
Error (task type)	1351.646	7	193.092	—	—
Entry method	17,564.625	2	8782.313	2.923	0.087
Error (entry method)	42,067.708	14	3004.836	—	—
Task type*entry method	114.875	2	57.438	0.759	0.486
Error (task type*entry method)	1058.792	14	75.628	—	—

Interpretation:

- The task type has a significant impact on the time spent to complete he task ($F(1, 7) = 14.217$, $p < 0.01$).
- There is no significant difference among the three text entry methods ($F(2, 14) = 2.923$, n.s.).
- The interaction effect between the two independent variables is not significant either ($F(2, 14) = 0.759$, n.s.).

Split-Plot ANOVA

- Involves both between-group and within-group factors
- Example: recruit two groups of participants. One group completes transcription tasks using all three data-entry methods. The other group completes composition tasks using all three data-entry methods

	Keyboard	**Prediction**	**Speech**
Transcription	Group 1	Group 1	Group 1
Composition	Group 2	Group 2	Group 2

Task Type	Participant Number	Task Type Coding	Standard	Prediction	Speech
Transcription	Participant 1	0	245	246	178
Transcription	Participant 2	0	236	213	289
Transcription	Participant 3	0	321	265	222
Transcription	Participant 4	0	212	189	189
Transcription	Participant 5	0	267	201	245
Transcription	Participant 6	0	334	197	311
Transcription	Participant 7	0	287	289	267
Transcription	Participant 8	0	259	224	197
Composition	Participant 9	1	256	265	189
Composition	Participant 10	1	269	232	321
Composition	Participant 11	1	333	254	202
Composition	Participant 12	1	246	199	198
Composition	Participant 13	1	259	194	278
Composition	Participant 14	1	357	221	341
Composition	Participant 15	1	301	302	279
Composition	Participant 16	1	278	243	229

Split-Plot ANOVA

Source	Sum of Square	*df*	Mean Square	*F*	Significance
Task type	2745.187	1	2745.187	0.995	0.335
Error	38,625.125	14	2758.937	—	—

Results of the split-plot test for the between-group variable (task type), there is no significant difference between participants who complete composition or transcription tasks ($F(1, 14) = 0.995$, n.s.)

Source	Sum of Square	*df*	Mean Square	*F*	Significance
Entry method	17,564.625	2	8782.313	5.702	0.008
Entry method*task type	114.875	2	57.437	0.037	0.963
Error (entry method)	43,126.5	28	1540.232	—	—

Results of the split-plot test for the within-group variable (text entry method), there is a significant difference among the three text entry methods ($F(2, 28) = 5.702$, $p < 0.01$). The interaction effect between task types and text entry methods is not significant ($F(2, 28) = 0.037$, n.s.).

Non-parametric tests

- Non-parametric tests are used when:
  - The data is not normally distributed
  - The distances between any two data units are not equal

I am satisfied with the time it took to complete the task

1	2	3	4	5
highly disagree	disagree	neutral	agree	highly agree

  - The variance of error is not equal

Non-parametric tests

- χ-squared test: used to analyze categorical data, e.g., race or gender, yes or no
- No parametric tests for comparing means
  - comparing two groups of data:
    - For between-group design:
      - Mann-Whitney U test: non-parametric alternative to the independent-samples t test
      - Wald-Wolfowitz runs test: nonparametric alternative to the paired-samples t test
    - For within-group design: Wilcoxon signed ranks test
  - Comparing three or more groups of data
    - For between-group design:
      - Kruskal-Wallis one-way analysis of variance by ranks: nonparametric alternative to the one-way ANOVA
    - For within-group design:
      - Friedman's two-way analysis of variance test: Nonparametric alternative to the Repeated Measures ANOVA

Chi-squared test

- You are examining the impact of age on users' preferences toward two target selection devices: a mouse and a touchscreen. You recruit two groups of users. One group consists of 20 adult users who are younger than 65 and the other consists of 20 users who are 65 or older. After completing a series of target selection tasks using both the mouse and the touchscreen, participants specify the type of device that they prefer to use.

Age	Preferred Device	
	Mouse	Touchscreen
<65	14	6
≥65	4	16

$\chi^2(1) = 10.1, p < 0.005$. It suggests that the probability of the difference between the rows and columns occurring by chance is less than 0.005. Using the 95% confidence interval, you reject the null hypothesis and conclude that there is a relationship between age and preferred pointing device.

Finally

- H0: There is no difference between the traditional and pad-bad TMT in the time spent completing task A.

Experiment Design	**Independent Variables (IV)**	**Conditions for each IV**	**Types of Test**
Between-group	1	2	Independent-samples *t* test
	1	3 or more	One-way ANOVA
	2 or more	2 or more	Factorial ANOVA
Within-group	1	2	Paired-samples *t* test
	1	3 or more	Repeated measures ANOVA
	2 or more	2 or more	Repeated measures ANOVA
Between- and within-group	2 or more	2 or more	Split-plot ANOVA

Standard procedure for running a study

1. Identify a hypothesis and measures for that hypothesis
2. Specify a study design
3. Build/make any forms/docs/materials/instruments
4. Submit the study design forms to ethics procedures
5. Run a pilot
6. Recruit participants
7. Run the study
8. Analyse the data
9. Write a paper (or dissertation/etc) acknowledging the limitations and biases of the work (Later sessions)