

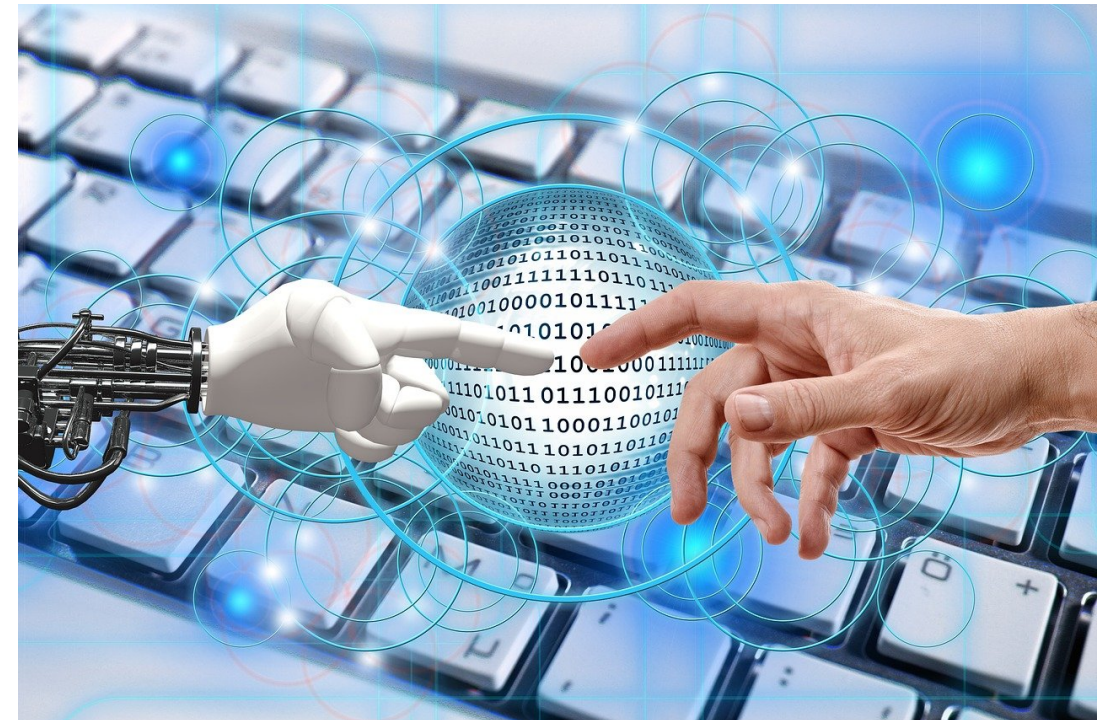
Power Analysis

Prof. Dr. Teena Hassan
teena.hassan@h-brs.de

Department of Computer Science

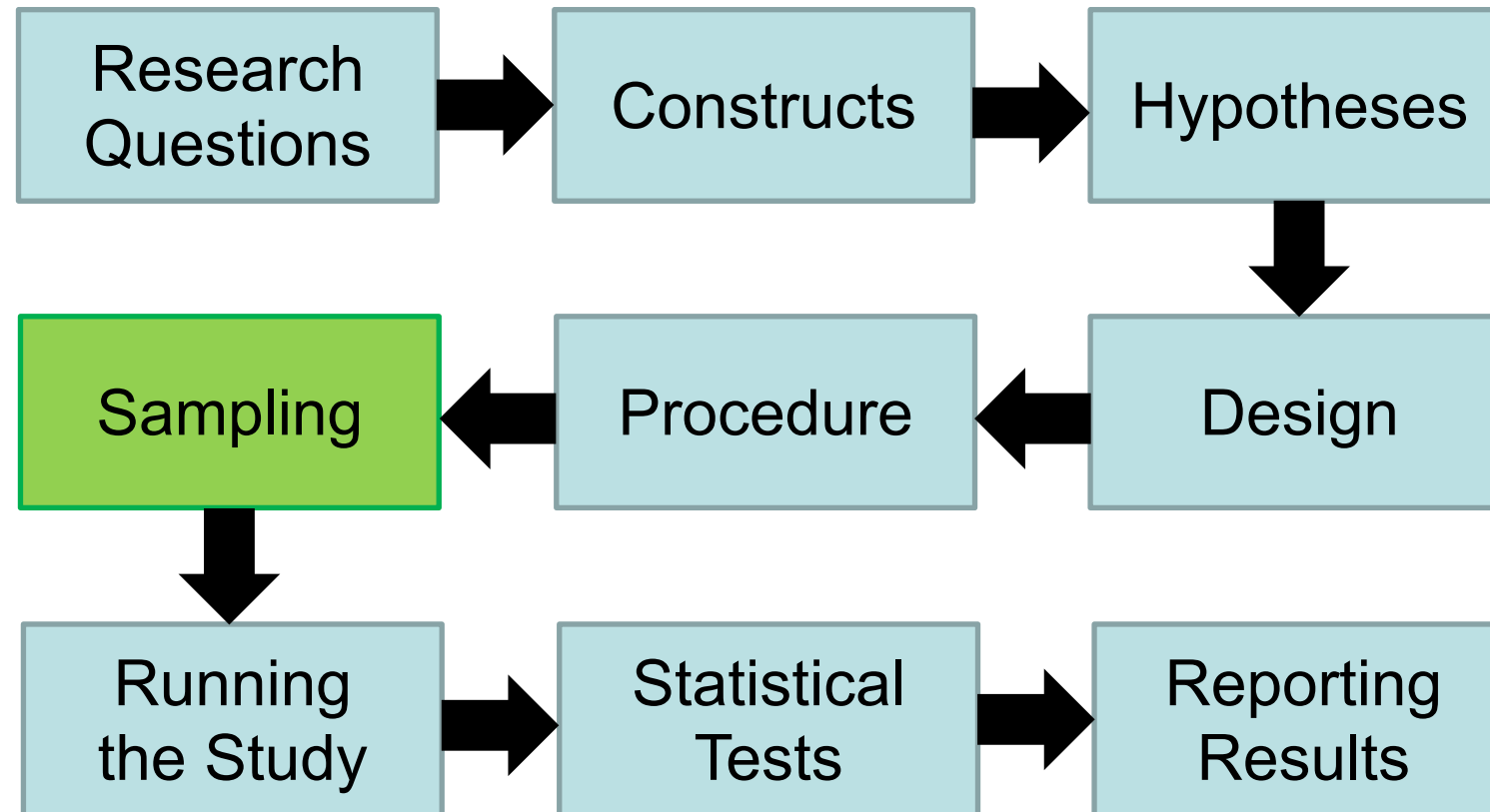
Hochschule Bonn-Rhein-Sieg
Sankt Augustin

27 June 2024



- At the end of today's lecture, you will be able to:
 1. Identify the factors that influence the **sample size** of participants required for an experiment.
 2. Apply **power analysis** to determine sample size for validating hypotheses in HRI experiments.

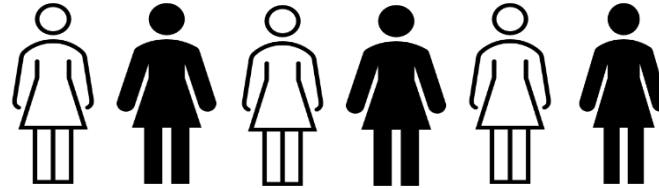
- Guy Hoffman and Xuan Zhao. 2020. A Primer for Conducting Experiments in Human Robot Interaction ACM Trans. Hum. Robot Interact . 10, 1, Article 6 (October 2020), 31 pages <https://doi.org/10.1145/3412374>
- Your reading assignment for last week!



(Hoffman & Zhao, 2020)

➤ Which design would require more samples to achieve the same statistical power?

A. Within-subjects design

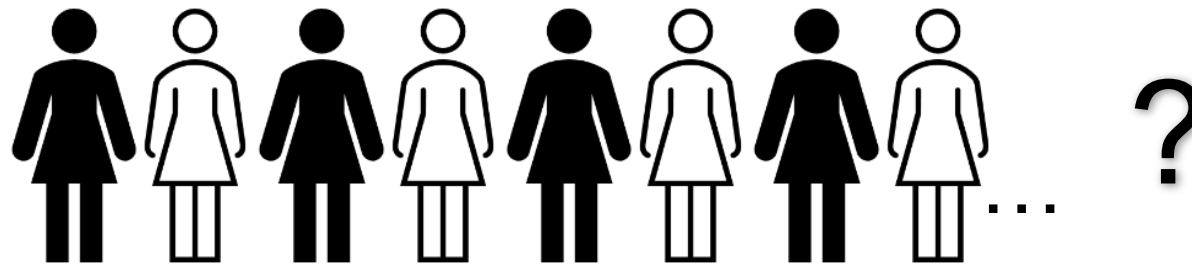


B. Between-subjects design



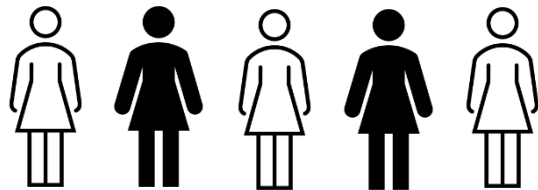
How Many Samples are Enough?

- Very crucial question in empirical research.
- How many samples are needed to draw scientifically sound conclusions?



Power Analysis

Simply put, a **statistically sound** method to determine **the sample size** and **power** of an experiment.



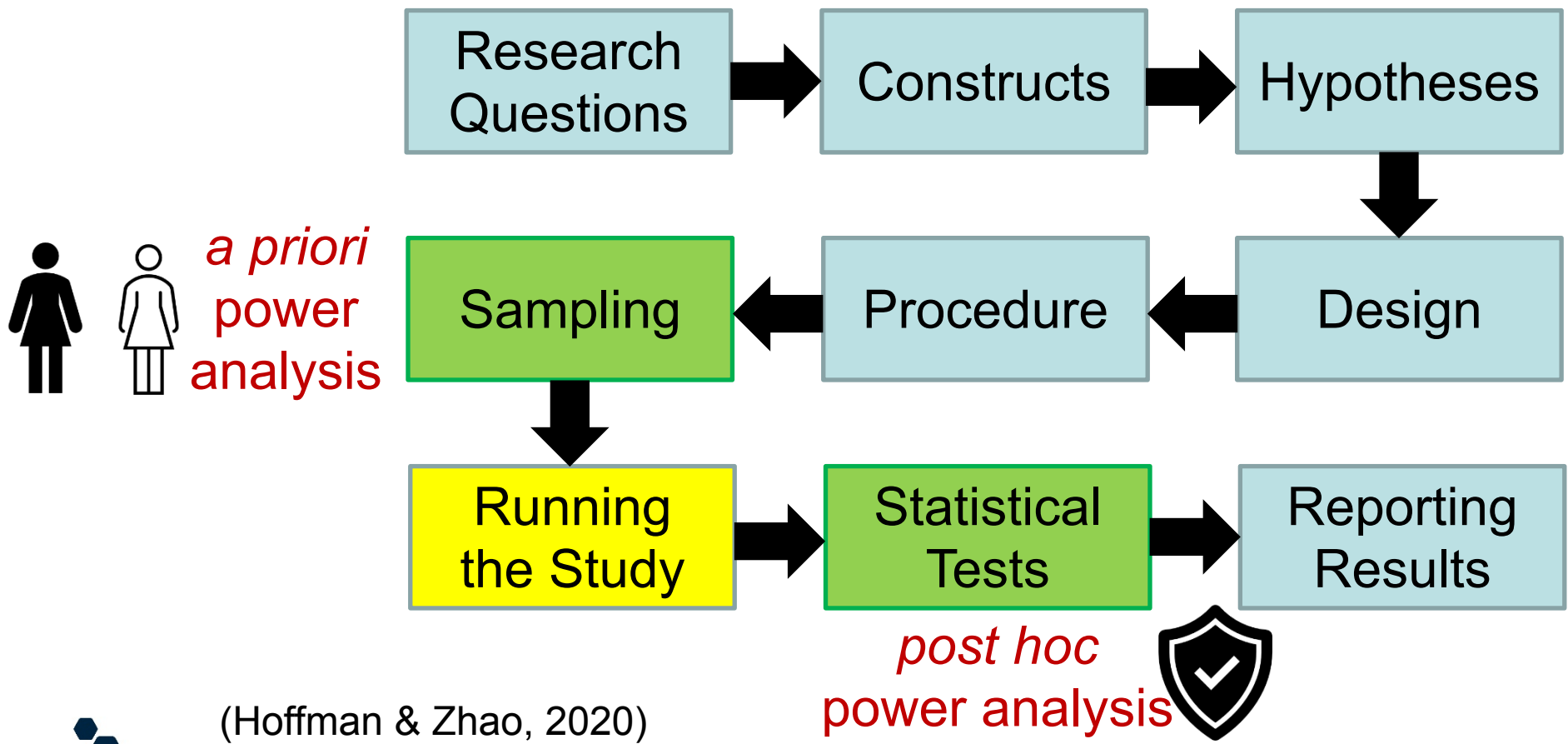
Sample size



Power

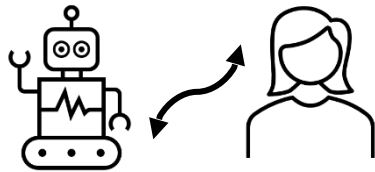
- Humans are **unpredictable** and **exhibit variability** in their behavior.
 - Inter- and intrapersonal variability
- When we conduct an experiment:
 - How can we be sure that the **effects that we observed are meaningful** and not noise?
- **We should perform a priori and post hoc power analyses.**

(Bartlett et al., 2022)



(Hoffman & Zhao, 2020)

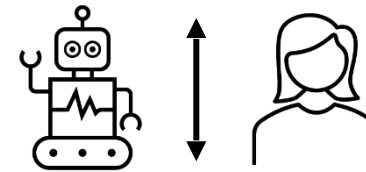
Level of users' trust in:



Cond-A

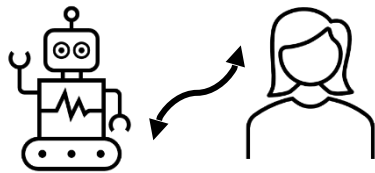
>

Level of users' trust in:



Cond-B

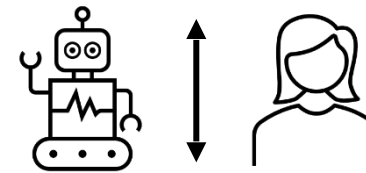
Level of users' trust in:



Cond-A

=

Level of users' trust in:

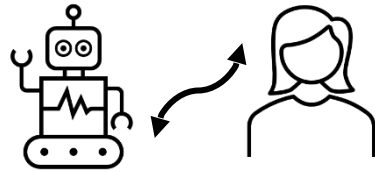


Cond-B

- **H_0** posits there is no difference between users' level of trust in cond-A and cond-B.

- Can we reject H0:

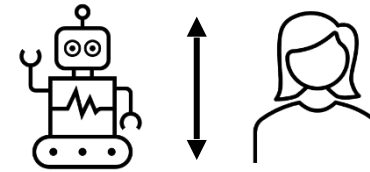
Level of users' trust in:



Cond-A

=

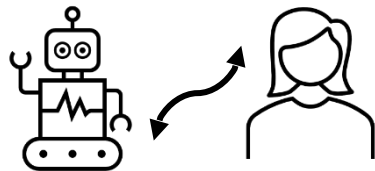
Level of users' trust in:



Cond-B

- And accept H1:

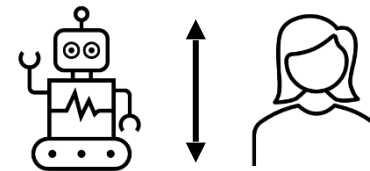
Level of users' trust in:



Cond-A

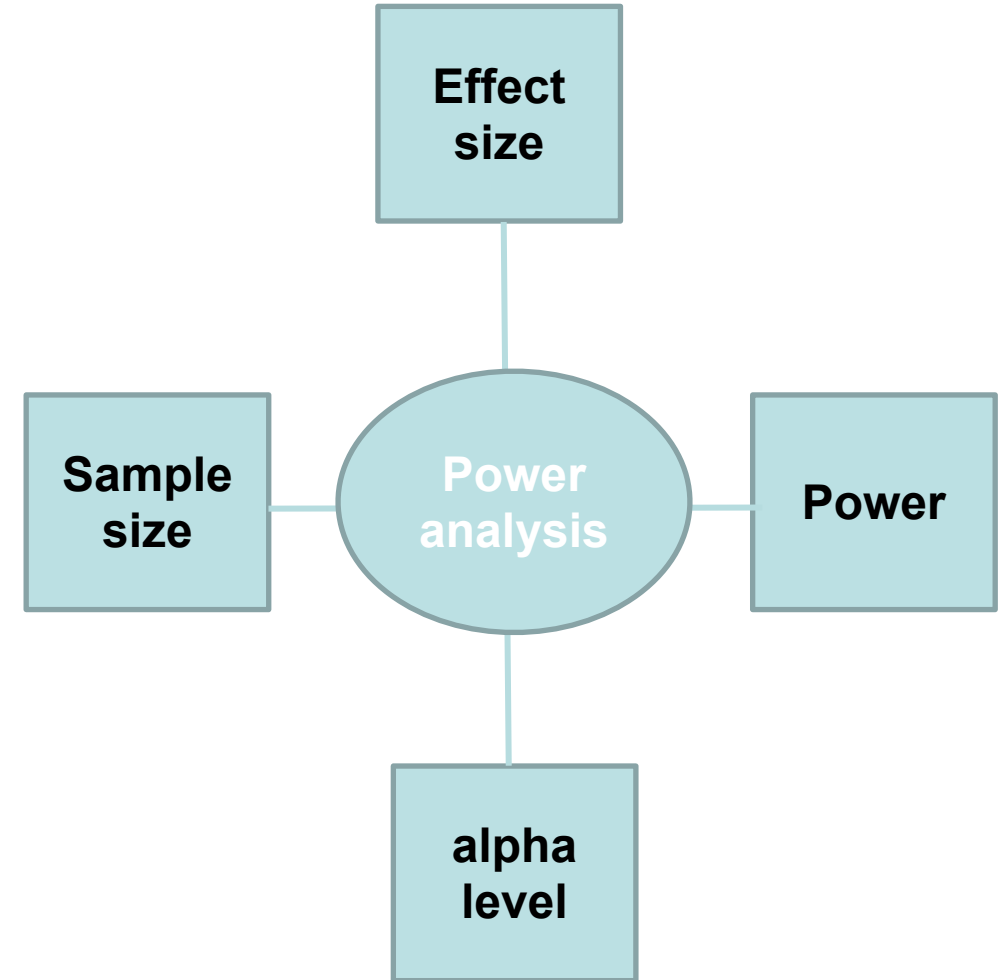
>

Level of users' trust in:

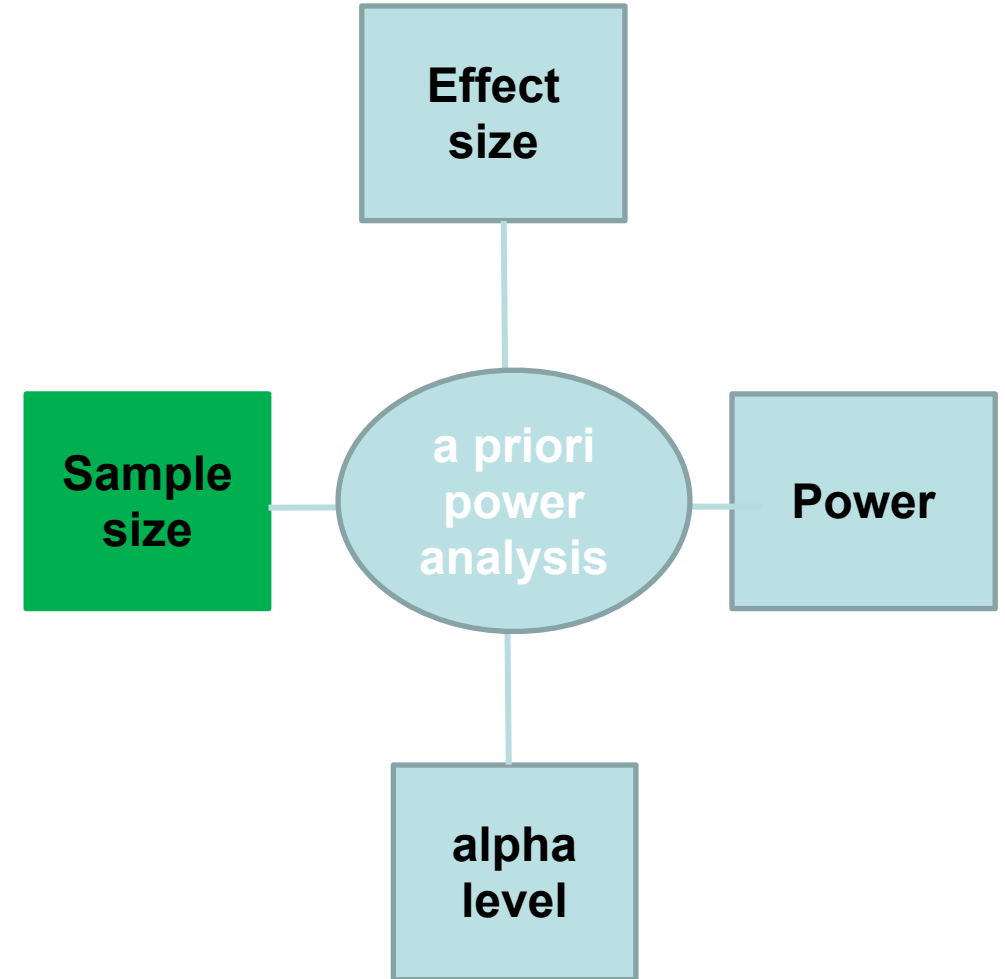


Cond-B

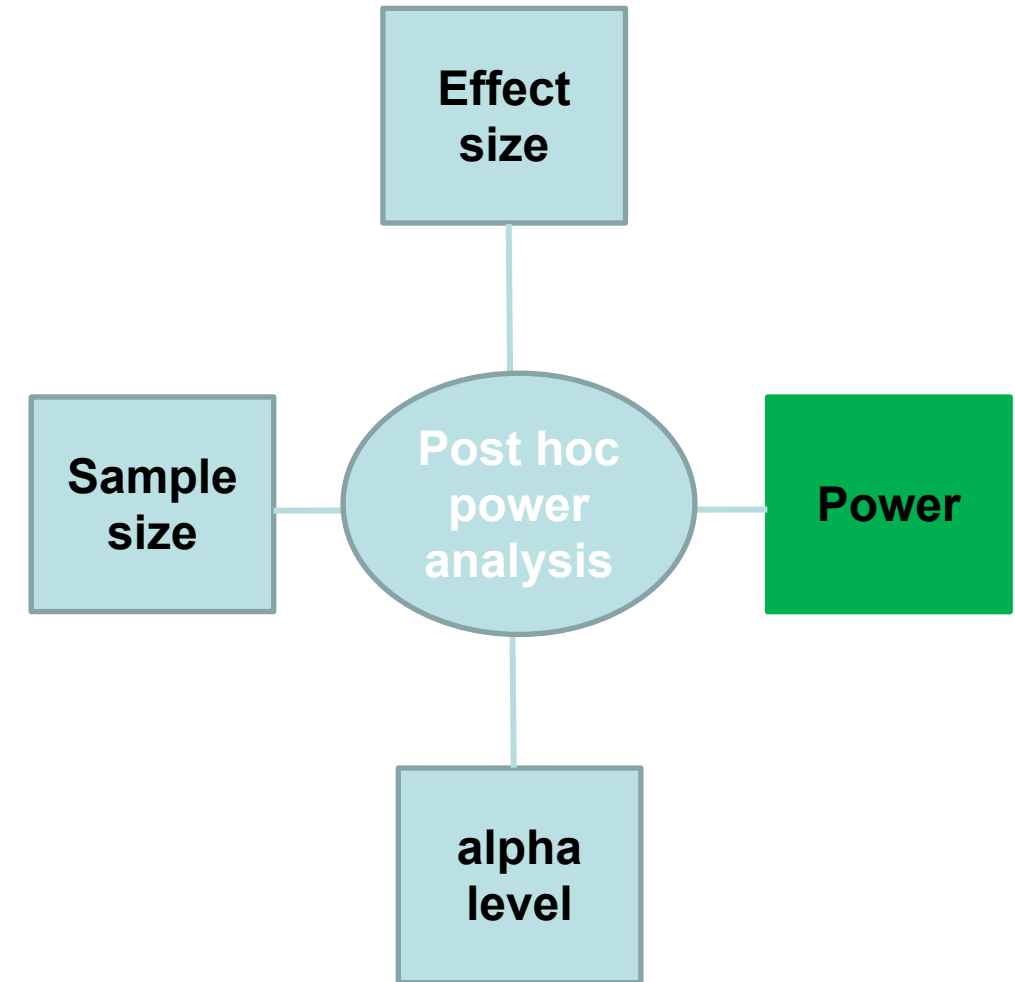
- Examines the relationship between **four parameters**:
 1. alpha level
 2. Power
 3. Effect size
 4. Sample size
- Given any three of these parameters, we can predict the fourth.

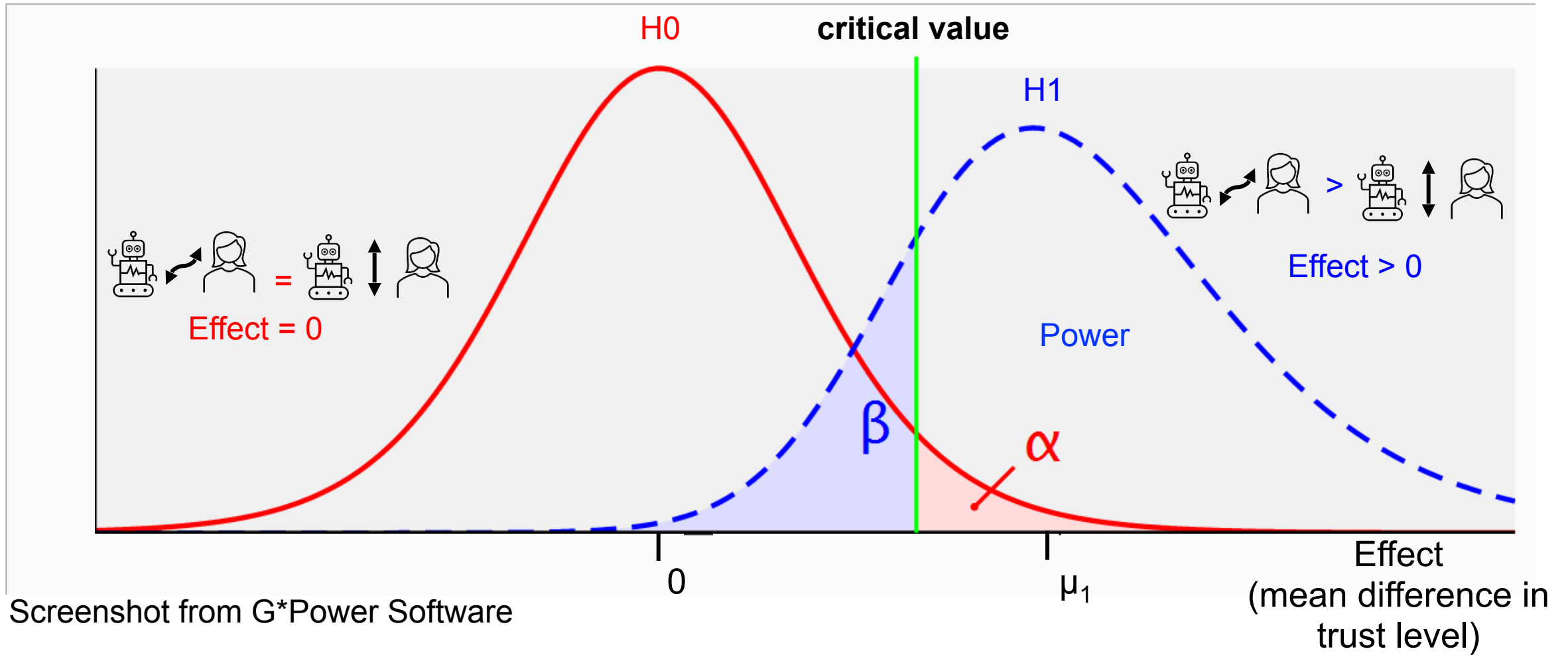


- Given:
 - Expected or desired values for:
 1. **alpha level**
 2. **Power**
 3. **Effect size**
- Predict, **sample size** needed for the experiment.



- Done **after** the experiment.
- Given:
 - Values computed from experimental data:
 1. **Effect size**
 2. **Sample size**
 3. **alpha level**
- Predict **statistical power** of the experiment.





Screenshot from G*Power Software

To get **the same power** after reducing alpha, we should **increase the sample size**.

	Reject H0 Accept H1	Accept H0 Reject H1
H0 is True H1 is False	α False positive rate (Type I error)	$1 - \alpha$ True negative rate
H0 is False H1 is True	$1 - \beta$ True positive rate (POWER)	β False negative rate (Type II error)

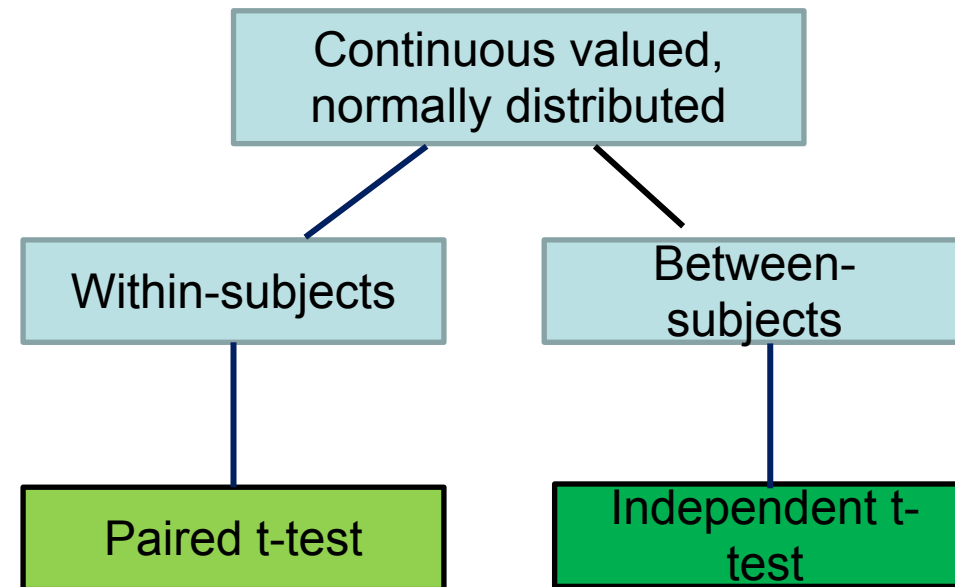


Holds in the general population



Holds in the sampled population

- **Effect size is the difference between outcomes measured in different conditions.**
- Effect size computation depends on the method chosen for the statistical test.



- Independent t-test

- Between subject
 - Two separate groups of subjects A and B
 - Each group exposed to only one condition
 - Two sampling distributions
- Computes means of dependent variable in either distribution: μ_A , μ_B
- Difference between means: $\Delta = \mu_A - \mu_B$
- Effect: Compares Δ to zero.
- Effect size (Cohen's d): Δ / σ_{pooled}

- Paired t-test

- Within subject
 - One group of subjects
 - Each subject exposed to both conditions A and B
- One sampling distribution S based on difference Δ_s in response of each subject s in condition A and condition B.
 - $\Delta_s = a_s - b_s$
 - $\mu_S = \text{Mean } \Delta_s \text{ over all } s$
- Effect: Compares μ_S to zero.
- Effect size (Cohen's d): μ_S / σ_S

$$\sigma_{pooled} = \sqrt{(\sum(a - \mu_A)^2 + \sum(b - \mu_B)^2) / (n_A + n_B - 2)}$$

- Commonly used values:
 - alpha: 0.05 or 0.01
 - power: 0.8 (HRI), 0.95 (clinical)
 - effect-size (for t-tests): 0.5

	Small effect size	Medium effect size	Large effect size
Cohen's d	0.2	0.5	0.8

Jacob Cohen. 1992. A power primer. Psychol. Bull. 112, 1 (1992).

- In this lecture, you learned to:
 1. Identify the factors that influence the **sample size** of participants required for an experiment.
 2. Apply **power analysis** to determine sample size for validating hypotheses in HRI experiments.

- Guy Hoffman and Xuan Zhao. 2020. **A Primer for Conducting Experiments in Human–Robot Interaction**. ACM Trans. Hum.-Robot Interact. 10, 1, Article 6 (October 2020), 31 pages. <https://doi.org/10.1145/3412374>
- Madeleine E. Bartlett, C. E. R. Edmunds, Tony Belpaeme, and Serge Thill. 2022. **Have I Got the Power? Analysing and Reporting Statistical Power in HRI**. *J. Hum.-Robot Interact.* 11, 2, Article 16 (June 2022), 16 pages. <https://doi.org/10.1145/3495246>

- Lecture on “Statistical Tests”
 - Thursday, 4.07.2024 at 9 am in B060, Sankt Augustin