Two types of power analysis:

1. A priori power analysis – determines required sample size
2. Sensitive power analysis – determines what level of effect you could find with the subjects you have

Software needed:

* G\*power - <https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.html>
* Conversion calculator (section 14) - <https://www.psychometrica.de/effect_size>

**A priori power analysis using G\*power**

Used to estimate the number of participants

* More participants = more power

Big effect = low power needed (The bigger an object is, the easier it is to find it)

Small effect = high power needed (The smaller an object is, the harder it is to find it)

**G\*power inputs/assumptions:**

* Power level: 80% (0.80) is a default.

80% power means that you have a 20% chance for a type 2 error\*. So go with a higher power when you can (i.e., **95%)** (Lakens, 2013)

* Alpha (significance threshold: default = **0.05**)
* Effect size estimate (the effect size you are looking for; see appendix 2)
* Study design and planned test(s): e.g., type of analysis, number of conditions/groups, number of predictors, etc.

\*Type 2 error: Failing to find a significant effect where there is one. The higher the power, the lower the chance for a type 2 error.

**A priori power analysis in G\*power: ANOVA walkthrough**

**One-way ANOVA:**

1. Test family: F tests
2. Statistical test: ANOVA: Fixed effects, omnibus, one-way
3. Type of power analysis: A priori…
4. Effect size: eta-squared (η2) (see appendix 2)

* Click determine – click on ‘Effect size from means’ and switch to ‘Effect size from variance’
* Click on ‘Direct’ – input the value of Partial eta-squared (ηp2)
* Click on ‘Calculate and transfer to main window’ (this will transfer your η2 to Cohen’s f).

1. Alpha error probability = 0.05
2. Power = 0.8 or 0.95 (depends on how accurate you want to be, both are fine)
3. Number of groups = number of conditions in your study
4. Click on ‘calculate’ - look at ‘Total sample size’ to find the required number of participants.

**Factorial (or two-way) ANOVA:**

1. Test family: F tests
2. Statistical test: ANOVA: Fixed effects, special, main effects and interactions
3. Type of power analysis: A priori
4. Effect size: eta-squared (η2) (see appendix 2)

* Click ‘determine’ – click on ‘Effect size from means’ and switch to ‘Effect size from variance’
* Click on ‘Direct’ – input the value of Partial η2 (ηp2)
* Click on ‘Calculate and transfer to main window’

1. Alpha error probability = 0.05
2. Power = 0.8 or 0.95
3. Numerator df: use the highest value calculated (see appendix 3)
4. Number of groups = number of conditions
5. Click on calculate - look at ‘Total sample size’ to find the required number of participants.

**One-way repeated measures ANOVA (within-factors):**

1. Test family: F tests
2. Statistical test: ANOVA: Repeated measures, within factors
3. Type of power analysis: A priori
4. Click ‘options’ – select ‘as in SPSS’ – click ok
5. Effect size: eta-squared (η2) (see appendix 2)

* Click ‘determine’ – click on ‘Effect size from means’ and switch to ‘Effect size from variance’
* Click on ‘Direct’ – input the value of Partial η2 (ηp2)
* Click on ‘Calculate and transfer to main window’

1. Alpha error probability = 0.05
2. Power = 0.8 or 0.95
3. Number of groups = 1
4. Number of measurements = Number of dependent variables
5. Nonsphericity correction e = If you have <2 measures then put 1, if you have >2 then put 0.75 (see appendix 4)
6. Click on calculate - look at ‘Total sample size’ to find the required number of participants.

**Mixed ANOVA (repeated measures ANOVA with between-factors):**

1. Test family: F tests
2. Statistical test: ANOVA: Repeated measures, within-between interaction
3. Type of power analysis: A priori
4. Click ‘options’ – select ‘as in SPSS’ – click ok
5. Effect size: eta-squared (η2) (see appendix 2)

* Click ‘determine’ – click on ‘Effect size from means’ and switch to ‘Effect size from variance’
* Click on ‘Direct’ – input the value of Partial η2 (ηp2)
* Click on ‘Calculate and transfer to main window’

1. Power = 0.8 or 0.95
2. Number of groups = enter the number of groups/conditions that you have
3. Number of measurements = number of dependent variables
4. Nonsphericity correction e = If you have <2 measures then put 1, if you have >2 then put 0.75 (see appendix 4)

**A priori power analysis in G\*power: t-tests walkthrough**

**Independent-samples t-test (between-subjects):**

1. Test family: t tests
2. Statistical test: Means: Difference between two independent means (two groups)
3. Type of power analysis: A priori
4. Tail(s): two (see appendix 5)
5. Effect size: d (see appendix 2)
6. Alpha error probability = 0.05
7. Power = 0.8 or 0.95
8. Allocation ratio N2/N/1: We can assume that we want equally sized sample groups (an allocation ratio of 1)

**Dependent-samples t-tests (within-subjects)**

1. Test family: t tests
2. Statistical test: Means: Difference between two dependent means (matched pairs)
3. Type of power analysis: A priori
4. Tail(s): two (see appendix 5)
5. Effect size: d (See appendix 2)
6. Alpha error probability = 0.05
7. Power = 0.8 or 0.95

**A priori power analysis in G\*power: Linear multiple regression walkthrough**

**Linear multiple regression**

1. Test family: F tests
2. Statistical test: Linear multiple regression: Fixed model, R2 increase
3. Type of power analysis: A priori
4. Effect size: eta-squared (η2) (See appendix 2)

* Click ‘determine’ – click on ‘Direct’
* Input eta-squared value
* Click ‘calculate’ – click ‘calculate and transfer to main window’

1. Alpha error probability = 0.05
2. Power = 0.8 or 0.95
3. Number of tested predictors: these are your dependent variables that are also used as predictors
4. Total number of predictors: these include both predictors form independent and dependent variables

Predictors: these are usually your independent variables, but in a way, they can also be dependent variables if you are looking to see how a measured effect influences another measured effect.

**Linear multiple regression t-tests**

T-tests are also used for assessing the significance of individual predictors in a multiple regression analysis.

1. Test family: t tests
2. Statistical test: linear multiple regression: Fixed model, single regression coefficient
3. Type of power analysis: A priori
4. Tail(s): two (see appendix 5)
5. Effect size: eta-squared (η2) (see appendix 2)

* Click ‘determine’ – click on ‘Direct’
* Input eta-squared value
* Click ‘calculate’ – click ‘calculate and transfer to main window’

1. Alpha error probability = 0.05
2. Power = 0.8 or 0.95
3. Number of predictors

**Sensitivity power analysis in G\*power**

* Used to determine what level of effect size you can detect given a specified number of participants.
* Basically, the sensitivity power analysis detects the effect size for a given power.

It does the opposite of a priori power analysis.

* A priori: detects the sample size for a given effect size
* Sensitivity: detects the effect size for a given sample size

**The utility of a sensitivity power analysis:**

If you ran a sensitivity power analysis with 80% power level, you could take the effect size produced and compare it to the effect size that you got from your actual analysis (when analyzing your study).

If your observed effect size from your study is **smaller** from the effect size of the power analysis, then your study had less than 80% power to detect that effect (it was underpowered – did not have enough participants).

If your study had less than 80% power then it means that the chance for a type 2 error was higher than 20%, meaning that your findings will be harder to replicate (if you re-run the study with the same sample).

**G\*power inputs/assumptions:**

* Power level: 80% or 95%
* Alpha: 0.05
* Study design and planned test(s): e.g. type of analysis, number of conditions/groups, number of predictors, etc.
* Obtained sample size

The procedure is very similar to the one used in a priori power analysis. The only differences are:

* It asks you to input the number of participants instead of effect size
* Type of power analysis: Sensitivity: Compute required effect size – given a, power, and sample size
* You will need to use the conversion calculator to convert the effect sizes (shown below).

**For ANOVAs**

Effect size produced: ‘f’ – convert to eta-squared (η2)

* Go to the website (<https://www.psychometrica.de/effect_size>) in section 14 (transformation of the effect sizes d, r, f…)
* In the ‘effect size’ select ‘f’
* Copy the ‘effect size f’ value you got from G\*power and paste it into the calculator
* Look for ‘η2’ - that is your effect size (see appendix 2 for what the effect means)

**For t-tests**

Effect size produced: ‘d’ – no need to convert

**For Linear multiple regressions**

Effect size produced: ‘f2’– convert to r2

* Calculate the square root of f2 to find f.
* Go to the website (<https://www.psychometrica.de/effect_size>) in section 14 (transformation of the effect sizes d, r, f…)
* In the ‘effect size’ select ‘f’
* Copy the f value you calculated and paste it into the calculator
* Look for ‘r’

Simply put, R is the correlation between the predicted values and the observed values. R square is the square of this coefficient and indicates the percentage of variation explained by your regression line out of the total variation.

**References**

Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology, 4*, 863.

Other sources:

<https://www.youtube.com/watch?v=5J4IKzDfT0w&feature=youtu.be>

<http://www.mormonsandscience.com/gpower-guide.html>

**Appendix 1**

**How do you know what effect size to look for before you do your study?**

You have 3 options…

1. Based on effect sizes of previous studies

* However, there may not be any comparable studies/effects
* Published effect sizes often overestimate the ‘true’ population effect (the true effect is probably smaller due to publishing bias for significant effects). Therefore, if you use a published estimate your study may end up underpowered.

1. Use Cohen’s (1988) guidelines

* Cohen’s **f** - Small effect= 0.1; Medium effect= 0.3; Large effect= 0.5 (but lower values may be needed; see (Hemphill, 2003))
* The type of effect size changes based on the statistical test you use (e.g., t-tests use Cohen’s d; ANOVAs use eta-squared)

To prevent your study from being underpowered, consider whether the ‘true’ effect might be smaller than the cut-offs from Cohen’s guidelines. Basically, if you look for a small effect size, you will also catch any possible medium and large effect sizes because your study will not be underpowered (a safety net). However, the number of participants needed increases as the effect size gets smaller.

1. Decide on your smallest effect size of interest

* Decide what is the smallest effect size you would consider meaningful.

**Appendix 2**

**ANOVAs:**

**Effect sizes used =** (η2) not (f) like shown in G\*power.

Effect sizes for eta-squared: Small (η2 = 0.01), medium (η2 = 0.06), and large (η2 = 0.14) (Lakens, 2013).

**T-tests:**

**Effect sizes**

* For independent t-tests (between-subjects) you will see that the effect size is shown with the letter ‘d’ (you may also see it as ds).
* For dependent t-tests (or paired samples; within-subjects) you will see it as ‘dz’.

For quantifying Cohen’s d effect sizes, the same rules apply for both:

Effect sizes for Cohen’s d**:** Small (d = 0.2), medium (d = 0.5), and large (d = 0.8+) (Lakens, 2013)

**Linear multiple regression:**

**Effect sizes used = (**partial r2) not (f2)like shown in G\*power. Partial r2 follows the same rules as eta-squared.

Effect sizes for eta-squared: Small (η2 = 0.01), medium (η2 = 0.06), and large (η2 = 0.14) (Lakens, 2013).

**Appendix 3**

**Calculating Numerator df**

The main effects have different degrees of freedom (df) than the interaction effects. Depending on the effect you want to investigate, main effect or interaction, the required sample size will differ.

Main effect: the effect of one independent variable on the dependent variable

Interaction effect: one independent variable interacts with another on a dependent variable (e.g., the effect of autistic traits and gender on empathy)

Calculating main effect:

Numerator df = (number of groups -1).

e.g., if you have 3 conditions then 3-1= 2 df

Calculating interaction effect:

Numerator df = (Number of levels in factor A -1) x (Number of levels in factor B -1)

e.g., in a 3x4 design, the numerator df is (3-1) x (4-1) = 6 df

**How do you know what your design is?**

This is based on how many factors (i.e., variables) you got and the levels within them.

2x3 design example:

We want to investigate the effect of two treatments on depression severity.

Independent variable: treatments = 2 levels (psychotherapy and antidepressants)

Dependent variable: depression = 3 levels (minor, moderate, and major depression)

**Which df value should you use in G\*power?**

After you have calculated both the main effect and interaction effect df, use the one that has the highest value.

**Appendix 4**

**What is the Sphericity Assumption?**

If your experimental design relies on matching (between-factors) rather than repeated measurements (within-factors), then you can assume sphericity, as violations are unlikely.

**The assumption of sphericity states** that the variances of the difference between all combinations of related conditions (levels) are equal. Violation of sphericity is when the variances of the differences between all combinations of related groups are not equal.

A “nonsphericity correction” value of <1.0 balances out the likelihood of a Type I error\* that goes up when sphericity is not met. Note that a “nonsphericity correction” value of <1.0 will increase the sample size requirement because it raises the critical cut-off value.

\*Type 1 error: Finding an effect where there is none (falsely rejecting the null hypothesis).

**How do you select the value for nonsphericity?**

Sphericity is assumed when it has a value of 1. If your study relies on repeated measures do **not** assume sphericity. A default is 0.8 when sphericity is not assumed but if you want to play it safe you can go for the stricter value.

This means that you are less likely to commit a type II error as you may be overestimating the power you need, but this may not be feasible if you have less resources.

The lower bound can be calculated by: 1/ (number of conditions -1).

e.g., if we had 5 conditions it would be 1/ (5-1) = 0.25 **but we would use 0.26 because the value needs to be bigger than the lower bound.**

**Appendix 5**

Hypothesis testing is run to determine whether a claim is true or not. One-tailed and two-tailed tests are hypothesis testing tools. Analyses such as ANOVA and chi-square tests do not have a one-tailed vs two-tailed option because the distributions they are based on only have one tail, but you do find these in t-tests.

**T-tests:**

Before you run a one- or two-tailed test, you need to establish a null hypothesis and an alternative hypothesis.

**Two-tailed test:**

* This is appropriate when you want to determine if there are any differences between the groups you are comparing
* It looks for positive **and** negative differences between the groups.
* However, it does not tell you in which direction the effect occurred.

e.g., if you compare scores from group A and B using a two-tailed test, it will tell you if there is a significant difference between them (in other words, it will tell you if your intervention had a positive or negative effect). The result could include one of the groups scoring higher or lower than the other, but you will not know which one. You will just know that there is a difference.

**One-tailed test:**

* This is appropriate when you want to determine if there are any differences between groups that occur in a specific direction
* This one looks for positive **or** negative differences **(not both)**

e.g., if you are interested to see whether group A scored higher than group B, and you are completely uninterested to see if group A scored lower than group B, then go for a one-tailed test.

**Which one to choose:**

When in doubt, it is almost always more appropriate to use a two-tailed test. A one-tailed test is only justified if you have a specific prediction about the direction of the difference (e.g., Group A scoring higher than Group B), and you are completely uninterested in the possibility that the opposite outcome could be true (e.g., Group A scoring lower than Group B).

However, you could use a one-tailed test in a follow up statistical analysis. So, if you found that there is a difference between groups using a two-tailed test, you can then perform a one-tailed test to see the direction of that difference. *– you should only do this when the results from the two-tailed test are significant!*