

**Poweranalyse für experimentelle Designs in R**  
Workshop auf der Nachwuchstagung der GEBF 2024

Sophie E. Stallasch  
(stallasch@uni-potsdam.de)

Literaturempfehlungen

**Einführung Poweranalyse**

- Aberson, C. L. (2019). Applied power analysis for the behavioral sciences (2nd Edition). Routledge, Taylor & Francis Group.
- Ahn, C., Heo, M., & Zhang, S. (2020). Sample size calculations for clustered and longitudinal outcomes in clinical research (First issued in paperback). CRC Press.
- Bausell, R. B., & Li, Y.-F. (2002). Power analysis for experimental research: A practical guide for the biological, medical and social sciences. Cambridge University Press.  
<https://doi.org/10.1017/CBO9780511541933>
- Hedberg, E. C. (2018). Introduction to Power Analysis: Two-Group Studies. SAGE Publications, Inc. <https://doi.org/10.4135/9781506343105>
- Hedges, L. V., & Rhoads, C. (2010a). Statistical power analysis in education research. National Center for Special Education Research.  
<https://ies.ed.gov/ncser/pubs/20103006/pdf/20103006.pdf>
- Hedges, L. V., & Rhoads, C. H. (2010b). Statistical power analysis. In International encyclopedia of education. Elsevier.  
<http://www.sciencedirect.com/science/referenceworks/9780080448947>
- Lipsey, M. W. (1990). Design sensitivity: Statistical power for experimental research. Sage.
- Lipsey, M. W., & Hurley, S. (2009). Design sensitivity: Statistical power for applied experimental research. In L. Bickman & D. Rog, The SAGE Handbook of Applied Social Research Methods (pp. 44–76). SAGE Publications, Inc.  
<https://doi.org/10.4135/9781483348858.n2>
- Liu, X. S. (2014). Statistical power analysis for the social and behavioral sciences: Basic and advanced techniques. Taylor&Francis. <http://site.ebrary.com/id/10801501>
- Murphy, M. M., Mazzocco, M. M. M., Hanich, L. B., & Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities*, 40(5), 458–478. psych. <https://doi.org/10.1177/00222194070400050901>
- Perugini, M., Gallucci, M., & Costantini, G. (2018). A practical primer to power analysis for simple experimental designs. *International Review of Social Psychology*, 31(1), 20.  
<https://doi.org/10.5334/irsp.181>

### **Designparameter und Benchmarks für Effektgrößen (Outcome: Schulleistung)**

- Brunner, M., Stallasch, S. E., & Lüdtke, O. (2023). Empirical benchmarks to interpret intervention effects on student achievement in elementary and secondary school: Meta-analytic results from Germany. *Journal of Research on Educational Effectiveness*, 1–39. <https://doi.org/10.1080/19345747.2023.2175753>
- Stallasch, S. E. (2023). Multides: R Tools for the “MULTI-DES” Project. R package version 0.0.0.9000. <https://github.com/sophiestallasch/multides>
- Stallasch, S. E., Lüdtke, O., Artelt, C., & Brunner, M. (2021). Multilevel design parameters to plan cluster-randomized intervention studies on student achievement in elementary and secondary school. *Journal of Research on Educational Effectiveness*, 14(1), 172–206. <https://doi.org/10.1080/19345747.2020.1823539>
- Stallasch, S. E., Lüdtke, O., Artelt, C., Hedges, L. V., & Brunner, M. (2023). Single- and multilevel perspectives on covariate selection in randomized intervention studies on student achievement. *PsyArXiv*. <https://doi.org/10.31234/osf.io/5ajmg>

### **Geschichte der Poweranalyse und Fisher vs. Neyman-Pearson**

- Pernet, C. (2016). Null hypothesis significance testing: A short tutorial. *F1000Research*, 4, 621. <https://doi.org/10.12688/f1000research.6963.3>
- Descôteaux, J. (2007). Statistical power: An historical introduction. *Tutorials in Quantitative Methods for Psychology*, 3(2), 28–34. <https://doi.org/10.20982/tqmp.03.2.p028>
- Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, 105(2), 309–316. <https://doi.org/10.1037/0033-2909.105.2.309>

### **Originalliteratur**

- Neyman, J., & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. *Biometrika*, 20A(1/2), 175. <https://doi.org/10.2307/2331945>
- Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A*, 231(694–706), 289–337. <https://doi.org/10.1098/rsta.1933.0009>
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *The Journal of Abnormal and Social Psychology*, 65(3), 145–153. <https://doi.org/10.1037/h0045186>
- Cohen, J. (1973). Brief notes: Statistical power analysis and research results. *American Educational Research Journal*, 10(3), 225–229. <https://doi.org/10.3102/00028312010003225>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). L. Erlbaum Associates. <https://www.utstat.toronto.edu/~brunner/oldclass/378f16/readings/CohenPower.pdf>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>

- Cohen, J. (1994). The earth is round ( $p < .05$ ). *American Psychologist*, 49, 997–1003.  
<https://doi.org/10/bm96wk>
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (Eds.). (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). L. Erlbaum Associates.

### **Bayesianische und simulationsbasierte Poweranalysen**

- Pek, J., & Park, J. (2019). Complexities in power analysis: Quantifying uncertainties with a Bayesian-classical hybrid approach. *Psychological Methods*, 24(5), 590–605.  
<https://doi.org/10.1037/met0000208>
- Stallach, S. E., Lüdtke, O., Artelt, C., Hedges, L. V., & Brunner, M. (2023). Single- and multilevel perspectives on covariate selection in randomized intervention studies on student achievement. *PsyArXiv*. <https://doi.org/10.31234/osf.io/5ajmg>
- Turner, R. M., Thompson, S. G., & Spiegelhalter, D. J. (2005). Prior distributions for the intracluster correlation coefficient, based on multiple previous estimates, and their application in cluster randomized trials. *Clinical Trials*, 2(2), 108–118.  
<https://doi.org/10.1191/1740774505cn072oa>
- Turner, R. M., Toby Prevoost, A., & Thompson, S. G. (2004). Allowing for imprecision of the intracluster correlation coefficient in the design of cluster randomized trials. *Statistics in Medicine*, 23(8), 1195–1214. <https://doi.org/10.1002/sim.1721>
- Williamson, S. F., Tishkovskaya, S. V., & Wilson, K. J. (2023). Hybrid sample size calculations for cluster randomised trials using assurance (arXiv:2308.11278). *arXiv*.  
<http://arxiv.org/abs/2308.11278>

### **Debatte um Alpha-Niveaus**

- Amrhein, V., & Greenland, S. (2017). Remove, rather than redefine, statistical significance. *Nature Human Behaviour*, 2, 4. <https://doi.org/10.1038/s41562-017-0224-0>
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Johnson, V. E. (2017). Redefine statistical significance. *Nature Human Behaviour*, 2, 6–10. <https://doi.org/10.1038/s41562-017-0189-z>
- Lakens, D., Adolphi, F. G., Albers, C. J., Anvari, F., Apps, M. A. J., Argamon, S. E., Zwaan, R. A. (2018). Justify your alpha. *Nature Human Behaviour*, 2(3), 168–171.  
<https://doi.org/10.1038/s41562-018-0311-x>