

Bibliographie

- Adams, B. G., Van de Vijver, F. J. R., de Bruin, G. P., & Bueno Torres, C. (2014). Identity in descriptions of others across ethnic groups in South Africa. *Journal of Cross-Cultural Psychology*, 45(9), 1411-1433. <https://doi.org/10.1177/0022022114542466>
- Aiken, L. S., West, S. G., & Millsap, R. E. (2008). Doctoral training in statistics, measurement, and methodology in psychology: Replication and extension of Aiken, West, Sechrest, and Reno's (1990) survey of PhD programs in North America. *American Psychologist*, 63(1), 32-50. <https://doi.org/10.1037/0003-066X.63.1.32>
- Algina, J., Keselman, H. J., & Penfield, R. D. (2006). Confidence intervals for an effect size when variances are not equal. *Journal of Modern Applied Statistical Methods*, 5(1), 2-13. <https://doi.org/10.22237/jmasm/1146456060>
- Altman, D. G. (2005). Why we need confidence intervals. *World Journal of Surgery*, 29(5), 554-556. <https://doi.org/10.1007/s00268-005-7911-0>
- American Psychological Association. (2010). *Publication manual of the American Psychological Association [APA]* (6^e éd.). Washington, DC: American Psychological Association.
- Andersen, M. B., McCullagh, P., & Wilson, G. J. (2007). But what do the numbers really tell us?: Arbitrary metrics and effect size reporting in sport psychology research. *Journal of Sport and Exercise Psychology*, 29(5), 664-672. <https://doi.org/10.1123/jsep.29.5.664>
- Anderson, S. F., & Maxwell, S. E. (2016). There's more than one way to conduct a replication study: Beyond statistical significance. *Psychological Methods*, 21(1), 1-12. <https://doi.org/10.1037/met0000051>
- Balluerka, N., Gómez, J., & Hidalgo, D. (2005). The controversy over null hypothesis significance testing revisited. *Methodology*, 1(2), 55-70. <https://doi.org/10.1027/1614-1881.1.2.55>
- Banerjee, A., Chitnis, U. B., Jadhav, S. L., Bhawalkar, J. S., & Chaudhury, S. (2009). Hypothesis testing, type I and type II errors. *Industrial psychiatry journal*, 18(2), 127-131. <https://doi.org/10.4103/0972-6748.62274>

- Baumgartner, H., & Steenkamp, J.-B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of marketing research*, 38(2), 143-156. <https://doi.org/10.1509/jmkr.38.2.143.18840>
- Beilmann, M., Mayer, B., Kasearu, K., & Realo, A. (2014). The relationship between adolescents' social capital and individualism-collectivism in Estonia, Germany, and Russia. *Child Indicators Research*, 7(3), 589-611. <https://doi.org/10.1007/s12187-014-9232-z>
- Blume, J. D., D'Agostino McGowan, L., Dupont, W. D., & Greevy, R. A. (2018). Second-generation p -values: Improved rigor, reproducibility, & transparency in statistical analyses. *PLoS One*, 13(3), 1-17. <https://doi.org/10.1371/journal.pone.0188299>
- Boneau, C. A. (1960). The effects of violations of assumptions underlying the t test. *Psychological Bulletin*, 57(1), 49-64.
- Boone, H. N., & Boone, D. A. (2012). Analyzing likert data. *Journal of Extension*, 50(2), 1-5. <https://doi.org/10.1214/aoms/1177728717>
- Box, G. E. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and of correlation between errors in the two-way classification. *The Annals of Mathematical Statistics*, 25(3), 484-498. <https://doi.org/10.1214/aoms/1177728717>
- Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31(2), 144-152. <https://doi.org/10.1111/j.2044-8317.1978.tb00581.x>
- Brown, M. B., & Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69(346), 364-367. <https://doi.org/10.2307/2285659>
- Bryk, A. S., & Raudenbush, S. W. (1988). Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. *Psychological Bulletin*, 104(3), 396-404. <https://doi.org/10.1037/0033-2909.104.3.396>
- Burriss, R. P., Troscianko, J., Lovell, P. G., Fulford, A. J., Stevens, M., Quigley, R., . . . Rowland, H. M. (2015). Changes in women's facial skin color over the ovulatory cycle are not detectable by the human visual system. *PLoS One*, 10(7), 1-16. <https://doi.org/10.1371/journal.pone.0130093>

- Button, K. S., Kounali, D., Thomas, L., Wiles, N. J., Peters, T. J., Welton, N. J., ... Lewis, G. (2015). Minimal clinically important difference on the Beck Depression Inventory-II according to the patient's perspective. *Psychological Medicine*, 45(15), 3269-3279. <https://doi.org/10.1017/S0033291715001270>
- Byrne, B. M. (1996). The status and role of quantitative methods in psychology: Past, present, and future perspectives. *Canadian Psychology/Psychologie canadienne*, 37(2), 76-80. <https://doi.org/10.1037/0708-5591.37.2.76>
- Cain, M. K., Zhang, Z., & Yuan, K.-H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods*, 49(5), 1716-1735. <https://doi.org/10.3758/s13428-016-0814-1>
- Carroll, R. J., & Schneider, H. (1985). A note on Levene's tests for equality of variances. *Statistics & probability letters*, 3(4), 191-194. [https://doi.org/10.1016/0167-7152\(85\)90016-1](https://doi.org/10.1016/0167-7152(85)90016-1)
- Church, A. T., Willmore, S. L., Anderson, A. T., Ochiai, M., Porter, N., Mateo, N. J., ... Ortiz, F. A. (2012). Cultural differences in implicit theories and self-perceptions of traitedness: Replication and extension with alternative measurement formats and cultural dimensions. *Journal of Cross-Cultural Psychology*, 43(8), 1268-1296. <https://doi.org/10.1177/0022022111428514>
- Coe, R. (2002). It's the effect size, stupid: What effect size is and why it is important. Papier présenté à la conférence annuelle de la British Educational Research Association, Université d'Exeter, Exeter, Royaume-Uni.
- Cohen, A. B., & Hill, P. C. (2007). Religion as culture: Religious individualism and collectivism among American Catholics, Jews, and Protestants. *Journal of Personality*, 75(4), 709-742. <https://doi.org/10.1111/j.1467-6494.2007.00454.x>
- Cohen, J. (1965). Some statistical issues in psychological research. Dans B. B. Wolmann (Éd.), *Handbook of Clinical Psychology* (pp. 95-121). New York, NY: McGraw-Hill.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioural sciences* (Erlbaum). Mahwah, NJ: (s.n.).

- Counsell, A., & Harlow, L. (2017). Reporting practices and use of quantitative methods in Canadian journal articles in psychology. *Canadian Psychology/Psychologie canadienne*, 58(2), 140-147. <https://doi.org/10.1037/cap0000074>
- Cowdry, R. W., Gardner, D. L., O'Leary, K. M., Leibenluft, E., & Rubinow, D. R. (1991). Mood variability: a study of four groups. *The American journal of psychiatry*, 148(11), 1505-1511. <https://doi.org/10.1176/ajp.148.11.1505>
- Croasmun, J. T., & Ostrom, L. (2011). Using likert-type scales in the social sciences. *Journal of Adult Education*, 40(1), 19-22.
- Cumming, G. (2013a). Cohen's d needs to be readily interpretable: Comment on Shieh (2013). *Behavior Research Methods*, 45(4), 968-971. <https://doi.org/10.3758/s13428-013-0392-4>
- Cumming, G. (2013b). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. (S.l.): Routledge.
- Cumming, G., Fidler, F., Kalinowski, P., & Lai, J. (2012). The statistical recommendations of the American Psychological Association Publication Manual: Effect sizes, confidence intervals, and meta-analysis. *Australian Journal of Psychology*, 64(3), 138-146. <https://doi.org/10.1111/j.1742-9536.2011.00037.x>
- Curtis, D. A., & Harwell, M. (1998). Training doctoral students in educational statistics in the United States: A national survey. *Journal of Statistics Education*, 6(1), 1-23. <https://doi.org/10.1080/10691898.1998.11910604>
- David, F. N., & Johnson, N. L. (1951). The effect of nonnormality on the power function of the F-test in the analysis of variance. *Biometrika*, 38(1-2), 43-57. <https://doi.org/10.2307/2332316>
- DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, 2(3), 292-307. <https://doi.org/10.1037//1082-989x.2.3.292>
- Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t -test instead of Student's t -test. *International Review of Social Psychology*, 30(1), 92-101. <https://doi.org/10.5334/irsp.82>

- Delacre, M., Leys, C., Mora, Y. L., & Lakens, D. (2019). Taking parametric assumptions seriously: Arguments for the use of Welch's F -test instead of the classical F -test in one-way ANOVA. *International Review of Social Psychology*, 32(1), 1-12. <https://doi.org/10.5334/irsp.198>
- Duran, R. P., Eisenhart, M. A., Erickson, F. D., Grant, C. A., Green, J. L., Hedges, L. V., & Schneider, B. L. (2006). Standards for reporting on empirical social science research in AERA publications: American Educational Research Association. *Educational Researcher*, 35(6), 33-40.
- Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.
- Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge, Royaume-Uni: Cambridge University Press.
- Erceg-Hurn, D. M., & Miroseovich, V. M. (2008). Modern robust statistical methods: an easy way to maximize the accuracy and power of your research. *American Psychologist*, 63(7), 591-601. <https://doi.org/10.1037/0003-066X.63.7.591>
- Everitt, B. S. (2001). *Statistics for psychologists: An intermediate course*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of educational research*, 62(1), 61-84. <https://doi.org/10.3102/00346543062001061>
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4^e éd.). Washington, DC: Sage.
- Finch, S., Cumming, G., & Thomason, N. (2001). Reporting of statistical inference in the Journal of Applied Psychology: Little evidence of reform. *Educational and Psychological Measurement*, 61(2), 181-210. <https://doi.org/10.1177/00131640121971167>
- Fraas, J. W., & Newman, I. (2000). Testing for Statistical and Practical Significance: A Suggested Technique Using a Randomization Test. Papier présenté à la réunion annuelle de la Mid-Western Educational Research Association, Chicago.

- Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2), 156-168. <https://doi.org/10.1177/2515245919847202>
- Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74-78. <https://doi.org/10.1016/j.paid.2016.06.069>
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in Social Research*. Beverly Hills, CA: Sage.
- Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, 42(3), 237-288. <https://doi.org/10.3102/00346543042003237>
- Goertzen, J. R., & Cribbie, R. A. (2010). Detecting a Lack of Association: An Equivalence Testing Approach. *British Journal of Mathematical and Statistical Psychology*, 63(3), 527-537. <https://doi.org/10.1348/000711009X475853>
- Golinski, C., & Cribbie, R. A. (2009). The expanding role of quantitative methodologists in advancing psychology. *Canadian Psychology/Psychologie canadienne*, 50(2), 83-90. <https://doi.org/10.1037/a0015180>
- Goulet-Pelletier, J.-C., & Cousineau, D. (2018). A review of effect sizes and their confidence intervals, Part I: The Cohen's *d* family. *The Quantitative Methods for Psychology*, 14(4), 242-265. <https://doi.org/10.20982/tqmp.14.4.p242>
- Green, E. G. T., Deschamps, J.-C., & Páez, D. (2005). Variation of individualism and collectivism within and between 20 countries: A typological analysis. *Journal of cross-cultural psychology*, 36(3), 321-339. <https://doi.org/10.1177/0022022104273654>
- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24(2), 95-112. <https://doi.org/10.1007/BF02289823>
- Grey, S., & Mathews, A. (2000). Effects of training on interpretation of emotional ambiguity. *the quarterly journal of experimental psychology*, 53(4), 1143-1162. <https://doi.org/10.1080/713755937>

- Grissom, R. J. (2000). Heterogeneity of variance in clinical data. *Journal of Consulting and Clinical Psychology*, 68(1), 155-165. <https://doi.org/10.1037/0022-006X.68.1.155>
- Grissom, R. J., & Kim, J. J. (2001). Review of assumptions and problems in the appropriate conceptualization of effect size. *Psychological Methods*, 6(2), 135-146. <https://doi.org/10.1037/1082-989X.6.2.135>
- Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Haar, J. M., Russo, M., Suñe, A., & Ollier-Malaterre, A. (2014). Outcomes of work-life balance on job satisfaction, life satisfaction and mental health: A study across seven cultures. *Journal of Vocational Behavior*, 85(3), 361-373. <https://doi.org/10.1016/j.jvb.2014.08.010>
- Hanushek, E. A., & Wößmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal*, 116(510), C63-C76. <https://doi.org/10.1111/j.1468-0297.2006.01076.x>
- Hartley, J. (2014). Some thoughts on Likert-type scales. *International Journal of Clinical and Health psychology*, 14(1), 83-86. [https://doi.org/10.1016/S1697-2600\(14\)70040-7](https://doi.org/10.1016/S1697-2600(14)70040-7)
- Harwell, M. R. (1992). Summarizing Monte Carlo results in methodological research. *Journal of Educational Statistics*, 17(4), 297-313. <https://doi.org/10.3102/10769986017004297>
- Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo Results in Methodological Research: The One- and Two-Factor Fixed Effects ANOVA Cases. *Journal of Educational Statistics*, 17(4), 315-339.
- Haslam, S. A., & McGarty, C. (2014). *Research methods and statistics in psychology* (2^e éd.). Londres, Royaume-Uni: Sage.
- Hauck, D. W. W., & Anderson, S. (1984). A New Statistical Procedure for Testing Equivalence in Two-Group Comparative Bioavailability Trials. *Journal of Pharmacokinetics and Biopharmaceutics*, 12(1), 83-91. <https://doi.org/10.1007/BF01063612>
- Hayes, A. F., & Cai, L. (2007). Further evaluating the conditional decision rule for comparing two independent means. *British Journal of Mathematical and Statistical Psychology*, 60(2), 217-244. <https://doi.org/10.1348/000711005X62576>

- Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466(7302), 29-29. <https://doi.org/10.1038/466029a>
- Heun, R., Burkart, M., Maier, W., & Bech, P. (1999). Internal and external validity of the WHO Well-Being Scale in the elderly general population. *Acta Psychiatrica Scandinavica*, 99(3), 171-178. <https://doi.org/10.1111/j.1600-0447.1999.tb00973.x>
- Hoekstra, R., Kiers, H., & Johnson, A. (2012). Are assumptions of well-known statistical techniques checked, and why (not)? *Frontiers in Psychology*, 3(137), 1-9. <https://doi.org/10.3389/fpsyg.2012.00137>
- Howitt, D., & Cramer, D. (2017). *Understanding statistics in psychology with SPSS* (7^e éd.). Édimbourg: Pearson Education.
- Hsu, T.-C., & Feldt, L. S. (1969). The Effect of Limitations on the Number of Criterion Score Values on the Significance Level of the F-Test. *American Educational Research Journal*, 6(4), 515-527. <https://doi.org/10.3102/00028312006004515>
- Huynh, C.-L. (1989). A Unified Approach to the Estimation of Effect Size in Meta-Analysis. Papier présenté à la réunion annuelle de l'American Educational Research Association, San Francisco, CA.
- Huynh, C.-L., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics*, 1(1), 69-82. <https://doi.org/10.3102/10769986001001069>
- Jamieson, S. (2004). Likert scales: How to (ab)use them? *Medical Education*, 38(12), 1217-1218. <https://doi.org/10.1111/j.1365-2929.2004.02012.x>
- Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *British Journal of Applied Science & Technology*, 7(4), 396-403. <https://doi.org/10.9734/BJAST/2015/14975>
- Judd, C. M., McClelland, G. H., Ryan, C. S., Muller, D., & Yzerbyt, V. (2010). *Analyse des données: une approche par comparaison de modèles* (7^e éd.). Bruxelles: De Boeck.

- Kelley, K. (2005). The effects of nonnormal distributions on confidence intervals around the standardized mean difference: Bootstrap and parametric confidence intervals. *Educational and Psychological Measurement*, 65(1), 51-69. <https://doi.org/10.1177/0013164404264850>
- Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3^e éd.). New Jersey: Prentice-Hall, Inc.
- Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (Prentice Hall). Upper Saddle River , New Jersey: (s.n.).
- Keselman, H. J., Algina, J., & Kowalchuk, R. K. (2001). The analysis of repeated measures designs: A review. *British Journal of Mathematical and Statistical Psychology*, 54(1), 1-20. <https://doi.org/10.1348/000711001159357>
- Keselman, H. J., Algina, J., Lix, L. M., Wilcox, R. R., & Deering, K. N. (2008). A generally robust approach for testing hypotheses and setting confidence intervals for effect sizes. *Psychological Methods*, 13(2), 110-129. <https://doi.org/10.1037/1082-989X.13.2.110>
- Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., ... Keselman, J. C. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research*, 68(3), 350-386. <https://doi.org/10.3102/00346543068003350>
- Keselman, H. J., Othman, A. R., Wilcox, R. R., & Fradette, K. (2004). The new and improved two-sample t test. *Psychological Science*, 15(1), 47-51. <https://doi.org/10.1111/j.0963-7214.2004.01501008.x>
- Keselman, H. J., & Rogan, J. C. (1980). Repeated measures *F* tests and psychophysiological research: Controlling the number of false positives. *Psychophysiology*, 17(5), 499-503. <https://doi.org/10.1111/j.1469-8986.1980.tb00190.x>
- Kester, S. W. (1969). *The communication of teacher expectations and their effects on the achievement and attitudes of secondary school pupils*. (S.l.): The University of Oklahoma.
- Koeser, S., & Sczesny, S. (2014). Promoting gender-fair language: The impact of arguments on language use, attitudes, and cognitions. *Journal of Language and Social Psychology*, 33(5), 548-560. <https://doi.org/10.1177/0261927x14541280>

- Kruschke, J. K. (2018). Rejecting or Accepting Parameter Values in Bayesian Estimation. *Advances in Methods and Practices in Psychological Science*, 1(2), 270-280. <https://doi.org/10.1177/2515245918771304>
- Kulinskaya, E., & Staudte, R. G. (2007). Confidence intervals for the standardized effect arising in the comparison of two normal populations. *Statistics in medicine*, 26(14), 2853-2871. <https://doi.org/10.1002/sim.2751>
- 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), 1-12. <https://doi.org/10.3389/fpsyg.2013.00863>
- Lakens, D. (2016, 9 décembre). The 20% Statistician: TOST equivalence testing R package (TOSTER) and spreadsheet. Repéré à <http://daniellakens.blogspot.com/2016/12/tost-equivalence-testing-r-package.html>
- Lakens, D. (2017). Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355-362. <https://doi.org/10.1177/1948550617697177>
- Lakens, D., Hilgard, J., & Staaks, J. (2016). On the reproducibility of meta-analyses: Six practical recommendations. *BMC psychology*, 4(1), 1-10. <https://doi.org/10.1186/s40359-016-0126-3>
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, 1(2), 259-269. <https://doi.org/10.1177/2515245918770963>
- Lane, D. (2016). The assumption of sphericity in repeated-measures designs: What it means and what to do when it is violated. *Quantitative Methods for Psychology*, 12(2), 114-122. <https://doi.org/10.20982/tqmp.12.2.p114>
- Lim, T.-S., & Loh, W.-Y. (1996). A comparison of tests of equality of variances. *Computational Statistics & Data Analysis*, 22(3), 287-301. [https://doi.org/10.1016/0167-9473\(95\)00054-2](https://doi.org/10.1016/0167-9473(95)00054-2)
- Liu, H. (2015). *Comparing Welch ANOVA, a Kruskal-Wallis test, and traditional ANOVA in case of heterogeneity of variance* (thèse de doctorat). Virginia Commonwealth University.

- Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of Assumption Violations Revisited: A Quantitative Review of Alternatives to the One-Way Analysis of Variance *F* Test. *Review of educational research*, 66(4), 579-619. <https://doi.org/10.3102/00346543066004579>
- Luh, W.-M., & Guo, J.-H. (2007). Approximate sample size formulas for the two-sample trimmed mean test with unequal variances. *British Journal of Mathematical and Statistical Psychology*, 60(1), 137-146. <https://doi.org/10.1348/000711006X100491>
- McCall, R. B., & Appelbaum, M. I. (1973). Bias in the analysis of repeated-measures designs: Some alternative approaches. *Child Development*, 401-415. <https://doi.org/10.2307/1127993>
- Meyners, M. (2012). Equivalence tests—A review. *Food Quality and Preference*, 26(2), 231-245. <https://doi.org/10.1016/j.foodqual.2012.05.003>
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105(1), 156-166. <https://doi.org/10.1037/0033-2909.105.1.156>
- Mills, L., Abdulla, E., & Cribbie, R. (2010). Quantitative methodology research: Is it on psychologists' reading lists? <https://doi.org/10.20982/tqmp.06.2.p052>
- Montoya, D. Y., & Briggs, E. (2013). Shared ethnicity effects on service encounters: A study across three US subcultures. *Journal of Business Research*, 66(3), 314-320. <https://doi.org/10.1016/j.jbusres.2011.08.011>
- Moser, B. K., & Stevens, G. R. (1992). Homogeneity of variance in the two-sample means test. *The American Statistician*, 46(1), 19-21.
- Murphy, K. R., Myers, B., & Wolach, A. H. (2014). *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests* (Fourth edition). New York: Routledge, Taylor & Francis Group. (01274).
- Nachar, N. (2008). The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution. *Tutorials in quantitative Methods for Psychology*, 4(1), 13-20.

- Neuhäuser, M., & Ruxton, G. D. (2009). Distribution-free two-sample comparisons in the case of heterogeneous variances. *Behavioral Ecology and Sociobiology*, 63(4), 617-623. <https://doi.org/10.1007/s00265-008-0683-4>
- Newman, I., Fraas, J. W., & Herbert, A. (2001). Testing Non-Null Hypotheses with *t* Tests of Group Means: A Monte Carlo Study. Papier présenté à la réunion annuelle de la Mid-Western Educational Research Association, Chicago, IL.
- Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological Methods*, 5(2), 241-301. <https://doi.org/10.1037/1082-989x.5.2.241>
- Nimon, K. F. (2012). Statistical assumptions of substantive analyses across the general linear model: A mini-review. *Frontiers in psychology*, 3(322), 1-5. <https://doi.org/10.3389/fpsyg.2012.00322>
- Nordstokke, D. W., & Zumbo, B. D. (2007). A Cautionary Tale about Levene's Tests for Equal Variances. *Journal of Educational Research & Policy Studies*, 7(1), 1-14.
- Nunnally, J. (1960). The place of statistics in psychology. *Educational and Psychological Measurement*, 20(4), 641-650. <https://doi.org/10.1177/001316446002000401>
- O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: an extensive primer. *Psychological Bulletin*, 97(2), 316-333. <https://doi.org/10.1037/0033-2909.97.2.316>
- Osborne, J. W., & Christianson, W. R. (2001). Educational Psychology from a Statistician's Perspective: A Review of the Quantitative Quality of Our Field. Papier présenté à la réunion annuelle de l'American Educational Research Association, Seattle, WA.
- Overall, J. E., Atlas, R. S., & Gibson, J. M. (1995). Tests that are robust against variance heterogeneity in k x 2 designs with unequal cell frequencies. *Psychological Reports*, 76(3), 1011-1017. <https://doi.org/10.2466/pr0.1995.76.3.1011>
- Pek, J., & Flora, D. B. (2018). Reporting effect sizes in original psychological research: A discussion and tutorial. *Psychological Methods*, 23(2), 208-225. <https://doi.org/10.1037/met0000126>

- Peng, C.-Y. J., & Chen, L.-T. (2014). Beyond Cohen's d : Alternative effect size measures for between-subject designs. *The Journal of Experimental Education*, 82(1), 22-50. <https://doi.org/10.1080/00220973.2012.745471>
- Peng, C.-Y. J., Chen, L.-T., Chiang, H.-M., & Chiang, Y.-C. (2013). The impact of APA and AERA guidelines on effect size reporting. *Educational Psychology Review*, 25(2), 157-209. <https://doi.org/10.1007/s10648-013-9218-2>
- Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin*, 112(1), 160-164. <https://doi.org/10.1037/0033-2909.112.1.160>
- Quensel, C.-E. (1947). The Validity of the χ^2 -Criterion when the Variates are taken from Different Normal Populations. *Scandinavian Actuarial Journal*, 30(1), 44-55. <https://doi.org/10.1080/03461238.1947.10419648>
- Quertemont, E. (2011). How to statistically show the absence of an effect. *Psychologica Belgica*, 51(2), 109-127. <https://doi.org/10.5334/pb-51-2-109>
- Quintana, S. M., & Maxwell, S. E. (1994). A Monte Carlo comparison of seven ε -adjustment procedures in repeated measures designs with small sample sizes. *Journal of Educational Statistics*, 19(1), 57-71. <https://doi.org/10.3102/10769986019001057>
- Rakotomalala, R. (2008). Comparaison de populations. Repéré à https://eric.univ-lyon2.fr/~ricco/cours/cours/Comp_Pop_Tests_Nonparametriques.pdf
- Rasch, D., Kubinger, K. D., & Moder, K. (2011). The two-sample t test: pre-testing its assumptions does not pay off. *Statistical papers*, 52(1), 219-231. <https://doi.org/10.1007/s00362-009-0224-x>
- Raviv, E. (2014, 2 juin). Bias vs. Consistency. Repéré à <https://eranraviv.com/bias-vs-consistency/>
- Rogers, J. L., Howard, K. I., & Vessey, J. T. (1993). Using significance tests to evaluate equivalence between two experimental groups. *Psychological Bulletin*, 113(3), 553-565. <https://doi.org/10.1037/0033-2909.113.3.553>

- Ruxton, G. D. (2006). The unequal variance t -test is an underused alternative to Student's t -test and the Mann–Whitney U test. *Behavioral Ecology*, 17(4), 688-690. <https://doi.org/10.1093/beheco/ark016>
- Schneider, P. J., & Penfield, D. A. (1997). Alexander and Govern's approximations: Providing an alternative to ANOVA under variance heterogeneity. *The Journal of Experimental Education*, 65(3), 271-286. <https://doi.org/10.1080/00220973.1997.9943459>
- Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, 15(6), 657-680.
- Seaman, M. A., & Serlin, R. C. (1998). Equivalence confidence intervals for two-group comparisons of means. *Psychological Methods*, 3(4), 403-411. <https://doi.org/10.1037/1082-989X.3.4.403>
- Serlin, R. C., & Lapsley, D. K. (1985). Rationality in Psychological Research: The Good-Enough Principle. *American Psychologist*, 40(1), 73-83. <https://doi.org/http://dx.doi.org/10.1037/0003-066X.40.1.73>
- Shaw, R. G., & Mitchell-Olds, T. (1993). ANOVA for unbalanced data: an overview. *Ecology*, 74(6), 1638-1645. <https://doi.org/10.2307/1939922>
- Sheskin, D. J. (2003). *Handbook of parametric and nonparametric statistical procedures* (3^e éd.). Boca Raton, Florida: CRC Press.
- Shieh, G. (2013). Confidence intervals and sample size calculations for the standardized mean difference effect size between two normal populations under heteroscedasticity. *Behavior Research Methods*, 45(4), 955-967. <https://doi.org/10.3758/s13428-013-0320-7>
- Shields, S. (1975). Functionalism, Darwinism, and the psychology of women. *American Psychologist*, 30(7), 739-754. <https://doi.org/10.1037/h0076948>
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P -curve: a key to the file-drawer. *Journal of Experimental Psychology: General*, 143(2), 534-547. <https://doi.org/10.1037/a0033242>

- Spiegelhalter, D. J., Freedman, L. S., & Parmar, M. K. (1994). Bayesian Approaches to Randomized Trials. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 157(3), 357-416. <https://doi.org/10.2307/2983527>
- Srivastava, A. B. L. (1959). Effects of non-normality on the power of the analysis of variance test. *Biometrika*, 46(1-2), 114-122. <https://doi.org/10.2307/2332813>
- Stout, D. E., & Ruble, T. L. (1995). Assessing the practical significance of empirical results in accounting education research: The use of effect size information. *Journal of Accounting Education*, 13(3), 281-298. [https://doi.org/10.1016/0748-5751\(95\)00010-J](https://doi.org/10.1016/0748-5751(95)00010-J)
- Student. (1908). The probable error of a mean. *Biometrika*, 6(1), 1-25.
- Subedi, B. P. (2016). Using Likert type data in social science research: Confusion, issues and challenges. *International journal of contemporary applied sciences*, 3(2), 36-49.
- Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the *P* value is not enough. *Journal of Graduate Medical Education*, 4(3), 279-282. <https://doi.org/10.4300/JGME-D-12-00156.1>
- Thompson, B., & Snyder, P. A. (1997). Statistical significance testing practices in the Journal of Experimental Education. *The Journal of Experimental Education*, 66(1), 75-83. <https://doi.org/10.1080/00220979709601396>
- Tiku, M. L. (1971). Power function of the *F*-test under non-normal situations. *Journal of the American Statistical Association*, 66(336), 913-916. <https://doi.org/10.1080/01621459.1971.10482371>
- Tomarken, A. J., & Serlin, R. C. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin*, 99(1), 90-99. <https://doi.org/10.1037//0033-2909.99.1.90>
- Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Sciences*, 21(1), 19-25.
- Vasey, M. W., & Thayer, J. F. (1987). The continuing problem of false positives in repeated measures ANOVA in psychophysiology: A multivariate solution. *Psychophysiology*, 24(4), 479-486. <https://doi.org/10.1111/j.1469-8986.1987.tb00324.x>

- Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2008). *Mathematical Statistics with Applications* (7^e éd.). Belmont, USA: Brooks/Cole.
- Wallenstein, S., Zucker, C. L., & Fleiss, J. L. (1980). Some statistical methods useful in circulation research. *Circulation Research*, 47(1), 1-9. <https://doi.org/10.1161/01.RES.47.1.1>
- Wang, H., Smith, K. P., Combs, E., Blake, T., Horsley, R. D., & Muehlbauer, G. J. (2012). Effect of population size and unbalanced data sets on QTL detection using genome-wide association mapping in barley breeding germplasm. *Theoretical and Applied Genetics*, 124(1), 111-124.
- Wasserman, B. D., & Weseley, A. J. (2009). 'Qué? Quoi? Do languages with grammatical gender promote sexist attitudes? *Sex Roles*, 61, 634-643. <https://doi.org/10.1007/s11199-009-9696-3>
- Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29(3), 350-362. <https://doi.org/10.2307/2332010>
- Wellek, S. (2010). *Testing Statistical Hypotheses of Equivalence and Noninferiority* (6^e éd.). Boca Raton: CRC Press.
- Westfall, P. H. (2014). Kurtosis as Peakedness, 1905-2014. R.I.P. *The American Statistician*, 68(3), 191-195. <https://doi.org/10.1080/00031305.2014.917055>
- Westlake, W. J. (1972). Use of Confidence Intervals in Analysis of Comparative Bioavailability Trials. *Journal of Pharmaceutical Sciences*, 61(8), 1340-1341. <https://doi.org/10.1002/JPS.2600610845>
- Wilcox, R. R. (1994). Some results on the Tukey-McLaughlin and Yuen methods for trimmed means when distributions are skewed. *Biometrical Journal*, 36(3), 259-273. <https://doi.org/10.1002/bimj.4710360302>
- Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist*, 53(3), 300-314. <https://doi.org/10.1037/0003-066X.53.3.300>
- Wilcox, R. R. (2005). Comparing Medians: An Overview Plus New Results on Dealing With Heavy-Tailed Distributions. *The Journal of Experimental Education*, 73(3), 249-263. <https://doi.org/10.3200/JEXE.73.3.249-263>

- Wilcox, R. R. (2011). *Introduction to robust estimation and hypothesis testing* (2^e éd.). Cambridge, Massachusetts, USA: Academic press. <https://doi.org/10.1016/B978-0-12-386983-8.00010-X>
- Wilcox, R. R. (2017). *Modern statistics for the social and behavioral sciences: A practical introduction*. New York: Chapman & Hall.
- Wilcox, R. R., Granger, D. A., & Clark, F. (2013). Modern robust statistical methods: Basics with illustrations using psychobiological data. *Universal Journal of Psychology*, 1(2), 21-31.
- Wilcox, R. R., & Keselman, H. J. (2003). Modern robust data analysis methods: measures of central tendency. *Psychological methods*, 8(3), 254-274. <https://doi.org/10.1037/1082-989X.8.3.254>
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594-604. <https://doi.org/10.1037/0003-066X.54.8.594>
- Yuan, K.-H., Bentler, P. M., & Chan, W. (2004). Structural equation modeling with heavy tailed distributions. *Psychometrika*, 69(3), 421-436.
- Yuen, K. K. (1974). The two-sample trimmed t for unequal population variances. *Biometrika*, 61(1), 165-170. <https://doi.org/10.1093/biomet/61.1.165>
- Zimmerman, D. W. (1996). Some properties of preliminary tests of equality of variances in the two-sample location problem. *The Journal of General Psychology*, 123(3), 217-231. <https://doi.org/10.1080/00221309.1996.9921274>
- Zimmerman, D. W. (2000). Statistical significance levels of nonparametric tests biased by heterogeneous variances of treatment groups. *The Journal of General Psychology*, 127(4), 354-364. <https://doi.org/10.1080/00221300009598589>
- Zimmerman, D. W. (2004). A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology*, 57(1), 173-181. <https://doi.org/10.1348/000711004849222>

Zumbo, B. D., & Coulombe, D. (1997). Investigation of the robust rank-order test for non-normal populations with unequal variances: The case of reaction time. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 51(2), 139-150.
<https://doi.org/10.1037/1196-1961.51.2.139>