|  |
| --- |
| List of effect sizes included in the application |
| Independent Groups |
| Parametric methods |
| Assuming normality and homoscedasticity |
| Cohen’s *d*1,2,3,4,5,7,8,10,11,12,15  With being  Mean difference standardized by the pooled standard deviation of the two groups. serves as an estimate of a true common standard deviation of the groups.  This estimator has a bias – it systematically overestimates the size of the true population standardized mean difference6. |
| Hedges’ g1,2,3,4,6,8,10,11,15  With *df* being  Mean difference standardized by the pooled standard deviation of the two groups – corrected for bias. |
| Mahalanobis’ *D*1:  With and being the vectors of means for groups a and b,  being the pooled covariance matrix,  *R* being the pooled correlation matrix, and  ***d*** being the vector of *d* values  Difference between the centroids of groups a and b standardized by the standard deviation of their discriminant axis.  This estimator has a bias – it systematically overestimates the size of the true population value of *D*. |
| Adjusted Mahalanobis’ *D*1:  With *k* being the number of variables  Difference between the centroids of groups a and b standardized by the standard deviation of their discriminant axis – corrected for bias. |
| Common Language Effect Size1:  Probability that a randomly picked individual from the group with the higher mean will outscore a randomly picked individual from the other group  parametric:    Φ(∙): normal cumulative distribution function (CDF)  confidence interval calculated with the same formula just with lower and upper ci of |d| |
| Overlapping coefficient: Proportion of overlap relative to a single distribution1:    Φ(∙): normal cumulative distribution function (CDF)  Quantify the proportion of the distribution area therefore ranges between 0 and 1. |
| OVL2: Proportion of overlap relative to the joint distributions1: |
| U1: Proportion of nonoverlap relative to the joint distribution, i.e. percentage of scores that do not overlap in the two distributions1: |
| U3: Proportion of individuals in the group with the higher mean who exceed the median individual of the other group1: |
| Probability of correct classification (PCC)1:  The probability that a that a randomly picked individual will be correctly classified    This method is only a good approximation if assumptions of normality and equality of variances/covariances are satisfied  The probability of success greatly depends on the statistical model used to perform the classification task. |
| Tail ratios1,22:  With *d* being the value of Cohen’s *d*,  And *z* being the user selected number of (pooled) standard deviations  Proportion of members of group a relative to members of group b (since *d* is defined as ) that can be found z pooled standard deviations away from the mean of group b.  The tail ratio can also be reported on a logarithmic scale – this is recommended to ensure comparibility22.  A multivariate tail ratio can be defined with the help of the Mahalanobis’ distance *D* (which also assumes normality of homogeneity of variances and covariances):  Proportion of members of group a relative to members of group b that can be found z standard deviations (on the discriminant axis) from the centroid of group b. |
| Assuming normality but not homoscedasticity |
| Glass’ 2,3,4,8,10,11,13,15  Mean difference standardized by the standard deviation of either group.  This estimator has a bias – it systematically overestimates the size of the true population standardized mean difference6. |
| Adjusted Glass’ 2,3,7,8,15  With df being  Mean difference standardized by the standard deviation of either group – corrected for bias |
| Cohen’s d’2,3,7,10,11  With s’ being  Mean difference standardized by the square root of the unweighted mean of the two group’s sample variances - an estimate of the standard deviation of a hypothetical population whose standard deviation is between σa andσb. This formula estimates an effect size in a hypothetical population. |
| Kulinska and Staudte ES10:  Alternatively:  Mean difference standardized by a weighted sum of the group variances. Plagued by similar problems like Cohen’s d when variances and sample sizes are unequal. |
| Variance ratio (*VR*)1:  Ranging from 0 to (theoretically) ∞. Often the group with the (usually) higher variation is placed in the numerator.  Sometimes a logarithmized version is employed:  The benefit of the log-transformed VR is that it is no longer bounded. A VR of 1 corresponds to a log VR of 0. A VR > 1 corresponds to log VR > 0 and a VR < 1 corresponds to a log VR < 0. Theoretically possible values range from -∞ to ∞  Confidence interval for VR:  As an alternative: calculate 95 percentile bootstraps from the data – possible both for regular VR as well as log VR.  For the multivariate case, the ratio of the covariance matrices for groups a and b can be assessed:  With and being the covariance matrices of groups a and b. |
| Tail ratios1,22:  With *z* being the user selected number of standard deviations of the reference group (b).  Proportion of members of group a relative to members of group b that can be found z standard deviations away from the mean of group b.  The tail ratio can also be reported on a logarithmic scale – this is recommended to ensure comparibility22. |
| standardized difference effect sizes that are somewhat outlier resistant |
| 1.1.3.1 standardized median differences |
| Difference in medians standardized by the median absolute deviation from the median (MAD) of the baseline/control group2,3  With MAD being  The MAD has lower sampling variance and is more robust to outliers than the standard deviation. Some percentage of the scores might also be trimmed prior to the calculation of the medians and/or of the MAD2. |
| Difference in medians standardized by 2,3  is more outlier resistant than the standard deviation and at the same time approximates the standard deviation - under normality it approximates the standard deviation very well2 |
| Difference in medians standardized by the biweight standard deviation of the control/baseline group2 With being  With and being  The biweight standard deviation, also called the biweight midvariance, is cited as being the most robust to outliers out of several measures of variability.2 |
| 1.1.3.2. trimmed mean differences assuming homoscedasticity |
| The robust d statistic3,5,11,12 With and being the trimmed means of the two groups and,  With being the square root of the pooled winsorized variance  Trimmed mean difference standardized by the square root of the pooled winsorized variance. Under normality and homoscedasticity, the value of this robust statistic will differ from Cohen’s d. |
| The scaled robust d statistic3,5,11,12 A scaled version of the above statistic. Under normality and homoscedasticity, the value of this scaled robust statistic will be equal to the value of Cohen’s d. 0.642 is the value for the winsorized standard deviation for a standard normal distribution for 20% trimmed means. Thus, the rescaled winsorized variance is a consistent estimate of the population standard deviation . |
| 1.1.3.3. Trimmed mean differences not assuming homoscedasticity |
| The scaled robust statistic3,10,11,13: *Trimmed mean difference standardized by the squared root of the winsorized variance of the j-th group. Under normality, the value of this statistic will be equal to Glass’ . 0.642 is the value for the winsorized standard deviation for a standard normal distribution for 20% trimmed means. Thus, the rescaled winsorized variance is a consistent estimate of the population standard deviation* . |
| The robust d’ statistic10,11:  With being:  *Trimmed mean difference standardized by the square root of the unweighted mean of the two group’s winsorized sample variances.* |
| nonparametric methods |
| Nonparametric versions of Glass’ :  2,3,8  With being the proportion of members of one group (e.g. the control/baseline group) that lie below the “middle” of the other group (e.g. treatment group). Different definitions of the middle exist – one being the median.  With being the scores of the members of group a and being the scores of the members of group b.  An alternate nonparametric estimate of Glass *dG* can be attained if one has pre- and postscores from a 2-by-2 design – this method is advised if pre- and postscores are available8:  8  With being the proportion of individuals whose scores lie above a line through the graph origin of (,) and the point . This is the proportion of individuals in group a whose scores increased relatively more strongly from pre- to posttest than their counterparts in group b.  And with being the proportion of individuals in group b whose scores lie below the line through the graph origin of (, ) and the point . This is the proportion of individuals in group b whose scores increased relatively less strongly from pre- to posttest than their counterparts in group a.  With being the pretest scores of members of group a, being the pretest scores of members of group b, being the posttest scores of members of group a, and lastly being the posttest scores of members of group b.  When *n* is small or might be equal to 0 or 1. To avoid consequently extreme effect sizes is set to or respectively in such cases.  If the observations are independently normally distributed, these nonparametric effect sizes estimate their corresponding versions of Glass’ , i.e. and estimate , while and estimate . |
| Probability of Superiority based on Mann-Whitney2,3,19:    U is the number of wins of each element of group a against all possible elements of group b  ranging from 0 to 1.  Confidence interval based on Newcombe(2005) second method. |
| Generalized Odds ratio2,3,20:    Probability of superiority of group a compared to group b divided by the probability of superiority of group b compared to group a  Confidence Interval based on Browne et al., 2010  Question: How should ties in this context be handled? Both options(ignoring ties, counting them as 0.5 are implemented) |
| Dominance measure2,3,5:    Probability of superiority of group a compared to group b minus probability of group b compared to group a  Confidence interval based on Pratt‘s Method and the transformation formula: |
| Overlapping coefficient8:  Proportion of overlap relative to a single distribution  nonparametric estimation based on Schmid & Schmidt (2006):  1. Density estimation of the two distributions  2. Approximation of underlying function  3. Approximation of integral of minimum of both functions |
| U31:  Proportion of individuals in the group with the higher mean who exceed the median individual of the other group  nonparametric estimation:  Calculate exact value  1. Calculate means  2. Calculate amount of individuals in group of higher mean that exceed mean of other group  3. Determine proportion |
| Tail ratios1,22:   |  |  |  |  | | --- | --- | --- | --- | |  | Group a | Group b |  | | Tail region of interest |  |  |  | | Rest of the distribution |  |  |  | |  |  |  |  |   Proportion of members of group a relative to members of group b that can be found below/above a specified cut-off value, i.e. in the tail region of interest.  The tail ratio can also be reported on a logarithmic scale – this is recommended to ensure comparibility22. |
| Dependent groups |
| parametric methods |
| 2.1.1. Mean of difference scores standardized by the standard deviation of difference scores |
| Cohen’s 3,4,7,16: With **x** being the pre- and **y** being the posttreatment scores,  being the null-hypothesized mean of the difference scores (per default = 0),  and with being:  With N being the number of pairs. Mean of the difference scores standardized by the standard deviation of the difference scores. This statistic estimates the number of difference score standard deviations the mean of these difference scores lies from 0 (or more generally). This effect size has been developed for power analysis of a repeated measures t-test and many authors discourage its use as a population effect size estimate2,3,4. The reason being that the correlation between pre- and posttest scores influences the size of the standardizer and thus this correlation confounds the treatment effect. |
| Assuming normality and homoscedasticity |
| Cohen’s d (for dependent groups)2,3,410,15  With being  Since the number of pre and post measurement should be equal, can be written as:  In this case Cohen’s d will equal Cohen’s d’.  Mean difference standardized by the pooled standard deviation of the two measurements. serves as an estimate of a true common standard deviation of the measurements. Lakens4 refers to this effect size measure as :  With **x** being the pre- and **y** being the posttreatment scores.  Mean of difference scores standardized by the average of the pre- and posttreatment standard deviations. With equal sample variances and (as well as *d* in an independent groups design with the same values for groups means and standard deviations) will be identical. will be generally more similar to *d*, as long as the correlation between the two sets of measurements isn’t too high or low and sample variances are not too different. This effect size is biased – Hedges’ correction can be applied again:  With *df* being  Since the pretest and posttest measurements are pooled. |
| Adjusted Cohen’s d/d’/3:  Mean difference standardized by the pooled standard deviation of the two measurements – corrected for small sample bias. |
| Cohen’s 4,16:  With **x** being the pre- and **y** being the posttreatment scores,  *r* being the correlation between the two sets of measurements  being the null-hypothesized mean of the difference scores (per default = 0),  and with being:  With N being the number of pairs.  has an alternate formula:  Mean of difference scores standardized by the standard deviation of difference scores – controlled for the correlation between measurements. The correlation between the two sets of measurements influences the size of the standardizer of . With equal sample variances the mean difference is standardized by a value which is larger than the standardizer of an independent groups standardized mean difference by a factor of . This estimate is biased – Hedges’ correction can be applied:  With *df* being |
| Common Language Effect Size1:  Probability that a randomly picked individual from the group with the higher mean will outscore a randomly picked individual from the other group  parametric:    Φ(∙): normal cumulative distribution function (CDF)  confidence interval calculated with the same formula just with lower and upper ci of |d| |
| Overlapping coefficient: Proportion of overlap relative to a single distribution1:    Φ(∙): normal cumulative distribution function (CDF)  Quantify the proportion of the distribution area therefore ranges between 0 and 1. |
| OVL2: Proportion of overlap relative to the joint distributions1: |
| U1: Proportion of nonoverlap relative to the joint distribution, i.e. percentage of scores that do not overlap in the two distributions1: |
| U3: Proportion of individuals in the group with the higher mean who exceed the median individual of the other group1: |
| Probability of correct classification (PCC)1:  The probability that a that a randomly picked individual will be correctly classified    This method is only a good approximation if assumptions of normality and equality of variances/covariances are satisfied  The probability of success greatly depends on the statistical model used to perform the classification task. |
| Tail ratios1,22:  With *d* being the value of Cohen’s *d* (for dependent groups),  and *z* being the user selected number of (pooled) standard deviations.  Proportion of pretest scores relative to posttest scores (since *d* is defined as ) that can be found z pooled standard deviations away from the mean of the pretest scores.  The tail ratio can also be reported on a logarithmic scale – this is recommended to ensure comparibility22. |
| Assuming normality but not homoscedasticity |
| Glass’ (for dependent groups) 2,3, 10,15  Mean difference standardized by the standard deviation of either measurement. Since the variability is likely to increase after treatment, some authors advise using the standard deviation of the pre-treatment measurement, so as not to underestimate the effectiveness of the treatment2,3.  The positive bias of the estimate can be once again reduced3:  With df being |
| Variance ratio (*VR*)1:  Ranging from 0 to (theoretically) ∞. Often the group with the (usually) higher variation is placed in the numerator.  Sometimes a logarithmized version is employed:  The benefit of the log-transformed VR is that it is no longer bounded. A VR of 1 corresponds to a log VR of 0. A VR > 1 corresponds to log VR > 0 and a VR < 1 corresponds to a log VR < 0. Theoretically possible values range from -∞ to ∞ |
| Tail ratios1,22:  With *z* being the user selected number of standard deviations of the pretest scores.  Proportion of posttest scores relative to prestest scores that can be found z standard deviations away from the mean of the pretest scores. The tail ratio can also be reported on a logarithmic scale – this is recommended to ensure comparibility22. |
| Standardized difference effect sizes that are somewhat outlier resistant |
| 2.1.4.1 Trimmed mean differences assuming homoscedasticity |
| The scaled robust d statistic (for dependent groups) 10: With being the square root of the pooled winsorized variance:  Since the number of pre and post measurement should be equal, can be written as:  In this case will equal .  Trimmed mean difference standardized by the square root of the unweighted mean of the two group’s winsorized sample variances. |
| 2.1.4.2. Trimmed mean differences not assuming homoscedasticity |
| The scaled robust statistic (for dependent groups)3,10: Trimmed mean difference standardized by the squared root of the winsorized variance of the j-th group. |
| 2.1.4.3. Trimmed mean of difference scores standardized by the winsorized standard deviation of difference scores |
| The scaled robust statistic5: With **x** being the pre- and **y** being the posttreatment scores,  being the null-hypothesized mean of the difference scores (per default = 0),  and with being the winsorized standard deviation of the difference scores  *Trimmed Mean of the difference scores standardized by the winsorized standard deviation of the difference scores. This statistic estimates the number of winsorized difference score standard deviations the trimmed mean of these difference scores lies from 0 (or more generally).* |
| nonparametric methods |
| Nonparametric Glass’ \* (for dependent groups)8,9:  With being the proportion of pretreatment measurements that are lower than the median posttreatment measurement.  Either the median of all **y** values are considered or the median of specific **y** values – like the median of the posttreatment measures of those individuals whose pretreatment score was equal to either the median of all pretreatment scores or to the score one unit above or below this media9.  Alternatively, one could look at:  With being the proportion of posttreatment scores that lie above the median of the pretreatment measurements.  When *n* is small might be equal to 0 or 1. To avoid consequently extreme effect sizes is set to or respectively in such cases.  If the observations are independently normally distributed, these nonparametric statistics estimates Glass’ , i.e. estimates , and estimates . |
| Nonparametric Cohen’s dz8:With being the proportion of individuals whose scores increase from pretest to posttest: If ties between pre- and posttest scores are present, each tie should be counted as half an observation in each direction (**y** > **x** & **y** < **x**).  If the scores are normally distributed this statistic estimates Cohen’s *dz*. |
| Probability of superiority for independent groups2,3,5:    w is the number of wins of each subject of one condition/time  Confidence Interval based on Pratt‘s confidence interval in Wilcox5. |
| Generalized Odds ratio2,3,20:    Probability of superiority of group a compared to group b divided by the probability of superiority of group b compared to group a  Confidence Interval based on Browne et al., 2010  Question: How should ties in this context be handled? Both options (ignoring ties, counting them as 0.5 are implemented) |
| Dominance measure2,3,5:    Probability of superiority of group a compared to group b minus probability of group b compared to group a  Confidence interval based on Pratt‘s Method and the transformation formula: |
| Overlapping coefficient8:  Proportion of overlap relative to a single distribution  nonparametric estimation based on Schmid & Schmidt (2006):  1. Density estimation of the two distributions  2. Approximation of underlying function  3. Approximation of integral of minimum of both functions |
| U31:  Proportion of individuals in the group with the higher mean who exceed the median individual of the other group  nonparametric estimation:  Calculate exact value  1. Calculate means  2. Calculate number of individuals in group of higher mean that exceed mean of other group  3. Determine proportion |
| Tail ratios1,22:   |  |  |  |  | | --- | --- | --- | --- | |  | Posttest scores | Pretest scores |  | | Tail region of interest |  |  |  | | Rest of the distribution |  |  |  | |  |  |  |  |   Proportion of posttest scores relative to pretest scores that can be found below/above a specified cut-off value, i.e. in the tail region of interest.  The tail ratio can also be reported on a logarithmic scale – this is recommended to ensure comparibility22. |
| Pretest-posttest control-group designs |
| parametric methods |
| Standardized mean change difference 16,18: With being the pre- and being the posttreatment scores of groups a and b,  being the null-hypothesized mean of the difference scores (per default = 0),  and with being:  With N being the number of pairs.  Difference between the standardized mean change scores of groups a and b. Some authors consider the standardized mean change scores and thus the difference between them biased. Using the standard deviation of the difference scores as an effect size standardizer confounds the treatment effects with the correlation between pre- and posttest scores. |
| 3,16,17: Difference between the standardized mean change for groups a and b. The mean change is standardized by the standard deviation of the pretest scores. The individual standardized mean changes have a positive bias, which can be corrected for:  With being  Using the standard deviation of the pretest scores assures that in the case of an increased variance of posttest scores due to treatment the population effect size does not get underestimated. |
| 3,17: With being the pooled pretreatment standard deviation  Difference between the mean change for groups a and b. Homoscedasticity is assumed between (at least) the pretreatment measurements. In case of homoscedasticity the pooled standard deviation is a better estimate of the common population standard deviation. As a consequence of pooling, the sampling variance of this effect size estimate is smaller than that of . The positive bias of the estimate can be alliveated  With *df* being  Using the pooled standard deviation of the pretest scores assures that in the case of an increased variance of posttest scores due to treatment the population effect size does not get underestimated. |
| 3,17: With being the pooled standard deviation  Difference between the mean change for groups a and b standardized by the pooled standard deviation. Homoscedasticity is assumed between pre- and posttest scores for both populations. In case of homoscedasticity the pooled standard deviation is a better estimate of the common population standard deviation. Pooling all four sample variances results in a slight reduction of the sampling variance of the effect size estimate compared to .  This estimate will also have a positive bias, but an exact correction is not known, since pre- and posttest scores are dependent. Thus, while applying the usual correction is better than nothing, the bias will likely be underestimated, especially if pre- and posttest scores are highly correlated. In this case, the amount of information gained by pooling dependent pre- and posttest scores will be less than if the groups were independent.  With *df* being  Further, if homoscedasticity between pre- and posttest is not given due to an increased variability of posttest scores as a result of the treatment, this statistic would estimate an effect size that is different from the one it is assumed to estimate. |
| nonparametric methods |
| Nonparametric standardized mean change difference 8:With being the proportion of individuals whose scores increase from pretest to posttest in groups a and b: If ties between pre- and posttest scores are present, each tie should be counted as half an observation in each direction ( > & < ).  If the scores are normally distributed this statistic estimates the standardized mean change difference. |
| Nonparametric 8: With being the proportion of pretreatment measurements that are lower than the median posttreatment measurement.  Either the median of all values are considered or the median of specific values – like the median of the posttreatment measures of those individuals whose pretreatment score was equal to either the median of all pretreatment scores or to the score one unit above or below this media9.  When observations are independently normally distributed this statistic estimates . |
| Nonparametric alternative to 8: With being the proportion of posttreatment scores that lie above the median of the pretreatment measurements.  When observations are independently normally distributed this statistic estimates . |
| Miscellaneous |
| 4.1 The counternull value of an effect sizes |
| The counternull effect size:2,3  When null-hypothesized value of (the population effect size) equals zero:  The counternull interval: |

References

1Del Giudice, M. (in press). Measuring sex differences and similarities. In D. P. VanderLaan & W. I. Wong (Eds.), *Gender and sexuality development: Contemporary theory and research*. Springer.

2Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. Lawrence Erlbaum Associates Publishers.

3Grissom, R. J., & Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate applications* (2nd ed.). Taylor and Francis Group.

4Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, Article 863. <https://doi.org/10.3389/fpsyg.2013.00863>

5Wilcox, R. R. (2017). *Introduction to robust estimation and hypothesis testing* (4th ed.). Elsevier.

6Hedges, L. V. (1981). Distribution theory of glass’s Estimator of effect size and related estimators. Journal of Educational and Behavioral Statistics, 6(2), 107-128. <https://doi.org/10.3102%2F10769986006002107>

7Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>

**8**Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

9Kraemer, H. C., Andrews, G. (1982). A nonparametric technique for meta-analysis effect size calculation. *Psychological Bulletin, 9*(2), 404-412. <https://doi.org/10.1037/0033-2909.91.2.404>

10Keselmna, H. J., Algina, J., Lix, M. L., 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>

11Peng, 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>

12Algina, J., Keselmna, H. J., & Penfield, R. (2005). An alternative to Cohen’s standardized mean difference effect size: A robust parameter and confidence interval in the two independent groups case. Psychological Methods, 10(3), 317-328. <https://doi.org/10.1037/1082-989X.10.3.317>

13Algina, J., Keselman, H. J., & Penfield, R. (2006). Confidence intervals for an effect size when variances are not equal. Journal of Modern Applied Statisstical Methods, 5(1), 2-13. <https://doi.org/10.22237/jmasm/1146456060>

14Algina, J., Keselman, H. J., & Penfield, R. (2005). Effect sizes and their intervals: The tow-level repeated measures case. Educational and Psychological Measurement, 65(2), 241-258. <https://doi.org/10.1177/0013164404268675>

15Cumming, G. (2012). Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis. Routledge.

16Morris, S. B., DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychological Methods, 7(1), 105-125. https://doi.org/10.1037//1082-989X.7.1.105

17Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control-group designs. Organizational Research Methods, 11(2), 364-386. https://doi.org/10.1177/1094428106291059

18Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods, 14*(1), 43-53. https://doi.org/10.1037/a0014699

19Newcombe, R. G. (2005). Confidence intervals for an effect size measure based on the Mann–Whitney statistic. Part 2: Asymptotic methods and evaluation. *Statistics in Medicine*, *25*(4), 559–573. <https://doi.org/10.1002/sim.2324>

20Browne, R. H. (2010). *The t-Test p Value and Its Relationship to the Effect Size and P(X>Y)*. 5.

21Anderson, G., Linton, O., & Whang, Y.-J. (2012). Nonparametric estimation and inference about the overlap of two distributions. *Journal of Econometrics*, *171*(1), 1–23. <https://doi.org/10.1016/j.jeconom.2012.05.001>

22Voracek, M. (2013). On the importance of tail ratios for psychological science. *Psychological Reports: Measures and Statistics, 112*(3), 872-886. https://doi.org/10.2466/03.PR0.112.3.872-886