# 9 Technical Information

This chapter outlines various equations and procedures used in ***Meta-Stat***. The equations are based on generally accepted statistical principles and on methods described in Glass, McGaw, and Smith (1981), Hunter and Schmidt (1990), Cohen (1977), Veldman (1967), Hedges and Olkin (1985), Draper and Smith (1966), and SPSS (1991). The reader is referred to these excellent materials for a fuller development of the technical approaches used in ***Meta-Stat***.

# Conversion formulas

Meta-analysis is based on placing the results from different research studies on a common metric. ***Meta-Stat*** allows the user to use either effect size or correlation as that common metric. To calculate effect size from a ***t, F,*** $\chi^2$, or ***p-value***, the reported statistics are first converted to a correlation (***r***) and the correlation is converted to an effect size (***d***). Taylor and White (1993) found that the alternative methods for computing effect size yield very comparable results.

The following formula is used to convert from an ***r*** to effect size:

$$d = \sqrt{\frac{4r^2}{1-r^2}}$$

Correlational meta-analysis simply uses the ***r*** as the common metric.

- To derive ***r*** from a ***t***-statistic:

$$r = \sqrt{\frac{t^2/(1-r_{xy})^2}{t^2/(1-r_{xy})^2 + df}}$$

- To derive ***r*** from an ***F***-statistic:

$$r = \sqrt{\frac{F}{F + df}}$$

Here *df* refers to the denominator degrees of freedom. The equation is only appropriate when there is 1 degree of freedom in the numerator.

- To derive ***r*** from a $\chi^2$-statistic with 1 degree of freedom:

$$r = \sqrt{\chi^2/N}$$

- To derive ***r*** from a probability level (***p***), ***p*** is first converted to a ***z*** and the ***z*** is then converted to an ***r***:

$$z = a - \frac{2.515517 + .802853a + .010328a^2}{1 + 1.432788a + .189269a^2 + .001308a^3}$$

where

$$a = \sqrt{\ln\,(1\ /\ p^2)}$$

$$r = \sqrt{z^2\ /\ N}$$

! To compute effect size as the standardized mean difference:

$$d = \frac{\bar{x}_e - \bar{x}_c}{\sigma_c \sqrt{1 - r_{ec}^2}}$$

! To compute effect size as parametric gain scores:

$$d = \frac{\bar{x}_{e,post} - \bar{x}_{e,pre}}{\sigma_{e,pre}} - \frac{\bar{x}_{c,post} - \bar{x}_{c,pre}}{\sigma_{c,pre}}$$

! To compute effect size from a paired comparison t-test

$$d = \bar{D}\,\frac{\sqrt{N}}{t}$$

where $\bar{D}$ is the mean difference between the observed pairs.

This last equation is from Gibbons, Hedeker, and Davis (1993).

# Pre-coded variables

There are precoded equations in both the effect size and correlational meta-analysis modules. Here, we discuss these equations and outline their use.

## Effect size meta-analysis

Two pre-coded variables are used in the effect size meta-analysis module, UNBIASED, and WEIGHT. Both of these variables are based on the work of Hedges and Olkin (1985).

The unbiased effect size, $d'$, is used in several homogeneity calculations and as the bases for the effect size plot. It should be noted that $d'$ usually differs only slightly from *d*, especially as N gets large (see Hedges and Olkin, 1985, pp 78-79.). Taylor and White (1993) found that $d'$ and *d* produce the same conclusions.

$$d' = d\left(1 - \frac{3}{4N_t - 9}\right)$$

where $N_t$ is the total sample size and *d* is the original effect-size.

The variable WEIGHT is the inverse of the variance. The variance is used in the effect size plot. The optimal weight when analyzing $d'$ and estimating the true effect size is the inverse of the variance. The most precise studies are given the greatest weight. The more precise studies are give higher weights (see Hedges and Olkin, 1985, pp 302-304).

$$WEIGHT = \frac{2N_t N_e N_c}{2N_t^2 + N_e N_c (d')^2}$$

## Correlational meta-analysis

The correlational meta-analysis modules have numerous precoded variables, including adjusted correlation, weights, and attenuation formulas.

## Adjusted correlation and weights

The adjusted correlation is the unadjusted correlation divided by the product of the correction factors. This is the what the correlation would be if it were not for artifacts in the data.

Two suggested weights are provided based on discussions by Hunter and Schmidt (1990, pages 145-150). WEIGHTB is the optimal weight presented by Hedges and Olkin (1985):

$$W_b = CF^2 \frac{N_t - 1}{(1 - r^2)^2}$$

CF, the product of the correction factors (attenuation factors to use the terminology of Hunter and Schmidt), is discussed in the next section.

In the case of multiple independent variables, $W_b$ is multiplied by

$$1 + r^2 + \left(\frac{\sigma}{\sigma_{ref}} - 1\right)$$

Hunter and Schmidt show that WEIGHTA = $N_t\ CF^2$ can differ only trivially from WEIGHTB and has the desired effect -- the greater the correction needed for a study, the less weight it is given. As with the effect size weight above, you may want to use one of these weights to obtain an improved true value estimate of the corrected (adjusted) correlation.

## Attenuation Formulas

Attenuation formulas are used in correlational meta-analysis to adjust for characteristics of the data (***artifacts*** to use the terminology of Hunter and Schmidt). They provide a theoretical estimate of the attenuation in the true correlation due to data characteristics.

If $r_o$ is the observed correlation, ***p*** is the true correlation, and ***CF*** is a correction factor for some data characteristic, then $p = r_o/CF$. Correction factors can be computed for multiple data characteristics and are assumed to be independent.

- To correct for unreliability in either the dependent or independent variable:

$$CF = \sqrt{r_{xx}}$$

where $r_{xx}$ is the reliability of the measure used to assess the dependent or independent variable.

- To correct for variable dichotomization (i.e., to assess the correlation for a continuous variable given a point biserial correlation):

$$CF = \frac{\phi(c)}{\sqrt{P(1-P)}}$$

where ***c*** is the point in the normal distribution that divides the distribution into proportions ***P*** and ***1-P***, i.e., the inverse of the cdf (values for ***c*** are derived using a look-up table), and

$$\phi(c) = \frac{e^{-\frac{1}{2}c^2}}{\sqrt{2\pi}}$$

i.e., the normal ordinate at ***c***.

- To correct for the use of a proxy dependent or independent variable:

$$CF = r_{xy}\sqrt{r_{xx}}$$

where $r_{xy}$ is the correlation of the proxy variable with the variable it is replacing and $r_{xx}$ is the reliability of the proxy variable.

- To unpartial a partial correlation:

$$CF = 1 / \sqrt{1-r_{zy}^2}$$

where $r_{zy}$ is the correlation of the extraneous variable with the dependent variable.

- To correct for restriction in range for the independent variable:

$$CF = \sqrt{u^2 + r_o^2(1 - u^2)}$$

where ***u*** is the ratio of the study standard deviation to the reference standard deviation.

# Probability

Probabilities for various statistics are based on Kendall's (1955) normalizing transformation of the F distribution:

$$z = \frac{\left(1-\frac{2}{9B}\right)F^{1/3} - \left(1-\frac{2}{9A}\right)}{\sqrt{\frac{2}{9B}F^{2/3} + \frac{2}{9A}}}$$

where A is the df for the numerator of the F ratio, and B is the df for the denominator.

If B<4 then ***Z*** is transformed using Kelley's (1947) correction

$$z' = z + \frac{.08z^5}{B^3}$$

The ***z*** or ***z'*** then provides the basis for approximating the ***p-level*** using:

$$p = .5/(1+.196854z+.115194z^2+.000344z^3+.019527z^4)^4$$

The procedure is used to provide the exact probability of other statistics:

***z*** Since $F_{1,\infty} = z^2 = t^2_\infty$, the probability for a ***z*** is computed using ***A=1, B=1000,*** $F=z^2$.

$\chi^2$ Since $F = \chi^2/df$ with $(df,\infty)$ degrees of freedom, the probability for a chi-square statistic is computed using ***A=df, B=1000,*** $F=\chi^2/df$.

***t*** Since $F = t^2$ with (1,df) degrees of freedom, the probability for a t-statistic is computed using ***A=1, B=df,*** $F=t^2$.

# Statistical Analyses

The following identifies the equations used in the analysis module of ***Meta-Stat***

## Descriptive statistics

This procedure provided the weighted and unweighted mean, median, variance, standard deviation, minimum value, maximum value, and range given a specified variable and an optional weighting variable.

The computational formula used to derive the mean and variance are:

$$Mean = \sum (x_i w_i) \; / \; \sum w_i$$

where $x_i$ is dependent variable data and $w_i$ is the weight for observation $i$ ($w_i$ is set equal to 1 for all i if a weighting variable is not specified)

$$\sigma^2 = \frac{\sum w_i (x_i)^2}{\sum w_i} - \left( \frac{\sum x_i w_i}{\sum w_i} \right)^2$$

The standard deviation is simply the positive square root of the variance. This is the standard deviation of the presented data. To treat this data as a sample and to obtain an unbiased estimate of the population variance, you can multiply the sample variance by ***n/(n-1)***.

If the dependent variable is the unbiased effect size then a variety of additional statistics are available. Hedges's $Q_T$ can be used to test the homogeneity of the adjusted effect sizes.

$$Q_T = \sum \frac{(d' - \overline{d'})^2}{\sigma^2_{d'}}$$

where $\sigma^2_{d'}$ is the variance of $d'$ and is computed as the inverse of WEIGHT.

The computational formula used by ***Meta-Stat*** is

$$Q_T = \sum \frac{d'^2}{\sigma^2_{d'}} - \frac{\left( \sum \frac{d'}{\sigma^2_{d'}} \right)^2}{\sum 1/\sigma^2_{d'}}$$

If you page down after the presentation of the basic descriptive statistics for the adjusted effect size, another page of meta-analysis statistics appears.

Mean ES or mean effect size is the mathematical average described above. It is repeated on this screen for reference. To compute the Fisher's Z, each effect size is first converted to an r using the reverse of the formula above:

$$r = \frac{d}{\sqrt{4 + d^2}}$$

The $r$ is then converted to an individual $Z$ using

$$Z = .5 \ln \frac{1+r}{1-r}$$

Fisher's Z is then computed as the weighted average of the individual Z's. The probability test the null hypothesize of no population effect. By converting the Fisher's Z back to an effect size or an r (equivalent es and equivalent r), we have a refined estimate of the population effect size and correlation. The Failsafe N identifies the number of studies with opposite conclusions that would be needed to overturn a rejected null hypothesis.

$$Fail\ safe\ N = \frac{N_t\ (Z^2 - 3.8416)}{3.8416}$$

The variance ratio is the ratio of the effect size sampling error variance $\sigma_e^2$ to the variance of the overall mean effect size. The error variance is computed as

$$\sigma_e^2 = \left(\frac{N-1}{N-3}\right)\left(\frac{4}{N}\right)\left(1+\frac{d^2}{8}\right)$$

where N is the average sample size, and *d* is the mean effect-size. If EFFECTSZ rather than UNBIASED is the criterion variance, *d* is corrected by dividing by *1 + .75/(N-3)*.

The sampling error variance is discussed by Hunter and Schmidt (1990, page 281-338).

## Regression

***Meta-Stat*** uses the iterative stepwise multiple regression approach developed by Greenberger and Ward (1956) and described in Veldman (1967, pages 294-307). The iteration process begins by selecting the variable with the highest correlation available from the set of predictor variables. In subsequent iterations, ***Meta-Stat*** either selects the predictor variable which will maximally increase the square of the multiple correlation when added to the already selected predictor variables or adjusts the weight of one of the previously selected predictor variables. The process is terminated when neither adjustments nor the addition of variables increases the squared multiple correlation by more than .00001.

The key advantage of this iterative approach is that it allows the computation of key regression statistics ($R^2$ and beta weights) without having to invert the correlation matrix. Thus the non-multicolinearity restriction is not needed for computing the multiple correlation and accuracy is improved. The standard errors for the beta weights, and hence the corresponding t-statistics, however, are based on the inverted the correlation matrix. The Gauss-Jordon procedure is used to do the inverting.

Some relevant equations are:

! Adjusted R-Square:

$$Adjusted\ R^2\ '\ R^2\ \&\frac{(1\&R^2)p}{N\&p\%1}$$

where N is the number of observations and p is the number of included predictor variables.

! Residual Sum of Squares:

$$SS_e\ '\ (1\&R^2)(C\&1)S_{yy}$$

where $S_{yy}$ is the sample variance for the criterion variable and C is the sum of the weights.

! Regression Sum of Squares:

$$SS_R\ '\ R^2(C\&1)S_{yy}$$

! Mean squares

$$MS\ '\ SS/df$$

! Standard error of the unstandardized regression weight:

$$\hat{o}_{\hat{A}_k}\ '\ \sqrt{\frac{aa_{kk}\ (1\&R^2)\ S^2_{yy}}{S^2_k\ (N\&p\&1)}}$$

where $aa_{kk}$ is the kth diagonal element of the inverted correlation matrix and $S_k$ is the sample variance for the kth predictor variable..

! t-Statistic for testing $H_o$: $\&_k = 0$

$$t_k = \frac{\hat{\beta}_k \frac{S_{yy}}{S_k}}{\hat{\sigma}_{\hat{\beta}_k}}$$

If the criterion variable is UNBIASED and the weighting variable is WEIGHT, then we follow the recommendations outlined on page 174 of Hedges and Olkin (1985). The weighted sum of squares about the regression is the chi-square statistic $Q_E$ and the weighted sum of squares due to the regression $Q_R$. The corresponding probabilities are reported. The standard errors of the regression weights are corrected by dividing by the square root of the mean square error.

### Group Means

This module provides means, standard deviations and confidence intervals by group for the selected criterion variable. Means and standard deviations are computed using the routines from the descriptive statistics discussed above. The 95% confidence intervals are computed using

$$c.i. = \bar{X} \pm t_{n-1,.05} \frac{s}{\sqrt{n}}$$

Values of *t* are obtained using a look-up table.

If the criterion variable is the UNBIASED effect size, ***Meta-Stat*** also computes Hedges's within, between and total homogeneity statistics (Hedges and Olkin, 1985, pp 153-165). The total homogeneity statistic tests whether the studies, regardless of the grouping variable, share the same effect size and is discussed above under descriptive statistics.

Hedges's between homogeneity test is analogous to the analysis of variance F-test examining whether the group means are the same. The tested hypothesis is that the group means are the same. The test statistic is:

$$Q_B = \sum_i \sum_j \frac{(d_{i.} - d_{..})^2}{\sigma^2(d_{ij})}$$

where $d_{i.}$ is the group mean weighted by the inverse of the variance, $d_{..}$ is the grand mean weighted by the inverse of the variance, and $\sigma^2(d_{ij})$ is the variance of the unbiased effect size estimator.

$Q_B$ is tested with p groups-1 degrees of freedom.

Hedges within homogeneity test examines the hypothesis that the effect sizes are homogeneous within group. The test statistic is

$$Q_W = \sum_i \sum_j \frac{(d_{ij} - d_{..})^2}{\sigma^2(d_{ij})}$$

$Q_W$ is tested with N-p degrees of freedom.

Hedges and Olkin (1985, p 156) note that if the samples sizes are at least 10 per group and if the effect sizes are not too large, then the actual significance of these test statistics will be sufficiently close to the nominal va;lues that would be obtained from a large sample distribution.

# Graphical Analysis

Descriptions of the various graphs can be found in Chapter 8. The following outlines some of the equations used in the program.

## Effect size plots

The approximate 95% confidence interval for the unbiased effect size is given by

$$d' - z_{\alpha/2}\sigma_{d'} \quad < \quad \delta \quad < \quad d' + z_{\alpha/2}\sigma_{d'}$$

where $z_{\alpha/2}$ is the two-tailed critical value of the standard normal distribution and $\sigma_{d'}$ is the positive square root of the variance of $d'$ computed as the square root of the inverse of WEIGHT.

## Mean Plots

The approximate 95% confidence interval about the mean is given by

$$\bar{x} - z_{\alpha/2}\sigma_{n-1} \quad < \quad \mu \quad < \quad \bar{x} + z_{\alpha/2}\sigma_{n-1}$$

where $z_{\alpha/2}$ is the two-tailed critical value of the standard normal distribution and $\sigma_{n-1}$ is the standard deviation estimate.