About Effect Size Estimation for BG and WG Designs

2-BG

```
      Variable   ACTYLVL    activity level - DV
   By Variable   COND       treatment conditio - IV
```

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	1	72.0000	72.0000	4.6452	.0467
Within Groups	16	248.0000	15.5000		
Total	17	320.0000			

df_{Error} (16) SS_{Effect} (72.0000) SS_{Error} (248.0000) MS_{Error} (15.5000)

n (Count)

Group	Count	Mean	Standard Deviation
Grp 1	9	28.6667	3.0822
Grp 2	9	32.6667	4.6368

r = η (there are both “df” and “n” based formulas)

$$\mathbf{r} = \sqrt{(SS_{effect} / SS_{rotal})} = \sqrt{(SS_{effect} / (SS_{effect} + SS_{error}))} = \sqrt{(72.00 / 320.00)} = \sqrt{[72.00 / (72.00 + 248)]} = \mathbf{.474}$$

$$\mathbf{r} = \eta = \sqrt{[F / (F + df)]} = \sqrt{[4.645 / (4.645 + 16)]} = \mathbf{.474}$$

$$\mathbf{r} = \eta = \sqrt{[F / (F + N)]} = \sqrt{[4.645 / (4.645 + 18)]} = \mathbf{.453}$$ "N" = number of data points = number of subs

d (there are both “df” and “n” based formulas)

$$\mathbf{d} = \frac{M_1 - M_2}{\sigma} \quad \text{where } \sigma = \sqrt{\left(\frac{[(n_1 - 1) * \sigma_1] + [(n_2 - 1) * \sigma_2]}{n_1 + n_2}\right)} = \sqrt{MS_{error}}$$

$$\mathbf{d} = (M_1 - M_2) / \sqrt{MSe} = (32.6667 - 28.6667) / \sqrt{15.5} = \mathbf{1.016}$$

$$\mathbf{d} = 2 * \sqrt{F} / \sqrt{N} = 2 * \sqrt{4.645} / \sqrt{18} = \mathbf{1.016}$$

$$\mathbf{d} = 2 * \sqrt{F} / \sqrt{df_{Error}} = 2 * \sqrt{4.645} / \sqrt{16} = \mathbf{1.076}$$

r & d

$$d = \sqrt{\left(\frac{4 * r^2}{1 - r^2}\right)} = \sqrt{\left(\frac{4 * .474^2}{1 - .474^2}\right)} = 1.077 \qquad r = \sqrt{\left(\frac{d^2}{d^2 + 4}\right)} = \sqrt{\left(\frac{1.016^2}{1.016^2 + 4}\right)} = .453$$

$$d = \frac{2 * r}{\sqrt{(1 - r^2)}} = \frac{2 * .474}{\sqrt{(1 - .474^2)}} = 1.077$$

So far so good -- the various "direct computation" formulas all match and the transformations are pretty close !! Notice that the transformation preserves the difference between the "df_{Error}" and the "N" versions -- though some references suggest this is "computational" or "approximation" error.

k-BG

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	2	1194.0000	597.0000	4.2838	.0222
Error → Within Groups	33	4599.0000	139.3636		
Total	35	5793.0000			

Some values used in the following calculations

k -1 (# IV conditions-1) **df** $_{\mathbf{error}}$ **SS**$_{\mathbf{Error}}$ **SS**$_{\mathbf{Effect}}$ **MS**$_{\mathbf{Error}}$

n (# subjects in each IV condition)

Group	Count	Mean	Standard Deviation
Grp 1	12	40.0000	10.7956
Grp 2	12	26.0000	15.2017
Grp 3	12	34.5000	8.3937
Total	36	33.5000	12.8652

Let's focus on the *pairwise* stuff -- group 1 vs. group 3

d = mean difference / $\sqrt{MS_{Error}}$ = $(40 - 34.5) / \sqrt{139.36}$ = **.466**

r -- without pairwise F-values (whether based on separate or full model error terms) we are limited to the d-to-r transformation (estimation) …

$$\mathbf{r} = \eta = \sqrt{[d^2 / (d^2 + 4)]} = \sqrt{[.466^2 / (.466^2 + 4)]} = .227$$

So, since the d-to-r transformations looked pretty good for the 2-BG model, we can consider using them as a way to get pairwise r values in a k-BG model when we don't have pairwise F-values (which takes us back to the full model question!).

So, here's the question… It would be convenient to use this same approach for k-WG designs. Namely, as a way to get r-values for pairwise comparisons in a WG model when all we have are the means and the error term.

One of the reasons for this is to do pairwise power analyses -- the idea being that the power of the overall study doesn't matter much if there isn't "enough" power for the smaller pairwise effects.

Here's were things get a bit dicey, because the "definitional formula" for d (starting with group mean and stds) doesn't work for WG designs -- the within-group std mixes error and subject variation that is separated by ANOVA when computing the MSerror. So, the likely approach is to use the mean difference & MSerror formula to compute the denominator of d and convert it to r. But ….

2-WG

	Mean	Std. Dev.	N
DV score - intermittent noise			
For entire sample	19.800	2.974	10
DV score - constant noise	23.100	2.685	10

$SS_{Subject}$ $MS_{Subject}$

Tests of Between-Subjects Effects.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	62.45	9	6.94		
CONSTANT	9202.05	1	9202.05	1326.16	.000

Tests involving 'NOISE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	82.05	9	9.12		
NOISE	54.45	1	54.45	5.97	.037

SS_{Error} SS_{IV} df_{IV} df_{Error} MS_{Error}

r

(brief editorial comment for your comment) Most meta analysts have embraced the use of the effect size r, which is the same as the upper bound of η as the effect size of choice, largely for four reasons: 1) it has the advantage that r and η are then defined the same for BG and WG designs, 2) it the equates the effect size with the significance test as much as possible, 3) it better equates r and d and 4) η and r can be computed directly from the within-groups F or t, as shown below. (Note: Other versions of these formulas use N rather than df. While the difference is usually small, using df gives a solution which is closer to that based on the SS formula shown above).

Lower bound $\eta = r = \sqrt{(SS_{effect} / SS_{total})} = \sqrt{(SS_{effect} / (SS_{effect} + SS_{sub} + SS_{error}))}$
$= \sqrt{54.45/(54.45 + 62.45 + 82.05)} = .523$

Upper bound η = **r** = $\sqrt{[SS_{effect} / (SS_{effect} + SS_{error})]} = \sqrt{[54.45 / (54.45 + 82.05)]}$ = **.631**

r = $\eta = \sqrt{[F / (F + df_{Error})]} = \sqrt{[5.970 / (5.970 + 9)]}$ = **.632**

r = $\eta = \sqrt{[F / (F + n)]} = \sqrt{[5.970 / (5.970 + 10)]}$ = **.611**

Notice that "n" is the "number of participants" or the "number of values contributing to each mean", not the "total number of data points". Though references vary!?!

d

d = $\sqrt{F} / \sqrt{n} = \sqrt{5.970} / \sqrt{10}$ = **.773**

d = $\sqrt{F} / \sqrt{df_{Error}} = \sqrt{5.970} / \sqrt{9}$ = **.814**

This is probably the most common formula for d. Dropping the "2*" (that is in the BG version of the formula) is an approximation. The "real" formula is $d = a * \sqrt{(F / df_{error})}$, where $a = \sqrt{[2(1-r_{12}) / r_{YY}]}$, with r_{12} = sample test-retest r and r_{YY} = DV test-retest reliability. Since neither of these values are available, the approximation of "1" is used.

When we try to apply this to k-WG designs we won't usually have pairwise F-values, so we need to be able to get to r via d. Which means we need a computation using mean dif & MSe that gets us a d that will take us back to the proper r. The one below is the most common -- it uses the same approximation as the one above.

d = $(M_1 - M_2) / \sqrt{(Mse * 2)} = (23.1 - 19.8) / \sqrt{(9.12*2)}$ = **.773** Notice it matches the "n" version from above.

Disturbingly (since it is important for getting pairwise effect sizes from k-WG designs) this is the formula that seems to vary most across references -- the difference being "2" vs. "√2" .

To transform d to r, first we "adjust" d for the difference between the BG and WG computations, then apply the d-to-r transformation. Notice that the result matches the r computed from n (rather than df), and isn't the r computed from the SS !!

$d_w = 2 * d = 2 * .773 = 1.546$

$$r = \sqrt{\frac{d^2}{d^2 + 4}} = \sqrt{\frac{1.546^2}{1.546^2 + 4}} = .611$$

So, we have formulas to apply to k-WG designs -- the goal !!