HUDM 5123 - Linear Models and Experimental Design
Notes 07 - Effect Size and Power

1 Introduction to Effect Size

An *effect size* is a quantity that measures the size of an effect as it exists in the population. The American Psychological Association (APA) and the American Educational Research Association (AERA) both advocate reporting effect sizes in addition to p values and t and F statistics for pairwise and multigroup comparisons, respectively. Consider the two group case where, for example, participants in group 1 were randomly assigned to a control group, while participants in group 2 were randomly assigned to a treatment group. Let Y_{ij} represents the outcome measurement for participant $i = 1, 2, \ldots, n_j$, in group $j = 1, 2$.

1.1 The t Statistic

The sample means for groups 1 and 2 are denoted as $\bar{Y}_{\cdot 1}$ and $\bar{Y}_{\cdot 2}$, and sample variances for groups 1 and 2 are denoted as s_1^2 and s_2^2, respectively, and defined as follows:

$$\bar{Y}_{\cdot 1} = \frac{1}{n_1}\sum_{i=1}^{n_1} Y_{i1} \qquad \bar{Y}_{\cdot 2} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_{i2}$$

$$s_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}\left(Y_{ij} - \bar{Y}_{\cdot 1}\right)^2 \qquad s_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}\left(Y_{ij} - \bar{Y}_{\cdot 2}\right)^2$$

Furthermore, the pooled variance, s_p, across both groups is defined as follows:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}.$$

The equation for the two-sample t statistic (assuming equal group variances) associated with the null hypothesis $H_0 : \mu_1 = \mu_2$ is given by

$$t = \frac{\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

If the data-generating populations are normally distributed, the statistic t has a t distribution with $n_1 + n_2 - 2$ degrees of freedom under the null hypothesis. As the group sample sizes go up, the normality assumption matters less and less due to the central limit theorem, which ensures that the sample mean is normally distributed in the limit.

1.2 The F Statistic

Let the sum of squared treatment deviations and mean squared treatment deviations be defined as follows, where the number of groups is $a = 2$:

$$SSTR = \sum_{j=1}^{2} n_j \left(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot}\right)^2$$

$$MSTR = \frac{SSTR}{a-1} = \frac{SSTR}{2-1} = SSTR.$$

Let the sum of squared error deviations and mean squared error deviations be defined as follows.

$$SSE = \sum_{j=1}^{2}\sum_{i=1}^{n_j} \left(Y_{ij} - \bar{Y}_{\cdot j}\right)^2$$

$$MSE = \frac{SSE}{n_1 + n_2 - a} = \frac{SSE}{n_1 + n_2 - 2}.$$

The equation for the F statistic (assuming equal group variances) is given by

$$F = \frac{MSTR}{MSE}$$

Again, if the data-generating populations are normally distributed, the statistic F has an F distribution with 1 numerator and $n_1 + n_2 - 2$ denominator degrees of freedom under the null hypothesis.

1.3 Equivalence of $t^2(df)$ and $F(1, df)$

We have discussed a few times that the F statistic with one numerator degree of freedom is equivalent to t^2. Here, we will prove that with some algebra. We will start with s_p^2 and show that it is identical to MSE.

$$\begin{aligned}
s_p^2 &= \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \\
&= \frac{(n_1-1)\frac{1}{n_1-1}\sum_{i=1}^{n_1}\left(Y_{ij} - \bar{Y}_{\cdot 1}\right)^2 + (n_2-1)\frac{1}{n_2-1}\sum_{i=1}^{n_2}\left(Y_{ij} - \bar{Y}_{\cdot 2}\right)^2}{n_1+n_2-2} \\
&= \frac{\sum_{i=1}^{n_1}\left(Y_{ij} - \bar{Y}_{\cdot 1}\right)^2 + \sum_{i=1}^{n_2}\left(Y_{ij} - \bar{Y}_{\cdot 2}\right)^2}{n_1+n_2-2} \\
&= \frac{\sum_{j=1}^{2}\sum_{i=1}^{n_1}\left(Y_{ij} - \bar{Y}_{\cdot j}\right)^2}{n_1+n_2-2} = \frac{SSE}{df_E} = MSE
\end{aligned}$$

We will next show that $\frac{\left(\bar{Y}_{\cdot 1}-\bar{Y}_{\cdot 2}\right)^2}{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$ is identical to $MSTR$.

$$
\begin{aligned}
\frac{\left(\bar{Y}_{\cdot 1}-\bar{Y}_{\cdot 2}\right)^2}{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \frac{\left(\bar{Y}_{\cdot 1}-\bar{Y}_{\cdot 2}\right)^2}{\left(\frac{n_1+n_2}{n_1 n_2}\right)} &= \frac{(n_1 n_2)\left(\bar{Y}_{\cdot 1}-\bar{Y}_{\cdot 2}\right)^2}{(n_1+n_2)} \\
&= \frac{(n_1 n_2)\left(\bar{Y}_{\cdot 1}^2 - 2\bar{Y}_{\cdot 1}\bar{Y}_{\cdot 2} + \bar{Y}_{\cdot 2}^2\right)}{(n_1+n_2)} = \frac{n_1 n_2 \bar{Y}_{\cdot 1}^2 - 2 n_1 n_2 \bar{Y}_{\cdot 1}\bar{Y}_{\cdot 2} + n_1 n_2 \bar{Y}_{\cdot 2}^2}{(n_1+n_2)} \\
&= \frac{n_1^2\bar{Y}_{\cdot 1}^2 + n_1 n_2 \bar{Y}_{\cdot 1}^2 + n_2^2\bar{Y}_{\cdot 2}^2 + n_1 n_2 \bar{Y}_{\cdot 2}^2 - n_1^2\bar{Y}_{\cdot 1}^2 - 2n_1 n_2 \bar{Y}_{\cdot 1}\bar{Y}_{\cdot 2} - n_2^2\bar{Y}_{\cdot 2}^2}{(n_1+n_2)} \\
&= \frac{n_1(n_1+n_2)\bar{Y}_{\cdot 1}^2 + n_2(n_1+n_2)\bar{Y}_{\cdot 2}^2 - \left(n_1\bar{Y}_{\cdot 1} + n_2\bar{Y}_{\cdot 2}\right)^2}{(n_1+n_2)} \\
&= n_1\bar{Y}_{\cdot 1}^2 + n_2\bar{Y}_{\cdot 2}^2 - \frac{\left[n_1\bar{Y}_{\cdot 1} + n_2\bar{Y}_{\cdot 2}\right]^2}{(n_1+n_2)} \\
&= n_1\bar{Y}_{\cdot 1}^2 + n_2\bar{Y}_{\cdot 2}^2 - \frac{1}{(n_1+n_2)}\left(\sum_{i=1}^{n_1} Y_{i1} + \sum_{i=1}^{n_2} Y_{i2}\right)^2 \\
&= n_1\bar{Y}_{\cdot 1}^2 + n_2\bar{Y}_{\cdot 2}^2 - \frac{1}{(n_1+n_2)}\left[\left(\frac{n_1+n_2}{n_1+n_2}\right)\left(\sum_{i=1}^{n_1} Y_{i1} + \sum_{i=1}^{n_2} Y_{i2}\right)\right]^2 \\
&= n_1\bar{Y}_{\cdot 1}^2 + n_2\bar{Y}_{\cdot 2}^2 - \frac{(n_1+n_2)^2}{(n_1+n_2)}\left[\left(\frac{1}{n_1+n_2}\right)\left(\sum_{j=1}^{2}\sum_{i=1}^{n_1} Y_{ij}\right)\right]^2 \\
&= n_1\bar{Y}_{\cdot 1}^2 + n_2\bar{Y}_{\cdot 2}^2 - (n_1+n_2)\,\bar{Y}_{\cdot\cdot}^2 \\
&= n_1\bar{Y}_{\cdot 1}^2 + n_2\bar{Y}_{\cdot 2}^2 - 2\,(n_1+n_2)\,\bar{Y}_{\cdot\cdot}^2 + (n_1+n_2)\,\bar{Y}_{\cdot\cdot}^2 \\
&= n_1\bar{Y}_{\cdot 1}^2 + n_2\bar{Y}_{\cdot 2}^2 - 2\,(n_1+n_2)\left[\left(\frac{1}{n_1+n_2}\right)\left(\sum_{j=1}^{2}\sum_{i=1}^{n_1} Y_{ij}\right)\right]\bar{Y}_{\cdot\cdot} + (n_1+n_2)\,\bar{Y}_{\cdot\cdot}^2 \\
&= n_1\bar{Y}_{\cdot 1}^2 + n_2\bar{Y}_{\cdot 2}^2 - 2\left[\sum_{j=1}^{2}\sum_{i=1}^{n_1} Y_{ij}\right]\bar{Y}_{\cdot\cdot} + (n_1+n_2)\,\bar{Y}_{\cdot\cdot}^2 \\
&= n_1\bar{Y}_{\cdot 1}^2 + n_2\bar{Y}_{\cdot 2}^2 - 2\left[\sum_{i=1}^{n_1} Y_{i1} + \sum_{i=1}^{n_2} Y_{i2}\right]\bar{Y}_{\cdot\cdot} + (n_1+n_2)\,\bar{Y}_{\cdot\cdot}^2 \\
&= n_1\bar{Y}_{\cdot 1}^2 + n_2\bar{Y}_{\cdot 2}^2 - 2n_1\bar{Y}_{\cdot 1}\bar{Y}_{\cdot\cdot} - 2n_2\bar{Y}_{\cdot 2}\bar{Y}_{\cdot\cdot} + n_1\bar{Y}_{\cdot\cdot}^2 + n_2\bar{Y}_{\cdot\cdot}^2 \\
&= n_1\left(\bar{Y}_{\cdot 1}^2 - 2\bar{Y}_{\cdot 1}\bar{Y}_{\cdot\cdot} + \bar{Y}_{\cdot\cdot}^2\right) + n_2\left(\bar{Y}_{\cdot 2}^2 - 2\bar{Y}_{\cdot 2}\bar{Y}_{\cdot\cdot} + \bar{Y}_{\cdot\cdot}^2\right) \\
&= n_1\left(\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot\cdot}\right)^2 + n_2\left(\bar{Y}_{\cdot 2} - \bar{Y}_{\cdot\cdot}\right)^2 \\
&= \sum_{j=1}^{2} n_j\left(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot}\right)^2
\end{aligned}
$$

Putting these facts together proves the identity.

$$t^2 = \frac{\left(\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}\right)^2}{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \frac{MSTR}{s_p^2} = \frac{MSTR}{MSE} = F$$

2 Effect Size Measures Based on Mean Differences

Importantly, the value of the t statistic (and, therefore, the F statistic) depends on the sample size. The problematic nature of this relationship can best be illustrated via example. Suppose the treatment and control groups, T and C, are both normally distributed and have identical variances of 225 (SDs = 15), but the mean of the treatment group is 110 and the mean of the control group is 100. Suppose, for simplicity, that the sample means and variances turn out to be exactly equal to the population values. For group sample sizes of size 10,

$$t = \frac{\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{10}{15\sqrt{\frac{1}{10} + \frac{1}{10}}} = 1.49$$

For group sample sizes of 100,

$$t = \frac{\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{10}{15\sqrt{\frac{1}{100} + \frac{1}{100}}} = 4.71$$

For group sample sizes of 1000,

$$t = \frac{\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{10}{15\sqrt{\frac{1}{1000} + \frac{1}{1000}}} = 14.91$$

and so on. The magnitude of the effect is the same in all three cases, even as the value of the t statistic gets larger and larger. Thus, the t statistic value (and, similarly, the F statistic value) is not a reasonable measure of effect size. Instead, we might consider using the numerator of the t statistic (i.e., the mean difference), $\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}$, which takes on the value 10 units no matter the sample size. A main motivation for using effect sizes is to have a measure of effect that can be meaningfully interpreted across studies. In this respect, the mean difference fails because the meaning of a ten unit difference across groups may vary widely for different studies that use variables on different scales. Imagine two scenarios to study the same treatment effect. In scenario 1, researchers use an outcome variable that has a standard deviation of only 1 in each group. In scenario 2, researchers use an outcome variable that has a standard deviation of 5 in each group. In both scenarios the mean difference is the same: 10 units. However, the magnitude of the effect in the scenario 1 is much larger than in scenario 2 relative to the variability. This is demonstrated in Figure 1, below, where the magnitude of the effect for groups represented by the black solid curves in Figure 1 is much larger than the magnitude of the effect for the groups represented by the dashed red curves even though the mean difference is 10 for both.

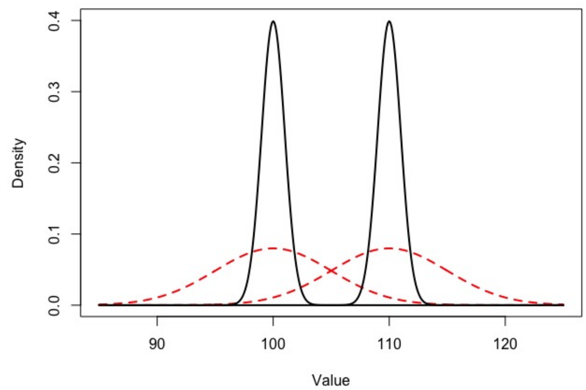

Figure 1: Density plots of two scenarios with the same group means but different group variances

A solution is to use the *standardized* mean difference, which is defined as the t statistic without the square root of the inverse sample sizes in the denominator. The standardized mean difference, sometimes referred to as Cohen's d, is defined as follows.

$$d = \frac{\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}}{s_p} = \frac{\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}}$$

Thus, the relationship between Cohen's d and the t statistic is simply that

$$d = t\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

Cohen's d, then, represents the number of pooled SDs by which two means differ. For the example above, for scenario 1, $d = 10/1 = 10$. For scenario 2, $d = 10/5 = 2$. In his foundational 1988 book on effect size and power analysis, Cohen suggested that 0.2, and 0.5, and 0.8 be used as benchmark values for small, medium, and large effects, respectively. In 1985, Hedges and Olkin showed that Cohen's d is a biased estimator for the population effect size. For medium to large samples, the bias is negligible. For small samples (i.e., $n < 25$ or so) the bias can be large. Hedge's g is identical to Cohen's d with the exception that d is multiplied by a correction term that renders it unbiased.

$$g = d\left(1 - \frac{3}{4(n_1+n_2)-9}\right)$$

3 Effect Size Measures Based on Association (Proportion of Variance Explained)

As you know, multiple R^2 is defined as the proportion of variance in the outcome that is explained by the multiple linear regression of the outcome on the predictors. That is,

$$R^2 = \frac{\text{TSS} - \text{RSS}}{\text{TSS}} = \frac{\text{RegSS}}{\text{TSS}}.$$

Cohen (1988) also provided guidelines for multiple R^2: $R^2 = .01$ is a small effect, $R^2 = .06$ is a medium effect, and $R^2 = .14$ is a large effect. In ANOVA designs, R^2 is given another name; it is called eta squared, η^2, even though it is defined identically to multiple R^2. The *complete* η^2 value for a categorical predictor, A, is calculated as the sum of squares for the factor A, divided by the total sum of squares.

$$\eta^2 = \frac{\text{SS}_A}{\text{TSS}}.$$

As an example, consider the one-way ANOVA table from the blood pressure data we examined in Notes 04.

Table 1: ANOVA table for the blood pressure data

Source	SS	*df*	MS	*F*	*p*
Treatment	335	3	111.67	1.66	.22
Error	1078	16	67.38		
Total	1413	19			

The value of complete η^2 may be calculated simply as $\frac{335}{335+1078} = 0.24$. According to Cohen's (1988) guidelines, this is a very large effect. In two-way (and higher-way) designs, it is possible to define *partial* η^2. The *partial* η^2 shares the same numerator as total η^2, but differs when it comes to the denominator. Instead of using the total sum of squares (TSS) in the denominator, partial η^2 only uses the sum of squares due to the factor of interest along with the residual sum of squares (SSE). Sums of squares due to any other effects in the model are not included in denominator.

$$\eta^2_{\text{partial}} = \frac{\text{SS}_A}{\text{SS}_A + \text{SSE}}.$$

As an example, consider the two-way ANOVA table from Data Set 1 from Lab 06. Note that I have added columns for complete and partial eta squared to the table.

Table 2: Two-way ANOVA table and complete and partial eta squared values for the fake Prestige data analyzed in Lab 06

Source	SS	*df*	MS	*F*	*p*	η^2	η^2_p
Income Category	1204	1	1204	11.25	.0013	.05	.15
Job Type	18695	2	9347.5	87.32	$<.0001$	.69	.73
Income:Type	71	2	35.5	0.33	.72	.003	.01
Error	7065	66					
Total	27035	72					

The overall η^2 for income category is $\eta^2 = 1204/27035 = .045$. The partial η^2 for income category is $\eta^2_p = 1204/(1204 + 7065) = .15$. As you can clearly see based on the definitions, partial eta squared will always be as large or larger than complete eta squared. Some properties of complete eta squared (from *Eta squared, partial eta squared, and misreporting of effect size in communications research*, by Levine and Hullett, 2002)

- Complete eta squared can be interpreted as the proportion of total variance in the outcome explained by the focal predictor.
- For factors with only two levels, the square root of complete eta squared represents the correlation, r, between the dichotomous factor and the outcome.
- The sum of the complete eta squared values for a factorial design (including error) is 1.
- As the number of factors in the model increases, the effect size for each particular effect decreases.
- Eta squared is more conservative than partial eta squared because eta squared is always less or equal to partial eta squared.

Unlike complete eta squared, partial eta squared (a) cannot be interpreted as the proportion of total sum of squares, (b) is not additive (i.e., partial eta squared values can sum to more than 1), (c) is not equivalent to multiple R^2, and (d) is more likely to over-estimate an effect. Thus, my general recommendation is that you use complete η^2 as an associational measure of effect size. That said, there is one case, discussed by Keppel and Wickens, in which it may make sense to use the partial version of eta squared. If you have a two way design where both factors are randomly assigned, and you are interested in determining the effect size of only one factor (say factor A) to use in a power analysis to plan a follow up study on only factor A (i.e., using a one-way ANOVA), then you are better off using the partial eta squared, because it will ignore any variability in the outcome *induced* by factor B, which is what you would observe in a one-way design with only factor A. Outside of this special case, my recommendation is that you use complete, rather than partial, measures of association. A final note on measures of association is that effect size measures like eta squared can also be used with contrasts. Consider a contrast for a one-way design with a factor that has r levels.

$$L = c_1\mu_1 + c_2\mu_2 + \cdots + c_r\mu_r.$$

Recall that the t statistic to test $H_0 : L = 0$ vs $H_0 : L \neq 0$ is given as follows.

$$t = \frac{\hat{L}}{s_{\hat{L}}},$$

where $\hat{L} = c_1\bar{Y}_1 + c_2\bar{Y}_2 + \cdots + c_r\bar{Y}_r$ and $s_{\hat{L}} = \sqrt{MSE\sum_{i=1}^{r}\left(\frac{c_i^2}{n_i}\right)}$. Putting it all together yields

$$t = \frac{\hat{L}}{\sqrt{MSE\sum_{i=1}^{r}\left(\frac{c_i^2}{n_i}\right)}}.$$

We can convert the t statistic to an F statistic by squaring it.

$$F = \frac{\hat{L}^2}{MSE\sum_{i=1}^{r}\left(\frac{c_i^2}{n_i}\right)}.$$

Further grouping yields

$$F = \frac{\left[\dfrac{\hat{L}^2}{\displaystyle\sum_{i=1}^{r}\left(\dfrac{c_i^2}{n_i}\right)}\right]}{MSE}.$$

The numerator here is the mean square associated with the contrast, L. But since contrasts only have 1 degree of freedom, the mean square is also the sum of squares. Thus, the formula for the sum of squares of a contrast is given by

$$\text{SS}_L = \frac{\hat{L}^2}{\displaystyle\sum_{i=1}^{r}\left(\dfrac{c_i^2}{n_i}\right)}$$

As an example, consider the positive psychology data from Labs 03 and 04. Recall that participants were randomly assigned to one of four groups. Three were grounded in positive psychology (signature strengths, three good things, and gratitude), and one not (early memories). Our interest for this contrast will center on testing the average of the positive psychology interventions against the fourth intervention (early memories). The one-way ANOVA table for this factor on the depression (CES-D total score) gain scores after three months:

Table 3: One-way ANOVA table for the positive psychology replication study data examined in Labs 03 and 04

Source	SS	*df*	MS	F	p
Treatment	96.0	3	32	0.40	.75
Error	10765.6	135	79.7		
Total	10861.6	138			

Next, to get the eta squared values for the contrast, we need to calculate the sum of squares for the contrast, SS_L.

$$\text{SS}_L = \frac{\hat{L}^2}{\displaystyle\sum_{i=1}^{r}\left(\dfrac{c_i^2}{n_i}\right)}$$

To estimate, we need to know group sample means and sample sizes.

```
table(rct_wide$int_fact)
```

```
 SS TGT  GV REM
 29  43  30  37
```

```
by(data = rct_wide$cesdGS, INDICES = rct_wide$int_fact, FUN = mean)
rct_wide$int_fact: SS
[1] -2.793103
------------------------------------------------------------------------
rct_wide$int_fact: TGT
[1] -3.023256
------------------------------------------------------------------------
rct_wide$int_fact: GV
[1] -4.5
------------------------------------------------------------------------
rct_wide$int_fact: REM
[1] -2.135135
```

$$\text{SS}_L = \frac{\hat{L}^2}{\sum_{i=1}^{r}\left(\frac{c_i^2}{n_i}\right)} = \frac{1/3(-2.79) + 1/3(-3.02) + 1.3(-4.5) - 1(-2.14)}{\frac{(1/3)^2}{29} + \frac{(1/3)^2}{43} + \frac{(1/3)^2}{30} + \frac{(-1)^2}{37}} = 49.82.$$

Thus, complete eta square for the contrast is $\eta^2 = 49.82/10861.6 = .005$, and partial eta squared for the contrast is $\eta_p^2 = 49.82/(49.82 + 10765.6) = .005$. Most analysts would not report an effect size for a non-significant contrast (here, this contrast is not significant), but we do so simply for demonstration.

4 Power and Sample Size

Recall the definitions of Type I and Type II error rate.

- **Type I error rate**: Pr(Reject H_0 given H_0 is true) $= \alpha$.
- **Type II error rate**: Pr(Do not reject H_0 given H_0 not true) $= \beta$.

Statistical power is defined in terms of the Type II error rate. The **power** of a test is the opposite of the Type II error rate. That is, the **power** of a test is the probability of rejecting the null hypothesis when it is false. Power $= 1 - \beta$. There are five factors that are all interrelated, and changing any one of them will affect the others; they are,

1. significance level (α),
2. power ($1 - \beta$),
3. effect size,
4. error variance, and
5. sample size.

An experimenter usually has direct control over the sample size and the Type I error rate, α, although .05 is the norm. An experimenter usually has no control (or very little control) over the effect size and the error variance. Sample size is usually used to control power. The purpose of a **power analysis** is typically to determine the minimum sample size that provides enough power for a given effect size, significance level, and error variance. Because we have to assume to know the truth about the effect size and error variance for power calculations, the effect sizes we use for power analysis are defined in terms of population parameters instead of sample quantities estimated from data. Thus, the names and formulas differ slightly from those presented above for Cohen's d and η_2.

4.1 Population-Based Effect Sizes

Recall from above the formula for Cohen's d, an effect size measure of the standardized mean difference between two groups:

$$d = \frac{\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}}{s_p} = \frac{\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}}$$

Note that the formula for d uses the best estimate for SE under the assumption of equal variances across groups, which is s_p, the pooled SE. Note that s_p is simply the square root of MSE for two groups. In a power analysis, however, we will need to assume that we know the population value of the SE, either based on a pilot study or theoretical considerations. Thus, the standardized mean difference most typically used in power analyses is also called Cohen's d, but uses the (assumed known) parameter value σ instead of s_p. That is,

$$d_{\text{population}} = \frac{\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2}}{\sigma}.$$

where σ is the square root of the error variance, σ^2, that is estimated by MSE in ANOVA designs. For measures of association like η^2, the effect size for the population is given a differ name, omega squared, ω^2. Like eta_2, both complete and partial versions can be defined. The complete ω^2 for factor A in an one-way ANOVA design:

$$\omega^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_{\text{Error}}^2}.$$

The complete ω^2 for a contrast, L, in an one-way ANOVA design:

$$\omega^2 = \frac{\sigma_L^2}{\sigma_A^2 + \sigma_{\text{Error}}^2}.$$

The partial ω^2 for a contrast, L, in an one-way ANOVA design:

$$\omega_p^2 = \frac{\sigma_L^2}{\sigma_L^2 + \sigma_{\text{Error}}^2}.$$

In a two-way design with factors A, B, and their interaction, the complete ω^2 for factor A:

$$\omega^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_B^2 + \sigma_{A:B}^2 + \sigma_{\text{Error}}^2},$$

and the partial ω^2 for factor A:

$$\omega^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_{\text{Error}}^2},$$

To make things slightly more confusing, Cohen, in his foundational 1988 text on power analysis and effect size, uses an effect size called f instead of ω^2, that is related to ω^2 as follows:

$$f = \sqrt{\frac{\omega^2}{(1-\omega^2)}} \quad \text{and} \quad \omega^2 = \frac{f^2}{(f^2+1)}.$$

Cohen's (1988) guidelines for f:

- $f = 0.10$ is a "small" effect. They tend not to be noticed except by statistical means.
- $f = 0.25$ is a "medium" effect. These effects are apparent to careful observation.
- $f = 0.40$ is a "large" effect. These effects are obvious to a superficial glance.

We will see some examples of power analyses in lab.