Chapter 6
Power Analysis for Categorical Moderator Models of Effect Size

Abstract This chapter provides methods for computing power with moderator models of effect size. The models discussed are analogues to one-way ANOVA models for effect sizes. Examples are provided for both fixed and random effects categorical moderator models. The power for meta-regression models requires knowledge of the values of the predictors for each study in the model, and is not provided here.

6.1 Background

The prior two chapters outlined the procedures for computing power to test the mean effect size and homogeneity in both fixed and random effects models. As discussed in Chap. 3, reviewers plan many syntheses to test theories about why effect sizes differ among studies. These theories are formally examined in the form of moderator analyses in meta-analysis. Given the variability we expect among studies especially in the social sciences, moderator analyses provide important information about, for example, how the effects of an intervention vary in different contexts and with different participants. Computing the power of moderator analyses is critical when planning a meta-analysis, since examining potential reasons for variance among effect sizes is an important focus of a systematic review. This chapter will provide the computations for power for categorical models of effect sizes that are analogous to one-way ANOVA. These models can be computed under both the fixed or random effects assumptions, and I discuss the power calculations under both assumptions.

As we have seen in the prior chapters, we face challenges in posing important values for the parameters needed in the power computations. For categorical models, we will need to arrive at values for the differences among the means of groups of effect sizes, and for the degree of heterogeneity within groups. When using random effects models, we also need to provide plausible values for the variance component, τ^2. Both fixed effects and random effects models will be discussed as well as strategies for providing values for the power parameters.

T.D. Pigott, *Advances in Meta-Analysis*, Statistics for Social and Behavioral Sciences,
DOI 10.1007/978-1-4614-2278-5_6,

6.2 Categorical Models of Effect Size: Fixed Effects One-Way ANOVA Models

The simplest moderator models for effect size are analogues to one-way ANOVA models where the researcher is interested in comparing the mean effect sizes across groups of studies. The groups of studies are formed from a small number of levels of a categorical factor. For example, one moderator included in many syntheses is whether the study used random assignment to place individuals into experimental groups. Computation of categorical models with a single factor in meta-analysis proceeds in the same manner as one-way ANOVA. In meta-analysis, instead of computing the sums of squares, we compute between-group and within-group homogeneity tests. Below I outline the procedures for computing the statistics for the one-way fixed effects ANOVA model in meta-analysis, followed by a discussion of how to compute power.

6.2.1 Tests in a Fixed Effects One-Way ANOVA Model

As discussed in Chap. 2, we refer to our set of effect sizes using $T_i,\ i = 1, 2, \ldots, k$, where k is the total number of studies. Now let us assume that our k studies fall into p groups as defined by a moderator variable. For example, the moderator variable might be the grade level where the intervention takes place, say elementary or high school. We assume that p is a small number of categories. We can then designate m_i as the number of studies in the ith group, so that k is the total number of studies, $k = m_1 + m_2 + \ldots + m_p$.

Our interest in a one-way ANOVA effect size model is to examine whether the means for the groups are equal. We will also want to know if the effect sizes within each group are homogeneous. We will compute two values of the Q statistic, Q_B, the test of between-group homogeneity, and Q_W, the test of within-group homogeneity. The Q_B is analogous to the F-test between groups, and is an omnibus test of whether all group means are equal. The Q_W is an overall test of within-group homogeneity, and is equal to the sum of the homogeneity tests within each group. Thus, $Q_W = Q_{W_1} + Q_{W_2} + \ldots + Q_{Wp}$ where $Q_{W_i}, i = 1, 2, \ldots, p$ are the tests of homogeneity within each group. In addition, the two tests of homogeneity, like the F-tests in ANOVA, sum to the overall homogeneity test across all effect sizes, or, $Q_T = Q_B + Q_W$. The power computations for these two tests of homogeneity are given below.

6.2.2 Power of the Test of Between-Group Homogeneity, Q_B, in Fixed Effects Models

The first concern in the ANOVA model is whether the group mean effect sizes are equal. When there are p group means, the omnibus test of the null hypothesis that the group mean effect sizes are equal is given by

$$H_0 : \bar{\theta}_{1\bullet} = \bar{\theta}_{2\bullet} = \ldots = \bar{\theta}_{p\bullet} \tag{6.1}$$

To test this hypothesis, we compute the between-groups homogeneity test given by

$$Q_B = \sum_{i=1}^{p} w_{i\bullet} \left(\bar{T}_{i\bullet} - \bar{T}_{\bullet\bullet}\right)^2 \tag{6.2}$$

where $w_{i\bullet}$ is the sum of the weights in the ith group, $\bar{T}_{i\bullet}$ is the mean effect size in the ith group, and $\bar{T}_{\bullet\bullet}$ is the overall mean effect size. These quantities are more formally given as

$$\begin{aligned} w_{i\bullet} &= \sum_{j=1}^{m_i} w_{ij}, \text{ where } i = 1, \ldots, p, \text{ and } j = 1, \ldots, m_i \\ \bar{T}_{i\bullet} &= \frac{\sum_{j=1}^{m_i} w_{ij} T_{ij}}{\sum_{j=1}^{m_i} w_{ij}} \\ \bar{T}_{\bullet\bullet} &= \frac{\sum_{i=1}^{p} \sum_{j=1}^{m_i} w_{ij} T_{ij}}{\sum_{i=1}^{p} \sum_{j=1}^{m_i} w_{ij}} \end{aligned} \tag{6.3}$$

When the null hypothesis in (6.1) is true, that is when all the group means are equal, Q_B has the chi-square distribution with $(p - 1)$ degrees of freedom. When $Q_B > c_\alpha$ where c_α is the $100(1 - \alpha)$ percentile point of the chi-square distribution with $(p - 1)$ degrees of freedom, we reject the null hypothesis. When the null hypothesis is false, that is when at least one of the means differs from the other group means, Q_B has a non-central chi-square distribution with $(k - 1)$ degrees of freedom and non-centrality parameter λ_B given by

$$\lambda_B = \sum_{i=1}^{p} w_{i\bullet} \left(\bar{\theta}_{i\bullet} - \bar{\theta}_{\bullet\bullet}\right)^2. \tag{6.4}$$

The power of the test of Q_B is

$$1 - F(c_\alpha \,|\, p - 1;\ \lambda_B) \tag{6.5}$$

where $F(c_\alpha \,|p\text{-}1;\lambda_B)$ is the cumulative distribution of the non-central chi-square with $(p - 1)$ degrees of freedom and non-centrality parameter, λ_B.

Table 6.1 Steps for computing power in fixed effects one-way ANOVA model

1. Establish a critical value for statistical significance, c_α
2. Decide on the magnitude of the difference between the group means in the effect size metric used. For example, with standardized mean differences, decide on the number of standard deviations that are substantively important. For correlations, decide on the difference in correlations that is important
3. Assign values to the group means, $\theta_1,\ldots,\theta_p$ corresponding to the differences between group means in (2)
4. Estimate the number of studies within each group, $m_1,\ldots,m_p$
5. Compute w_{ij}, the common value of the weights for each effect size, given "typical" values for the within-study sample sizes

6.2.3 Choosing Parameters for the Power of Q_B in Fixed Effects Models

As in prior chapters, our challenge is to pose important values for the parameters in the power computations. Our main interest in the test of between-group homogeneity is the difference among the group effect sizes. Thus, we can pose a substantively important difference that we would like to test among the group means. If we, for example, want to know if the mean effect size for studies that use random assignment is at least 0.5 standard deviations smaller than the mean effect size for quasi-experiments, we could assume that the group mean effect size for the randomized studies is 0.0, and that for the quasi-experimental studies is 0.5. Thus, to test the between-group homogeneity, we can pose values for the mean effect sizes in each group, or at least the difference we would like to test between groups. We then need to suggest the number of effect sizes we will have in each group, and compute values for w_{ij}, the weights for the individual effect sizes. In prior chapters, we have made the simplifying assumption that all studies have the same sample sizes to obtain a common value of the w_{ij}, by assuming a "typical" within-study sample size. Table 6.1 summarizes the values needed to compute power for a fixed effects categorical moderator model.

6.2.4 Example: Power of the Test of Between-Group Homogeneity in Fixed Effects Models

One of the data sets used in the book is based on Sirin (2005), a meta-analysis of the correlations between measures of socio-economic status and academic achievement. The studies included in this meta-analysis used samples of students at several grade levels. Let us say that we want to make sure our meta-analysis will have enough power to detect differences in the mean effect size for studies that use students at three different grade levels: elementary (K-3), middle (4–8), and high school (9–12). We may be interested in whether the difference in the mean correlation between elementary and middle school is 0.1, and between elementary and high school is 0.5 (implying a difference of 0.4 for middle versus high school). We have three groups, $p = 3$, and we can assume that our mean effect sizes are $\bar{\theta}_1 = 0.0$, $\bar{\theta}_2 = 0.1$, and $\bar{\theta}_3 = 0.5$. In terms of Fisher's z-transformations, these means would be equal to $\bar{\theta}_{z1} = 0.0$, $\bar{\theta}_{z2} = 0.10$, and $\bar{\theta}_{z3} = 0.55$, respectively. Let us also assume that we have 5 studies per group, and our common within-group sample size is $n = 15$. To compute the non-centrality parameter, λ_B, we first need our common value of w_{ij}. For Fisher's z, $w_{ij} = 1/v_{ij} = n_{ij}$ -3 (v_{ij} for correlations is given in 2.15). Thus, we can compute

$$w_{i\bullet} = \sum_{j=1}^{5} w_{ij} = \sum_{j=1}^{5} (15 - 3) = \sum_{j=1}^{5} 12 = 60$$

as given in (6.3). The overall mean effect size is $\bar{\theta}_{i\bullet} = (0.0 + 0.10 + 0.55)/3 = 0.22$, since all effect sizes have the same weight, and we have equal numbers of studies within each group. (With different weights for each effect size and different numbers of effect sizes within groups, we would need to use a weighted mean as given in (6.3). We can compute the non-centrality parameter as

$$\begin{aligned}\lambda_B &= \sum_{i=1}^{3} w_{i\bullet} (\bar{\theta}_{i\bullet} - \bar{\theta}_{\bullet\bullet})^2 \\ &= 60(0.0 - 0.22)^2 + 60(0.10 - 0.22)^2 + 60(0.55 - 0.22)^2 \\ &= 60(-0.22)^2 + 60(-0.12)^2 + 60(0.33)^2 \\ &= 10.302\end{aligned}$$

The central chi-square distribution with $p - 1 = 3-1 = 2$ degrees of freedom has a critical value equal to 5.99 with $\alpha = 0.05$. The power of the omnibus test that $H_0 : \bar{\theta}_{z1} = \bar{\theta}_{z2} = \bar{\theta}_{z3}$ is given by $1 - F(5.99 \mid 2; 10.302) = 1 - 0.17 = 0.83$. The Appendix in Chap. 5 provides options for computing values of the non-central chi-square distribution.

6.2.5 *Power of the Test of Within-Group Homogeneity, Q_W, in Fixed Effects Models*

If a reviewer finds that the effect size means do differ between groups, then a second question centers on whether the effect sizes within those groups are homogeneous. The rationale for this question is similar to that for the overall test of homogeneity. While the mean effect sizes may differ among groups, we also need to know if the effect sizes within the group are estimating a common mean effect size. Hedges and Pigott (2004) refer to the test of within-group homogeneity as a test of the goodness of fit of the fixed effects model. In other words, the ANOVA model proposed fits the data well if the effect sizes within groups are homogeneous, i.e., that the effect sizes within each group estimate a common mean value. The overall test of within-group homogeneity is the omnibus test that the effect sizes within each group estimate a common mean. We can write the null hypothesis as

$$H_0 \;:\; \theta_{ij} = \bar{\theta}_{i\bullet}\,,\;\; i = 1, \ldots, p;\, j = 1, \ldots, m_i \tag{6.6}$$

where the alternative hypothesis is that at least one of the effect sizes in group i differs from the group mean, and there are m_i effect sizes within the ith group. The test for overall within-group homogeneity is

$$Q_W = \sum_{i=1}^{p}\sum_{j=1}^{m_i} w_{ij}\left(T_{ij} - \bar{T}_{i\bullet}\right)^2 = \sum_{i=1}^{p} Q_{W_i}. \tag{6.7}$$

Note that this sum can also be written as the sum of the within-group homogeneity statistics. Q_{W_i}. When every group in the model is homogeneous, Q_W has the central chi-square distribution with $(k - p)$ degrees of freedom, where k is the number of effect sizes and p is the number of groups. We reject the null hypothesis when, $Q_W > c_\alpha$ where c_α is the $100(1 - \alpha)$ percentile point of the chi-square distribution with $(k - p)$ degrees of freedom. Rejecting the null hypothesis indicates that at least one group is heterogeneous, i.e., that at least one effect size differs significantly from its group mean.

When the null hypothesis is false, the statistic Q_W has a non-central chi-square distribution with $(k - p)$ degrees of freedom, and non-centrality parameter λ_W given by

$$\lambda_W = \sum_{i=1}^{p}\sum_{j=1}^{m_i} w_{ij}\left(\theta_{ij} - \bar{\theta}_{i\bullet}\right)^2. \tag{6.8}$$

The power of the test of Q_W is given by

$$1 - F(c_\alpha \,|\, k - p;\; \lambda_W\,), \tag{6.9}$$

where $F(c_\alpha \mid k - p; \lambda_W)$ is the cumulative distribution of the non-central chi-square with $(k - p)$ degrees of freedom and noncentrality parameter λ_W, evaluated at c_α, the desired critical value of the central χ^2.

6.2.6 *Choosing Parameters for the Test of Q_W in Fixed Effects Models*

The difficulty in computing the power of within-group heterogeneity is in posing a substantively important value of heterogeneity. As we did in Chap. 5, we could decide on the amount of heterogeneity we would want to detect within groups based on the standard error of the group mean effect size. For example, we might want to determine the power to detect heterogeneity if one of the groups had effect sizes that differed from the mean by 3 standard errors of the mean. Below I illustrate an example of this strategy.

6.2.7 Example: Power of the Test of Within-Group Homogeneity in Fixed Effects Models

Let us continue with the example in Sect. 6.2.4. In that example, we have three groups, $p = 3$, corresponding to studies with elementary, middle and high school students, respectively. The mean Fisher z-transformations for each group are given by $\bar{\theta}_{z1} = 0.0$, $\bar{\theta}_{z2} = 0.10$, and $\bar{\theta}_{z3} = 0.55$, respectively. Let us also assume that we have 5 studies per group (for a total $k = 15$), and our common within-group sample size is $n = 15$. Thus, we have a common effect size variance of $v = 1/(n-3) = 1/(15-3) = 1/12$, with a common weight equal to $w = 1/(1/12) = 12$. As we did in Chap. 5, we can decide on how much variation we expect within the groups for our power computations. Given our within-group sample sizes, we can compute the variance (and standard error) of our mean effect sizes as

$$v_{i\bullet} = \frac{1}{\sum_{j=1}^{5} w} = \frac{1}{\sum_{j=1}^{5} 12} = \frac{1}{60} = 0.017$$
$$\sqrt{v_{i\bullet}} = \sqrt{0.017} = 0.13$$

Let us say that the effect sizes using elementary school samples and those from middle school samples differ by one standard deviation on average from their group mean, while the effect sizes using high school samples differ by four standard deviations on average from its group mean. We can compute the non-centrality parameter, λ_W, from (6.8) as

$$\begin{aligned}
\lambda_W &= \sum_{i=1}^{p}\sum_{j=1}^{m_i} w_{ij}\left(\theta_{ij} - \bar{\theta}_{i\bullet}\right)^2 \\
&= \sum_{j=1}^{5} 12(0.13)^2 + \sum_{j=1}^{5} 12(0.13)^2 + \sum_{j=1}^{5} 12(4 * 0.13)^2 \\
&= 5 * 12(0.017) + 5 * 12(0.017) + 5 * 12(0.52)^2 \\
&= 60(0.017) + 60(0.017) + 60(0.27) \\
&= 1.02 + 1.02 + 16.20 \\
&= 18.24
\end{aligned}$$

In the computation above, we replace the difference between the individual effect sizes and their group mean, $(\theta_{ij} - \bar{\theta}_{i\bullet})^2$, by the proposed number of standard deviations, $\sqrt{v_{i\bullet}}$ among the means. To compute power, we need the $c_{0.05}$ critical value of a central chi-square with $k - p = 15 - 3 = 12$ degrees of freedom, which is equal to 21.03. Equation 6.9 gives the power as $1 - F(c_\alpha \mid k - p; \lambda_W) = 1 - F(21.03|12; 18.24) = 1 - 0.18 = 0.82$. In this example, we have adequate power to find that at least one of our groups of effect sizes is not homogeneous. The Appendix in Chap. 5 provides the program code needed to compute values for the non-central chi-square distribution.

6.3 Categorical Models of Effect Size: Random Effects One-Way ANOVA Models

As discussed in Chap. 3, a random effects model for effect sizes assumes that each study's effect size is a random draw from a population of effect sizes. Thus, each effect size differs from the overall mean effect size due to the underlying variance of the population, designated as τ^2, and due to within-study sampling variance, v_i. Our goal in the random effects analysis is to first compute the variance component, τ^2, and then to use the variance component to compute the random effects weighted mean effect size.

When we are interested in estimating a random effects categorical model for moderators, we will conduct a similar analysis for each group defined by our categorical factor. As indicated in Chap. 3, we will make the assumption that there is a common variance component, τ^2, across studies, regardless of the study's value for the categorical factor. Given that we will assume the same variance component across studies, we will only focus on the test of between-group heterogeneity. The test of the significance of the variance component given in Chap. 5 for the random effects model would apply to the case where we are assuming a common variance component across studies.

As in Sect. 6.2.1, our k studies fall into p groups as defined by a moderator variable where, $k = m_1 + m_2 + \ldots + m_p$. The effect size in the jth study in the ith group is then designated by T^*_{ij} with variance of v^*_{ij}. The variance of each effect size contains two components, one due to sampling error in an individual study, denoted by v_{ij} and one due to the variance component, τ^2. The random effects variance for the effect size T^*_{ij} can be written as $v^*_{ij} = v_{ij} + \tau^2$.

6.3.1 *Power of Test of Between-Group Homogeneity in the Random Effects Model*

Similar to the fixed effects case, our test for between-group mean differences is an omnibus test. The null hypothesis for the test that the random effects group means differ can be written as

$$H_0 : \theta^*_{1\bullet} = \theta^*_{2\bullet} = ... = \theta^*_{p\bullet} \tag{6.10}$$

where the $\theta^*_{i\bullet}$ the random effects means for the $i = 1,\ldots, p$ groups. We test this hypothesis by computing the between-groups random effects homogeneity test given by

$$Q^*_B = \sum_{i=1}^{p} w^*_{i\bullet} \left(\bar{T}^*_{i\bullet} - \bar{T}^*_{\bullet\bullet}\right)^2 . \tag{6.11}$$

where $w^*_{i\bullet}$ is the sum of the weights in the ith group, $\bar{T}^*_{i\bullet}$ is the mean effect size in the ith group, and $\bar{T}^*_{\bullet\bullet}$ is the overall mean effect size, all in the random effects model. These quantities are more formally given as

$$w^*_{i\bullet} = \sum_{j=1}^{m_1} w^*_{ij}\ , \text{ where } i = 1, ..., p, \text{ and } j = 1, ..., m_i$$

$$\bar{T}^*_{i\bullet} = \frac{\sum_{j=1}^{m_i} w^*_{ij} T^*_{ij}}{\sum_{j=1}^{m_i} w^*_{ij}}$$

$$\bar{T}^*_{\bullet\bullet} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{m_i} w^*_{ij} T^*_{ij}}{\sum_{i=1}^{p}\sum_{j=1}^{m_i} w^*_{ij}}$$

When the null hypothesis is true, i.e., when all the means are equal, Q^*_B is distributed as a central chi-square distribution with $(p - 1)$ degrees of freedom. When the value of Q^*_B exceeds the critical value c_α which is the $100(1 - \alpha)$ percentile point of the central chi-square distribution with $(p - 1)$ degrees of freedom, we assume that at least one of the random effects group means is significantly different from the rest of the means. In the case where we reject the null hypothesis, Q^*_B has a non-central chi-square distribution with $(p - 1)$ degrees of freedom, and non-centrality parameter λ^*_B given by

$$\lambda^*_B = \sum_{i=1}^{p} w^*_{i\bullet} \left(\theta^*_{i\bullet} - \theta^*_{\bullet\bullet}\right)^2. \tag{6.12}$$

The power of the test for the between-group mean difference in a random effects model is given as

$$1 - F(c_\alpha \,|\, p - 1;\ \lambda^*_B) \tag{6.13}$$

where $F(c_\alpha \,|\, p - 1;\ \lambda^*_B)$ is the cumulative distribution function at c_α of the non-central chi-square with $(p - 1)$ degrees of freedom and non-centrality parameter λ^*_B.

Table 6.2 Steps for computing power in random effect categorical moderator model

1. Establish a critical value for statistical significance, c_α
2. Decide on the magnitude of the difference between the group means in the effect size metric used. For example, with standardized mean differences, decide on the number of standard deviations that are substantively important. For correlations, decide on the difference in correlations that is important
3. Assign values to the group means, $\theta_1,\ldots,\theta_p$ corresponding to the differences between group means in (2)
4. Estimate the number of studies within each group, $m_1,\ldots,m_p$
5. Compute v, the common value of the sampling variance for each effect size, given "typical" values for the within-study sample sizes
6. Compute τ^2 for levels of heterogeneity based on v. Large degree of heterogeneity is $\tau^2 = 3v$, moderate level is $\tau^2 = v$, and a small degree is $\tau^2 = (1/3)v$
7. Compute the values for the common weight, $w = 1/(v + \tau^2)$

6.3.2 *Choosing Parameters for the Test of Between-Group Homogeneity in Random Effects Models*

To compute the power of the test of between-group mean differences, we will need to decide on the size of the difference between the group means that is of substantive importance. In random effects models, we will also need to decide how much variation we have between studies, i.e., we need a value for τ^2. We can use the convention based on Higgins and Thompson (2002). Recall in Chap. 4 that we chose a value of τ^2 based on the value of v, the "typical" within-study sampling variance of effect sizes. A large degree of heterogeneity, and thus a large value of τ^2, was assumed to be $\tau^2 = 3v$, which corresponds to an I^2 value of .75. A moderate degree of heterogeneity is $\tau^2 = v$, corresponding to an I^2 value of .5, and a small degree of heterogeneity is $\tau^2 = (1/3)v$, corresponding to an I^2 value of .25. We can amend the steps in Table 6.1 for the power for differences of random effects means as seen in Table 6.2.

Thus, to compute the power of the test of the between-group random effects means, we need to pose values for the number of studies within each group, the within-study sample size, the degree of heterogeneity we expect, and the magnitude of the differences between the means that is substantively important.

6.3.3 *Example: Power of the Test of Between-Group Homogeneity in Random Effects Models*

Let us return to the example in Sect. 6.2.4. Recall that we have three groups, $p = 3$, and we can assume that our mean effect sizes are $\bar{\theta}_1 = 0.0$, $\bar{\theta}_2 = 0.1$, and $\bar{\theta}_3 = 0.5$. In terms of Fisher's z-transformations, these means would be equal to $\bar{\theta}_{z1} = 0.0$, $\bar{\theta}_{z2} = 0.10$, and $\bar{\theta}_{z3} = 0.55$, respectively. Let us also assume that we have 5 studies per group (with $k = 15$), and our common within-group sample size is $n = 15$. To compute the non-centrality parameter λ_B^*, we first need to posit a value of the common value of the variance component, τ^2, across studies, based on our common value of v. For Fisher's z-transformation, the common value of the within-group sampling variance in this example is $v = 1/(n - 3) = 1/12 = 0.083$. For a large degree of heterogeneity, the variance component would equal $\tau^2 = 3(0.083) = 0.25$, a moderate degree of heterogeneity, $\tau^2 = 0.083$, and a small degree of heterogeneity, $\tau^2 = (1/3)0.083 = 0.028$. Thus, our value for the random effects variance, v^*, is given as $v^* = 0.083 + \tau^2$. Assuming a small degree of heterogeneity, the value of the weight for the random effects means, $w_{i\bullet}^*$, is given as

$$w_{i\bullet}^* = \frac{1}{\sum_{i=1}^{5} \frac{1}{v + \tau^2}} = \frac{1}{\sum_{i=1}^{5} \frac{1}{0.083 + 0.028}} = \frac{1}{5 * 9.01} = 0.022$$

Given our value of $w_{i\bullet}^*$, we can compute the non-centrality parameter λ_B^* as

$$\begin{aligned}\lambda_B^* &= \sum_{i=1}^{3} w_{i\bullet}^* \left(\theta_i^* - \theta_\bullet^*\right)^2 \\ &= 0.022(0 - 0.22)^2 + 0.022(0.1 - 0.22)^2 + 0.022(0.55 - 0.22)^2 \\ &= 0.022(0.048) + 0.022(0.014) + 0.022(0.11) \\ &= 0.004\end{aligned}$$

The power of the test of the between-group differences is given as $1 - F \times (c_\alpha \mid p - 1;\ \lambda_B^*)$ in (6.13). With $p - 1 = 2$ degrees of freedom, the $c_{0.05}$ critical value of a central chi-square distribution is 5.99. Thus, the power in this example with a small degree of heterogeneity is $1 - F(5.99 \mid 2;\ 0.004) = 1 - 0.95 = 0.05$.

We can also compute the power with a large degree of heterogeneity as a comparison. With a large degree of heterogeneity, we have a value for the random effects variance, v^*, as $v^* = 0.083 + 0.25 = 0.33$. Assuming a large degree of heterogeneity, the value of the weight for the random effects means, $w_{i\bullet}^*$, is given as

$$w_{i\bullet}^* = \frac{1}{\sum_{i=1}^{5} \frac{1}{v + \tau^2}} = \frac{1}{\sum_{i=1}^{5} \frac{1}{0.33}} = \frac{1}{5 * 3.033} = \frac{1}{15.01} = 0.067$$

Given our value of $w_{i\bullet}^*$, we can compute the non-centrality parameter λ_B^* as

$$\begin{aligned}\lambda_B^* &= \sum_{i=1}^{3} w_{i\bullet}^* \left(\theta_i^* - \theta_\bullet^*\right)^2 \\ &= 0.067(0 - 0.22)^2 + 0.067(0.1 - 0.22)^2 + 0.067(0.55 - 0.22)^2 \\ &= 0.067(0.048) + 0.067(0.014) + 0.067(0.11) \\ &= 0.012\end{aligned}$$

With $p - 1 = 2$ degrees of freedom, the $c_{0.05}$ critical value of a central chi-square distribution is 5.99. Thus, the power in this example with a large degree of heterogeneity is $1 - F(5.99 \mid 2;\ 0.012) = 1 - 0.95 = 0.05$. In the random effects model, we have little power to detect a difference among the mean effect sizes.

6.4 Linear Models of Effect Size (Meta-regression)

When a reviewer wishes to test a model with a number of moderators, including continuous predictors, the relatively simple ANOVA models discussed above have limited utility. Large-scale reviews often include multiple moderators that reviewers would like to test simultaneously in a linear model, commonly called a meta-regression. The use of meta-regression models instead of a series of one-way ANOVA models helps reviewers avoid conducting too many statistical tests. However, computing the power of meta-regression models a priori is problematic. As seen in Hedges and Pigott (2004), we need to know the exact values of the moderators for each study in order to compute power for tests in meta-regression. These values cannot be guessed a priori, and thus, I will not present the details of these tests here.

References

Hedges, L.V., and T.D. Pigott. 2004. The power of statistical tests for moderators in meta-analysis. *Psychological Methods* 9(4): 426–445.

Higgins, J.P.T., and S.G. Thompson. 2002. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 21: 1539–1558.

Sirin, S.R. 2005. Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research* 75(3): 417–453. doi:10.3102/00346543075003417.