7.1.2 Definitions of proportions of explained variance in two-level models

In multiple linear regression, the customary R^2 parameter can be introduced in several ways: e.g., as the maximal squared correlation coefficient between the dependent variable and some linear combination of the predictor variables, or as the proportional reduction in the residual variance parameter due to the joint predictor variables. A very appealing principle to define measures of modeled (or explained) variation is the principle of ***proportional reduction of prediction error***. This is one of the definitions of R^2 in multiple linear regression, and can be described as follows. A population of values is given for the explanatory and the dependent variables, $(X_{1i}, ..., X_{qi}, Y_i)$, with a known joint probability distribution; β is the value for the vector v for which the expected squared error

$$\mathcal{E}(Y_i - \sum_{h=0}^{q} v_h X_{hi})^2$$

is minimal. (This is the definition of the ordinary least squares ('OLS') estimation criterion.) (In this equation, v_0 is defined as the intercept and $X_{0ij} = 1$ for all i, j.) If, for a certain case i, the values of $X_{1i}, ..., X_{qi}$ are unknown, then the best predictor for Y_i is its expectation $\mathcal{E}(Y)$, with mean squared prediction error $\text{var}(Y_i)$; if the values $X_{1i}, ..., X_{qi}$ are given, then the linear predictor of Y_i with minimum squared error is the regression value $\sum_h \beta_h X_{hi}$. The difference between the observed value Y_i and the predicted value $\sum_h \beta_h X_{hi}$ is the prediction error. Accordingly, the mean squared prediction error is defined as

$$\mathcal{E}(Y_i - \sum_{h} \beta_h X_{hi})^2 .$$

The proportional reduction of the mean squared error of prediction is the same as the proportional reduction in the unexplained variance, due to the use of the variables X_1 to X_q. In a formula, it can be expressed by

$$R^2 = \frac{\text{var}(Y_i) - \text{var}(Y_i - \sum_h \beta_h X_{hi})}{\text{var}(Y_i)}$$

$$= 1 - \frac{\text{var}(Y_i - \sum_h \beta_h X_{hi})}{\text{var}(Y_i)} ;$$

this formula expresses one of the equivalent ways to define R^2.

The same principle can be used to define 'explained proportion of variance' in the hierarchical linear model. For this model, however, there are several options with respect to what one wishes to predict. Let us consider a two-level model with dependent variable Y. In such a model, one can choose between predicting an individual value Y_{ij} at the lowest level, or a group mean $\overline{Y}_{.j}$. On the basis of this distinction, two concepts of explained proportion of variance in a two-level model can be defined. The first, and most important, is the ***proportional reduction of error for predicting an individual outcome***. The second is the ***proportional reduction of error for predicting a group mean***. To elaborate these concepts more specifically, first consider a two-level random effects model with a random intercept and some predictor variables with fixed effects but no other random effects:

$$Y_{ij} = \gamma_0 + \sum_{h=1}^{q} \gamma_h X_{hij} + U_{0j} + R_{ij} \ . \tag{7.1}$$

Since we wish to discuss the definition of 'explained proportion of variance' as a population parameter, we assume temporarily that the vector γ of regression coefficients is known.

Level one

For the level-one explained proportion of variance, we consider the prediction of Y_{ij} for a randomly drawn level-one unit i within a randomly drawn level-two unit j. If the values of the predictors X_{ij} are unknown, then the best predictor for Y_{ij} is its expectation; the associated mean squared prediction error is $\text{var}(Y_{ij})$. If the value of the predictor vector X_{ij} for the given unit is known, then the best linear predictor for Y_{ij} is the regression value $\sum_{h=0}^{q} \gamma_h X_{hij}$ (where X_{h0j} is defined as 1 for all h, j.) The associated mean squared prediction error is

$$\text{var}(Y_{ij} - \sum_h \gamma_h X_{hij}) = \sigma^2 + \tau_0^2 \ .$$

The level-one explained proportion of variance is defined as the proportional reduction in mean squared prediction error:

$$R_1^2 = 1 - \frac{\text{var}(Y_{ij} - \sum_h \gamma_h X_{hij})}{\text{var}(Y_{ij})} \ . \tag{7.2}$$

Now let us proceed from the population to the data. The most straightforward way to estimate R_1^2 is to consider $\hat{\sigma}^2 + \hat{\tau}_0^2$ for the empty model,

$$Y_{ij} = \gamma_0 + U_{0j} + E_{ij} \ , \tag{7.3}$$

as well as for the fitted model (7.1), and compute 1 minus the ratio of these values. In other words, R_1^2 is just the proportional reduction in the value of $\hat{\sigma}^2 + \hat{\tau}_0^2$ due to including the X-variables in the model. For a sequence of nested models, the contributions to the estimated value of (7.2) due to adding new predictors can be considered to be the contribution of these predictors to the explained variance at level one.

To illustrate this, we once again use the data from the first (balanced) example, and estimate the proportional reduction of prediction error for a model where within-group and between-groups regression coefficients may be different.

Table 7.2 Estimating the level-one explained variance (balanced data).

	$\hat{\sigma}^2$	$\hat{\tau}_0^2$
A. $Y_{ij} = \beta_0 + U_{0j} + E_{ij}$	8.694	2.271
D. $Y_{ij} = \beta_0 + \beta_1(X_{ij} - \overline{X}_{.j}) + \beta_2 \overline{X}_{.j} + U_{0j} + E_{ij}$	6.973	0.991

From Table 7.2 we see that $\hat{\sigma}^2 + \hat{\tau}_0^2$ for model (A) amounts to 10.965, and for model (D) to 7.964. R_1^2 is thus estimated to be $1 - (7.964/10.965) = 0.274$.

Level two

The level-two explained proportion of variance can be defined as the proportional reduction in mean squared prediction error, for the prediction of $\overline{Y}_{.j}$ for a randomly drawn level-two unit j. If the values of the predictors X_{hij} for the level-one units i within level-two unit j are completely unknown, then the best predictor for $\overline{Y}_{.j}$ is its expectation; the associated mean squared prediction error is $\text{var}(\overline{Y}_{.j})$. If the values of all predictors X_{hij} for all i in this particular group j are known, then the best linear predictor for $\overline{Y}_{.j}$ is the regression value $\sum_h \gamma_h \overline{X}_{h.j}$; the associated mean square prediction error is

$$\text{var}(\overline{Y}_{.j} - \sum_h \gamma_h \overline{X}_{h.j}) = \frac{\sigma^2}{n} + \tau_0^2 \, ,$$

where n is the number of level-one units on which the average is based. The level-two explained proportion of variance is now defined as the proportional reduction in mean squared prediction error for $\overline{Y}_{.j}$:

$$R_2^2 = 1 - \frac{\text{var}(\overline{Y}_{.j} - \sum_h \gamma_h \overline{X}_{h.j})}{\text{var}(\overline{Y}_{.j})} \, . \tag{7.4}$$

To estimate the level-two explained proportion of variance, we follow a similar approach as above: R_2^2 is estimated as the proportional reduction in the value of $\hat{\sigma}^2/n + \hat{\tau}_0^2$, where n is a representative value for the group size.

In the example given earlier, let us use for n a usual group size of $n = 30$. For model (a) the value of $\hat{\sigma}^2/n + \hat{\tau}_0^2$ is $8.694/30 + 2.271 = 2.561$, whereas for model (b) this amounts to $6.973/30 + 0.991 = 1.223$. R_2^2 is thus estimated at $1 - (1.223/2.561) = 0.52$.

It is natural that the mean squared error for predicting a group mean should depend on the group size. Often one can use for n a value which is deemed *a priori* to be 'representative'. For example, if a normal class size is considered to be 30, and even if because of missing data the values of n_j in the data set are on average less than 30, it is advisable to use the representative value $n = 30$. In the case of varying group sizes, if it is unclear how a representative group size should be chosen, one possibility is to use the harmonic mean, defined by $N/\{\sum_j (1/n_j)\}$.

It is quite common that the data within the groups are based on a sample and not the entire groups in the population are observed, so that the group sizes in the data set do not reflect the group sizes in the population. In such a case, since it is more relevant to predict population group averages than sample group averages, it is advisable to let n reflect the group sizes in the population rather than the sample group sizes. If group sizes are very large, e.g., when the level-two units are defined by municipalities or other regions and the level-one units by their inhabitants, this means that, practically speaking, R_2^2 is the proportional reduction in the intercept variance.