Chapter 8
Ecometrics: Using MLA to Construct Contextual Variables from Individual Data

Abstract Multilevel analysis can be used to construct characteristics of higher level units. This is done on the basis of systematic observations by several observers or of perceptions of respondents who describe, for example, their neighbourhood. By using MLA, we solve a number of problems associated with simple aggregation of data from the individual level to the higher level. The chapter starts by identifying these problems and then works step by step towards more elaborate models to measure latent characteristics of higher level units. Latent variable analysis in MLA is also called ‘ecometrics’, a new term for methods to measure characteristics of ecological units on the basis of multiple observations or responses.

Keywords Multilevel analysis · Ecometric analysis · Aggregation · Contextual variables · Latent variable analysis · Reliability · Patient safety

In Chap. 3, in the section on hypotheses, we mentioned different types of variables that can characterise higher level units. They could be either directly measured at the higher level or aggregated from characteristics of lower level units (such as individuals). Examples of variables that are measured at the higher level are the specialty of a hospital ward, or the total surface of green areas in a neighbourhood. Other characteristics can be measured by aggregation from individual-level characteristics, such as the average age or income of the inhabitants of a neighbourhood. Sometimes we are not dealing with separate variables but with composite scores, based on several variables or responses. Examples are questionnaires that ask people about neighbourhood contacts and that could be combined into a social capital measure or questionnaires to doctors and nurses in hospitals that ask about dealing with issues of safety which could be combined into a measure of patient safety culture.

We could just aggregate the individual variables or responses to items in questionnaires to the appropriate higher level. However, there are problems associated with doing this, and we can overcome these problems by applying MLA to construct our higher level variables. We then use MLA to estimate the higher level effect—e.g. a neighbourhood effect or a hospital effect—net of the individual variation at other levels. When applied to composite scores, this approach is now
known as ecometrics or latent variable analysis in MLA (Raudenbush 2003; Raudenbush and Sampson 1999).

The use of ecometrics in public health and health services research is becoming more frequent. It is therefore important to pay attention to this application of multilevel research. Following its name—ecometrics as the measurement of ecological characteristics—it is currently mainly used to construct variables to characterise neighbourhoods, such as social capital (Mohnen et al. 2011; Nyqvist et al. 2014; Prins et al. 2012). It is much less frequently used for other ecologies of humans, such as work places (Oksanen et al. 2013), schools (Gilreath et al. 2012) or healthcare institutions (van Schoten et al. 2014).

This chapter is based around an example of patient safety culture in hospital departments. We will discuss the multilevel model and its interpretation in ecometric analysis, and we will compare ecometrics with traditional methods. We end the chapter with a discussion of ecometric properties (comparable to psychometric properties), such as reliability.

Problems with Simple Aggregation

Simple aggregation of individual variables to a higher level unit is not wrong, but there can be a couple of problems with doing this. First, our individual-level variables are often derived from a sample of individuals. When we group these individuals into higher level units, the sample sizes may vary. Some higher level units might have many more individual observations than others, and this is especially likely to be the case when the study was not originally designed as a multilevel study. If we simply aggregate an individual-level variable using such data, our aggregated variable is then based on different numbers of observations. However, if we aggregate the data and use this aggregated variable in our model, all aggregated observations are treated in the same manner, irrespective of whether they were based on say 100 individual observations or just ten. The solution to this has already been presented in Chap. 3; the multilevel approach to estimating higher level effects or residuals takes into account whether the number of observations differs between higher level units. The estimated values for units with few observations are shrunk towards the overall mean.

The second problem is particularly important when the individual variables contain a subjective element. Examples are people’s responses to questions about their neighbourhood (‘how safe is your neighbourhood?’) or the hospital in which they were treated (‘were you treated with respect during your visit to this hospital?’). The responses to such questions supposedly indicate a characteristic of the higher level unit—the neighbourhood or the hospital—but part of the response is determined by individual differences in how people perceive their neighbourhood or hospital. The response may also be determined in part by incidental circumstances, such as what they read in the newspaper that morning. What we are really interested in is the common component in all responses about the same unit, net of the individual component. We can obtain this by partitioning the variance in individual responses into that attributable to higher level units and that attributable to
individuals. We do this using MLA as detailed in Chap. 6. Related to this is the argument that using ecometrics the effect of same source bias is reduced (as noted by de Jong et al. 2011). Same source bias originates from the fact that in survey research often both the independent variables and the dependent variables are asked from the same respondent in the same questionnaire.

A third problem is that the sample of individuals in higher level units may differ between these units as a result of selective non-response. Selective non-response might lead to there being more elderly respondents or more highly educated respondents in some hospitals or neighbourhoods. If these characteristics are related to the variables that we want to aggregate, simple aggregation would lead to the creation of a biased contextual variable. This might be the case if elderly people have a higher level of response in some neighbourhoods in a survey looking at neighbourhood safety, since we know that elderly people perceive their neighbourhood as less safe than younger people. Again the solution is to use MLA to control for the effects of differential neighbourhood composition. The idea that the estimation of contextual effects should take into account relevant compositional factors was discussed in Chap. 7.

Finally, there is a specific problem when the responses at the individual level form a scale where several questions or items together are supposed to measure a characteristic of the higher level unit. For example, rather than simply asking people who were treated in a hospital whether they were treated with respect, we could design some questions that when combined measure the unobserved variable ‘respectful treatment’. We could construct the scale at the individual level in a single-level model. However, if we did this we would lose information about the fact that the items are not only nested within the individuals that complete the questionnaire but also within the higher level units that we want to characterise. The solution here is to analyse the data using a multiple response model with items at the lowest level, nested within individuals and higher level units (latent variable analysis). We have described such data structures in Chap. 4.

This means that we can use MLA to construct a contextual variable, because we want to say something about higher level units, based on individual-level observations. We can use either a single variable (using a two-level model) or a number of variables collected at the individual level (using a three-level model) and combine these into a higher level variable. The first step is to construct a multilevel model including the variable(s) that we want to use to describe our higher level units as dependent variable(s). In the second step we then take the higher level residual (the higher level effect) and use this as an independent variable in a subsequent multilevel analysis, relating this to a dependent variable (such as self-rated health).

**Single Variables**

We can use MLA both when we want to construct a characteristic of a higher level unit based on a single individual variable and when we wish to combine information from several related individual variables. We begin by considering the single-
variable case. From the problems that we discussed in the previous section, we can see that it is possible to make another distinction. Some individual variables indicate objective information, such as household income, whilst others indicate perceptions or evaluations of characteristics of the higher level units, such as perceived safety or the extent to which treatment was respectful.

When using objective information, such as household income per neighbourhood, we may have access to population data from municipalities or national statistical sources. However, this information is not always available, especially not when we have good reasons to deviate from a standard administrative definition of neighbourhoods (as discussed in Chap. 2). In such a case, we may have to use sample data. As discussed above, the sample size might differ between neighbourhoods, and we would have more confidence in an aggregated variable which is based on more information than one based on fewer observations. The estimated neighbourhood-level income from a multilevel analysis will be closer to the overall mean when the sample size (and thus the number of observations) in that neighbourhood is smaller.

When analysing individual perceptions or evaluations, multiple questions are often used. However, for research into patient experiences with healthcare providers, single questions are also often used. These could be used to compare healthcare providers, or they could be used as independent variables at the provider level in the analysis of an individual-level dependent variable. Research with the so-called consumer quality index on GP care showed very strong clustering of the single item from the questionnaire about privacy at the reception desk (‘Can people in the waiting room hear what is being discussed at the reception desk?’; Meuwissen and De Bakker 2008). They found the intra-class correlation to be 0.29; nearly a third of the variation in responses was associated with the level of the GP practice. Although there is still a lot of variation between the individual patients in how they answered this question, the answers clearly say something about the contexts. The GP practice residuals could subsequently be used in a separate analysis to predict individual satisfaction with GP care.

**Composite Variables: The Traditional Method**

As we said, usually perceptions will be based on composite variables. We will discuss both the ‘traditional approach’ and the ecometric approach. The example we will use is based on data from a study on patient safety culture in hospital wards (Smits 2009). Patient safety culture could be seen as an independent variable at the level of the hospital ward to predict adverse events among patients. Patient safety culture is measured by several items in a questionnaire for hospital personnel.

Other examples for comparable data structures and approaches in analysing data could be a questionnaire about social capital for inhabitants of neighbourhoods (Mohnen et al. 2011) or observations that are made concerning the disorderliness of streets within neighbourhoods, including items such as people drinking outside, graffiti and broken windows (Raudenbush and Sampson 1999).
The items from the patient safety questionnaire that we will be using relate to ‘feedback and learning from error’. The items are:

- We are informed about errors that happen in this unit.
- We are given feedback about changes put into place based on event reports.
- In this unit, we discuss ways to prevent errors from happening again.
- Mistakes have led to positive changes here.
- After we make changes to improve patient safety, we evaluate their effectiveness.
- We are actively doing things to improve patient safety.

The traditional approach would be to perform a psychometric analysis and combine the items into a scale, all within a single-level model. This would involve undertaking an analysis of the characteristics of the items, their inter-correlations, item total correlation and so on. We would calculate Cronbach’s alpha as a measure of the reliability of the scale. Finally we would actually calculate the scale and aggregate the individual scale values to ward level. This would be our independent variable for subsequent multilevel analysis of individual-level outcomes such as the occurrence of adverse events.

We will not go deeply into the psychometric properties of this scale (for more details, see Smits et al. 2008). The scale average in an analysis of 583 employees in four hospitals was 3.34; Cronbach’s alpha was 0.78; and the correlation of the scale with a grading of patient safety from excellent to failing was 0.40.

After aggregating the individual scale values to the level of hospital wards, we can rank the wards in terms of their patient safety culture. Some hospitals can be seen to have a more favourable patient safety culture than others.

**Composite Variables: A Simple Multilevel Model**

In this section we will take the analysis one step further but do not stray too far from the traditional approach: we will analyse the individual scale values in a multilevel model. In the following section, we will introduce the ecometric approach in which we treat the separate items that form the scale as the lowest level, with these responses nested within the individual.

In our example, we are not interested in individual variation in perceived safety culture, but only in the common variance at the ward (or hospital) level. When we theorise about patient safety culture, our hypothesis about variation would be that if something approximating patient safety culture exists, we should find significant clustering at the level of hospitals or wards since this is almost certain to vary between units. Culture as a concept implies a shared definition of the situation. And if we want to characterise the wards or hospitals, then we need to remove the individual variation.

In this example, the sample size varies between wards. The average sample was 22, but there was a minimum of only seven questionnaires and a maximum of 53. In this case we estimate a three-level model with the data structure shown in Fig. 8.1.
We use a three-level model because the hospital wards are themselves nested within hospitals, and the hospital itself may affect the safety culture within its wards. When analysing social capital in neighbourhoods, we could work with a two-level model involving neighbourhoods at the highest level and scale values for a social capital scale as the dependent variable at the level of the individual. The social capital scale would have been constructed from individuals who answered questions about their neighbourhood. In our example of patient safety culture, the minimum number of observations on a ward is relatively small (only seven observations in one ward). The question then is: how confident would we be about an estimate of the population parameter ‘patient safety culture’ derived from the ward that had this small sample size? How confident can we be of any difference from the population mean given the small sample size? This is the rationale for using an estimator that shrinks the estimate for this ward a bit closer to the overall mean.

The multilevel model we estimate is described in Eq. (8.1).

\[
y_{ijk} = \beta_0 + v_{0k} + u_{0jk} + e_{0ijk}
\]

\[
v_{0k} \sim N\left(0, \sigma^2_{v0}\right)
\]

\[
u_{0jk} \sim N\left(0, \sigma^2_{u0}\right)
\]

\[
e_{0ijk} \sim N\left(0, \sigma^2_{e0}\right)
\]

(8.1)

Here \(y_{ijk}\) is the response ‘feedback and learning from error’ for respondent (nurse or doctor) \(i\) in ward \(j\) and hospital \(k\), measured on a scale from 0 to 5 (the answers to the six items forming the scale were given values 0 through 5, then added and divided by the number of items). The random intercept model described partitions the variance between the individual, ward and hospital levels. The resulting estimates from this model are shown in Table 8.1.

The constant gives the scale average. In addition we have estimated three variance components: at hospital, ward and individual level. It is important to note...
that there is a significant variation at ward level, which is what we would expect given that the item in question is measuring an aspect of patient safety culture.

An advantage of the multilevel analysis over and above the traditional approach of simply aggregating the individual scale values is that we can adjust for compositional effects by including individual independent variables that may have an impact on individual responses but not necessarily hospital culture. The adjusted model is described in Eq. (8.2).

\[
y_{ijk} = \beta_0 + \beta_1 x_{1ijk} + \beta_2 x_{2ijk} + \beta_3 x_{3ijk} + \nu_{0k} + u_{0jk} + e_{0ijk}
\]

\[
v_{0k} \sim N(0, \sigma^2_{v0})
\]

\[
u_{0jk} \sim N(0, \sigma^2_{u0})
\]

\[
e_{0ijk} \sim N(0, \sigma^2_{e0})
\]

(8.2)

In addition to the constant, this model includes three variables enabling adjustment for the number of years an individual has worked in the ward, the number of hours he or she works per week and whether the respondent is a physician or a nurse. The estimates from this model are detailed in Table 8.2.

Although the adjusted model fits the data better than the empty model, in this dataset the variance components of the adjusted model are nearly the same as for the null model. This is of course not necessarily the case. Apparently employee characteristics, in this case the composition of our samples according to length of employment on this ward, the number of hours they work and whether they are nurses do not vary much enough between wards or hospitals to influence the results. In other datasets, there might be bigger effects of composition; it is not possible to know whether these will exist before undertaking the analysis.

As we have two higher levels, wards and hospitals, we can also calculate two variance partition coefficients for each model (see Table 8.3).

| Table 8.2 Estimates from a multilevel analysis of the individual scale values for the scale ‘feedback and learning from error’; empty model and adjusted model (simple multilevel model) |
|---|---|---|---|
| Parameter | Empty model | Adjusted model |
| | Coefficient (SD) | Coefficient (SD) |
| Fixed part | | |
| Constant | 3.372 (0.045) | 3.373 (0.046) |
| Years in this ward | 0.028 (0.010) | |
| Hours per week | −0.050 (0.025) | |
| Nurse (reference: physician) | 0.011 (0.004) | |
| Random part | | |
| Hospital-level variance | 0.021 (0.013) | 0.023 (0.013) |
| Ward-level variance | 0.052 (0.012) | 0.051 (0.011) |
| Individual-level variance | 0.293 (0.010) | 0.290 (0.010) |
| −2+log likelihood | 3080.7 | 3059.5 |
Twenty per cent of the total variation in this scale is above the level of the individual; this is a relatively strong clustering effect. The scale apparently measures something at the level of the contexts, as should be the case given that we intended to measure culture.

This all looks fine, but as we mentioned in the introduction to this chapter, there is still a problem with this approach. The items are nested within individuals, wards and hospitals. We should take ward-level correlations between items into account and because we want the scale to say something about wards, we would also like to know how reliable a measure it is of ‘feedback and learning from error’ at the ward level. For this reason, in the next section we move beyond the traditional approach or the simple multilevel model based on the individual scale values to a full ecometric approach.

**Ecometric Approach**

In the ecometric approach, we will estimate a more complicated model: a multiple response model with items at the lowest level, nested in individuals and higher level units. The data structure then looks as in Fig. 8.2.

The term ‘ecometrics’ was coined by Raudenbush. He describes ecometrics as a statistical method to evaluate the validity and reliability of imperfect measures of contextual properties (Raudenbush 2003). The term is analogous to psychometrics, the difference being that it does not aim to measure latent psychological characteristics of individuals but latent characteristics of ecological units. The data used in ecometrics are multiple observations on an ecological unit, made by trained observers or individuals (e.g. respondents in a survey) who are able to give information about characteristics of these units. As in psychometrics, the aim is to combine these multiple observations into a single scale or latent variable and to analyse the characteristics of the scale such as its reliability and validity. Mujahid et al. (2007) have illustrated the ecometric approach by using survey data to construct a number of scales that are relevant for health and health behaviour.
They include respondents’ perceptions of the walking environment, the availability of healthy food and social cohesion. An example based on observers’ evaluations of neighbourhood environment is given by Gauvin et al. (2005).

The basis of an ecometric analysis is a three-level model: the items or observations are the lowest level, nested within observers or individual respondents, and these nested again in higher level units. (In our example, the higher level units of interest are the hospital wards, but these are in turn nested within the hospitals to take the particular data structure and the possibility that hospitals influence the culture on wards into account.) The model is shown algebraically in Eq. (8.3).

\[
y_{ijkl} = \beta_0 + \beta_2 (x_{2ijkl} - \frac{1}{6}) + \beta_3 (x_{3ijkl} - \frac{1}{6}) + \beta_4 (x_{4ijkl} - \frac{1}{6}) + \beta_5 (x_{5ijkl} - \frac{1}{6}) + \beta_6 (x_{6ijkl} - \frac{1}{6}) + f_{0l} + v_{0kl} + u_{0jkl} + e_{1ijkl}x_{1ijkl} + e_{2ijkl}x_{2ijkl} + e_{3ijkl}x_{3ijkl} + e_{4ijkl}x_{4ijkl} + e_{5ijkl}x_{5ijkl} + e_{6ijkl}x_{6ijkl}
\]

\[
f_{0l} \sim N\left(0, \sigma_{f0}^2\right)
\]

\[
v_{0kl} \sim N\left(0, \sigma_{v0}^2\right)
\]

\[
u_{0jkl} \sim N\left(0, \sigma_{u0}^2\right)
\]

\[
e_{mijkl} \sim N\left(0, \sigma_{em}^2\right), \quad m = 1 \ldots 6
\]  

(8.3)

In this formulation, \(\beta_0\) is the scale average and \(\beta_2 \ldots \beta_6\) are the deviance scores for items 2–6, respectively. With six items, and therefore six responses per individual, we include only five dummy variables \(x_{2ijkl} \ldots x_{6ijkl}\) coded 1 if the response relates to that item and 0 otherwise. We subtract the reciprocal of the number of items—in this case \(\frac{1}{6}\)—from each of the dummy variables to ensure that we obtain the deviance scores. This amounts to scoring each variable equal to \(\frac{5}{6}\) if the response relates to that item and \(-\frac{1}{6}\) otherwise, meaning that each of these variables has a mean of 0. By doing so, the value obtained for \(\beta_0\) is comparable to the scale value of the original single-level model (between 0 and 5). Otherwise the scale value would be the average of the item that was left out. The response \(y_{ijkl}\) refers to item \(i\) for respondent \(j\) in ward \(k\) and hospital \(l\). There are variances associated with the hospital, ward and individual levels, whilst each of the six items is assumed to be independently normally distributed with its own variance \(\sigma_{e1}^2 \ldots \sigma_{e6}^2\).
**Application of the Ecometric Approach**

Applying this model to our data gives the results presented in Table 8.4.

We can start by pointing out the differences between the simple multilevel model shown in Table 8.1 and the model shown in Table 8.4. Apart from the constant, which is the scale average, we now have fixed effects for the different items. We did not have that for the simple model because the scale was first constructed at the individual level, and the scale value for each individual was taken as the dependent variable. There is also a difference in the random part. In addition to the individual-, ward- and hospital-level variances, we now also estimate a variance for each item.

When it comes to interpreting the model shown in Table 8.4, we first note that the average scale value obtained is almost identical whether we use the ecometric or the simple approach. This is because we use the item weights for the fixed effects as explained above. This is only necessary when we want an easily interpretable and comparable scale average.

The other fixed effects give the weights of the scale items. The average score of item 2, for example, is $3.375 + \left(\frac{0.394}{5/6}\right) = 3.703$. The fixed effects indicate how frequently individuals tend to agree with a statement, something called item difficulty in psychometric analysis. Item 3 was the item for which agreement was most common: ‘In this unit, we discuss ways to prevent errors from happening again’. Agreement was least common for item 6: ‘We are actively doing things to improve patient safety’. It appears to be easier to agree with item 3 than with item 6.

Then we move on to the variance components or the random part of the model. Each item has its own variance, indicating the measurement error. Item 1 has the biggest variance: ‘We are informed about errors that happen in this unit’. The ecometric analysis has separated the individual variance in the traditional approach

| Parameter                | Coefficient (SD)          |
|--------------------------|---------------------------|
| Fixed part               |                           |
| Constant                 | 3.375 (0.044)             |
| Item 1 (reference)       |                           |
| Item 2                   | 0.394 (0.026)             |
| Item 3                   | 0.643 (0.024)             |
| Item 4                   | 0.328 (0.025)             |
| Item 5                   | 0.265 (0.024)             |
| Item 6                   | −0.020 (0.026)            |
| Random part              |                           |
| Hospital-level variance  | 0.019 (0.012)             |
| Ward-level variance      | 0.049 (0.011)             |
| Individual-level variance| 0.201 (0.009)             |
| Item 1 variance          | 0.716 (0.026)             |
| Item 2 variance          | 0.536 (0.020)             |
| Item 3 variance          | 0.354 (0.014)             |
| Item 4 variance          | 0.450 (0.017)             |
| Item 5 variance          | 0.354 (0.014)             |
| Item 6 variance          | 0.516 (0.019)             |
into item-specific measurement error and variance associated with the individual. The item variance is used to calculate the reliability of the scale (see next section). The other variances can be used to calculate variance partition coefficients.

As with the simple model, we can estimate an ecometric model (Fig. 8.2) in which we adjust for individual characteristics. However, once again the empty and adjusted models do not differ much and so we have not shown these results.

Table 8.5 shows the variance components for the model presented in Table 8.4 and for a model adjusted by the number of years spent working on that ward, the number of hours worked per week and the type of employee (nurse or physician). As a consequence of removing the measurement error (item variance) from the individual variance, the intra-class correlations are higher compared to those obtained under the simple approach and shown in Table 8.3. The percentage of the variance at ward and hospital levels combined has increased from 20 to 25%.

So far we have analysed the scale ‘feedback and learning from error’, and we have estimated the variances at the ward and hospital levels. The final step is to calculate and save the ward residuals or effects. Whilst the hospital level is still in the analysis, the ward residuals show the departure from the hospital mean. The ward residuals can be used as an independent variable at ward level in a new analysis. Figure 8.3 shows the ranking of the hospital wards according to how they score on the scale ‘feedback and learning from error’.

**Table 8.5** Variance partition coefficients at hospital and ward level (ecometric approach)

| Level    | Empty model | Adjusted model |
|----------|-------------|----------------|
| Hospital | 0.072       | 0.078          |
| Ward     | 0.183       | 0.180          |

**Fig. 8.3** Ranking of hospital wards on the scale ‘feedback and learning from error’. Ward residuals from the empty model (including the hospital level + mean scale value)
The point estimates of the ward residuals can be used as an independent variable in a new analysis. We then have the ward effect, net of the individual variation in the perceptions of employees about patient safety culture. Some wards score significantly lower on patient safety culture, and some score significantly higher within their hospital. If we want an overall ward effect, we can omit the hospital level from the analysis meaning that the hospital-level variance would all go to the ward level (as described in Chap. 6: Apportioning variation in multilevel models).

**Comparison of the Traditional and Ecometric Approach**

In an analysis of neighbourhood disorder, Steenbeek (2011) compared simple aggregation of the individual-level scale values and ecometric analysis. In his analysis of 71 Dutch neighbourhoods, only 6% had exactly the same rank. Nearly 30% of neighbourhoods moved ranks between the two analyses by more than five positions. With some exceptions, agreement between the two methods was greatest at the extremes, and there were notable differences in ‘average’ neighbourhoods. In a similar manner, we can compare two sets of rankings of the wards in our example, one based on the simple aggregation of scale values and the other based on the ecometric analysis. Figure 8.4 compares the ranks obtained under the two methods.

![Figure 8.4](image)

**Fig. 8.4** Comparison of ranking of hospital wards based on ecometric analysis and the traditional method (aggregated scale values)
The top panel of Fig. 8.4 shows on the horizontal axis the ranking of the wards based on the aggregated scale values, or the ‘traditional method’, and on the vertical axis the ranking based on the ecometric analysis. We note from the top panel that in this particular example the two rankings are fairly consistent. Most of the hospital wards are very close to the diagonal, and this is true at the extremes more than in the middle. Secondly, the lower panel shows the distribution of the differences of the rankings, with the number of wards on the vertical axis and the difference between the ecometric and the traditional ranking on the horizontal axis. In approximately a quarter of the wards, the ranking is the same. In 40% of the wards the difference is two or more ranks out of the 87 wards. The correlation between the scores produced by the traditional method and the ecometric approach is consequently very high. As yet little has been published that cites the correlation between the two approaches. Steenbeek et al. (2012) also found high correlations between the two methods (over 0.90), and Mohnen et al. (2011), in an analysis of neighbourhood social capital, found a correlation of $r = 0.80$. These differences are not very big, but if these were to be used for information relating to public performance, especially when the results are presented by grouping constituents into three or five categories, then even a difference of one or two ranks could move a unit from an ‘average’ category to one described as performing ‘below average’.

Further Ecometric Properties of the Scale

In psychometric analysis, reliability is usually expressed by means of Cronbach’s alpha. There is an equivalent to Cronbach’s alpha in ecometric analysis which takes into account how much agreement there is between observers or respondents evaluating the same ecological unit (the extent of inter-subject agreement), the number of informants or respondents sampled, and the number of items. This is shown in Eq. (8.4).

\[
\text{Reliability} = \frac{\sigma_{v0}^2}{\sigma_{v0}^2 + \sigma_{u0}^2 / \bar{n}_J + \sum_{m=1}^{n_i} \sigma_{em}^2 / n_j \bar{n}_J} \tag{8.4}
\]

In our example $\sigma_{v0}^2$ is the ward-level variance, $\sigma_{u0}^2$ the individual-level variance, $\sum_{m=1}^{n_i} \sigma_{em}^2$ is the item consistency (the sum of the error variances at item level, also known as the measurement error), $\bar{n}_J$ is the average number of individual respondents in a ward, and $n_j$ is the number of items. As the model still includes the hospital level, the ward-level variance relates to the departure from the hospital means.

Using Eq. (8.4), the reliability of the scale ‘feedback and learning from error’ at the ward level can be calculated as follows.
Reliability = \frac{0.049}{0.049 + 0.201/22 + 2.926/(6 \times 22)} = 0.61

This reliability is adequate but not very high. It is a lower reliability than Cronbach’s alpha at the level of the employees which would be calculated using a traditional approach (which was 0.78). It is also much lower than the ward-level reliability, which we would calculate by first aggregating the individual items to ward level and then performing a reliability analysis, giving a value of Cronbach’s alpha of 0.90. This means that a failure to take into account the structure of the data would result in an overestimation of reliability at the ward level.

From Eq. (8.4) it is clear that the average number of observers or respondents per higher level unit is an important determinant. We can see this relationship in Fig. 8.5; reliability increases sharply with the number of observers or respondents per ecological unit. Raudenbush and Sampson (1999), Steenbeek (2011) and, in the field of public health research, Corsi et al. (2012) give graphs like this based on their own data. The form of the relationship is the same. Such graphs can inform us as to the appropriate number of observers or respondents per ecological unit when we want to apply an ecometric analysis. At about 30–40 respondents per ecological unit, the reliability is usually above 0.70. An important cause of low reliability is a small sample size; see, for example, Riva et al. (2011).

The item inter-correlations inform us about whether some of the items might be redundant; very high item inter-correlations suggest that we could have done with fewer items since they appear to measure the same thing. If the item inter-correlations are very low, then the items do not appear to relate to the same latent variable meaning that we could increase reliability by omitting uncorrelated items when constructing the scale. In an ecometric analysis, we can compare the item inter-correlations at the
individual level and at the ward level. In our example, these range between 0.61 and 0.94 at ward level and between 0.26 and 0.49 at the level of the individual employees within the wards. Judging from the ward-level item inter-correlations, we could probably have used fewer items. However, we could not have known this in advance. If we were to develop a measurement instrument for use at an ecological level, the development of this instrument would include an analysis of the item inter-correlations. Based on the results of this analysis, we would be in a position to reconsider the items measured.

We can assess the construct validity of the scale at ward level by examining associations with other contextual measures. As an example, we calculated the correlation of the scale with self-reported frequency of event reporting at ward level. This correlation is 0.63, a moderately high correlation. In wards that have higher scores on the scale ‘feedback and learning from error’, more people tend to say that they frequently report events. However, these are not necessarily the same people who (at individual level) say that they receive feedback about errors. The correlation at the individual level between the scores on the scale ‘feedback and learning from error’ and the self-reported frequency of event reporting is only 0.36.

Conclusions

Ecometrics is a statistical method used to combine multiple data items collected from individuals, be these respondents in a survey or trained observers, about higher level units. This combination of individual responses is used to ascertain properties of the higher level units. We can take into account varying sample sizes associated with the higher level units and consequent reliability, shrinking the estimates for units with few observations towards the overall mean. We can also take into account the composition of the sample and adjust for compositional differences. Ecometrics also allows us to analyse interesting properties of the data, such as the extent of clustering at different levels, and the reliability and difficulty of items.

References

Corsi DJ, Subramanian SV, McKee M, Li W, Swaminathan S, Lopez-Jaramillo P, Avezum A, Lear SA, Dagenais G, Rangarajan S, Teo K, Yusuf S, Chow CK (2012) Environmental Profile of a Community’s Health (EPOCH): an ecometric assessment of measures of the community environment based on individual perception. PLoS One 7:e44410
Gauvin L, Richard L, Craig CL, Spivock M, Riva M, Forster M, Laforest S, Laberge S, Fournel MC, Gagnon H, Gagné S, Potvin L (2005) From walkability to active living potential: an “ecometric” validation study. Am J Prev Med 28(suppl 2):126–133
Gilreath TD, Chaix B, King G, Matthews S, Flisher AJ (2012) Multi-level influence of school norms on tobacco use in South Africa: an ecometric consideration of group differences. Cancer Causes Control 23(suppl 1):27–36
de Jong K, Albin M, Skärbäck E, Grahn P, Wadbro J, Merlo J, Björk J (2011) Area-aggregated assessments of perceived environmental attributes may overcome single-source bias in studies of green environments and health: results from a cross-sectional survey in southern Sweden. Environ Health 10:4

Meuwissen LE, De Bakker DH (2008) [CQ-index GP care: development of the measurement instrument]. NIVEL, Utrecht

Mohnen SM, Groenewegen PP, Völker B, Flap HD (2011) Neighborhood social capital and individual health. Soc Sci Med 72:660–667

Mujahid MS, Diez-Roux AV, Morenoff JD, Raghunathan T (2007) Assessing the measurement properties of neighborhood scales: from psychometrics to ecometrics. Am J Epidemiol 165:858–867

Nyqvist F, Nygård M, Steenbeek W (2014) Social capital and self-rated health amongst older people in Western Finland and Northern Sweden: a multi-level analysis. Int J Behav Med 21:337–347

Oksanen T, Kawachi I, Kouvonen A, Takao S, Suzuki E, Virtanen M, Pentti J, Kivimäki M, Vahtera J (2013) Workplace determinants of social capital: cross-sectional and longitudinal evidence from a Finnish cohort study. PLoS One 8:e65846

Prins RG, Mohnen SM, van Lenthe FJ, Brug J, Oenema A (2012) Are neighbourhood social capital and availability of sports facilities related to sports participation among Dutch adolescents? Int J Behav Nutr Phys Act 9:90

Raudenbush SW (2003) The quantitative assessment of neighborhood social environments. In: Kawachi I, Berkman LF (eds) Neighborhoods and health. Oxford University Press, Oxford

Raudenbush SW, Sampson RJ (1999) Ecometrics: toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. Sociol Methodol 29:1–41

Riva M, Bamba C, Curtis S, Gauvin L (2011) Collective resources or local social inequalities? Examining the social determinants of mental health in rural areas. Eur J Public Health 21:197–203

van Schoten SM, Baines RJ, Spreeuwenberg P, de Bruijne MC, Groenewegen PP, Groeneweg J, Wagner C (2014) The ecometric properties of a measurement instrument for prospective risk analysis in hospital departments. BMC Health Serv Res 14:103

Smits M (2009) Unintended events in hospitals: causes and the role of patient safety culture. PhD thesis (Free University, Amsterdam). NIVEL, Utrecht

Smits M, Christiaans-Dingelhoff I, Wagner C, van der Wal G, Groenewegen PP (2008) The psychometric properties of the ‘Hospital Survey on Patient Safety Culture’ in Dutch hospitals. BMC Health Serv Res 8:230

Steenbeek W (2011) Social and physical disorder: how community, business presence and entrepreneurs influence disorder in Dutch neighbourhoods. PhD thesis, ICS, Utrecht

Steenbeek W, Völker B, Flap H, van Oort F (2012) Local businesses as attractors or preventers of neighborhood disorder. J Res Crime Delinq 49:213–248

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.