This is the second video outlining the basics of multilevel modelling. In this session I'll introduce the two-level random intercept model for continuous data. This is probably the most widely used multilevel model and allows us to account for basic clustering, whilst also identifying relevant group level predictors. I'll start by introducing the simplest multilevel model for continuous data, a two-level model with no explanatory variables, sometimes called the variance components model. I'll then outline how to quantify group differences using the variance partition coefficient. Finally, I'll include explanatory variables in the model, starting with variables measured at the lowest level between individuals, before looking at variables measured at the group level.

Two level random intercept models extend standard linear regression models by re-partitioning the residual error between an individual and a group component. This can allow us to gain an initial picture of the importance of groups, where no explanatory variables are included. To explain the logic behind the random intercept model, I want to discuss a very simple example with just eight data points, that each record a person's height. We can represent the distribution of heights, with the mean height across the sample, and the variance, or standard deviation which captures the spread of heights. We can also think of this as a basic regression model with no explanatory variables. Here the height for any individual, YI, is equal to the mean height across the population, beta naught, and the individuals residual difference from this mean. So, the height for person 1, Y1, is equal to beta naught plus individual 1's residual difference from the mean, E1. We then assume that the residuals are approximately normal with mean zero in variant Sigma Squared E. But if we know the individuals belong to different groups, or families in this case, we can capitalize on this additional information and improve our estimates of the height of any individual. So here our 8 data points come from two families, one represented by green triangles and one by red circles. Now we can refer to YIJ, there's the height for individual I from Group J, we still have an estimate of the mean height beta naught, which now refers the average height of people across all families; and we include an additional set of residuals labelled UJ, which refer to the group mean differences in height from the overall mean. So, people from Group 1 are generally taller than average, and the average height of family 1 is equal to the overall average beta naught, plus the residual U1, whilst people from group 2 are generally shorter than average. Like the individual residuals in the single level model, these group residuals are assumed approximately normal with zero mean and variance which we label Sigma Squared U.

And now we also have individual residual differences around the family specific means, which we also assume are approximately normal with mean 0 and variance Sigma Squared E. Now we've re-partitioned the variance between an individual component, Sigma Squared E, and a group component Sigma Squared U. Now we can produce an initial assessment of the importance of groups with the variance partition coefficient. This looks at what proportion of the total variance, made up of Sigma Squared U and Sigma Squared E, can be attributable to differences between groups Sigma Squared U. This ranges from 0, when there is no group effect, to 1 where there are within group differences.

Turning to some real data, here we have measurements of fear of crime, which is a standardized scale, where higher scores mean more fear, for a total of 27,764 residents that live in 3,390 areas of England. There is an average of 8 residents in each area with a maximum of 47. Here we have our empty model which partitions the variation in fear between individuals, with a variance of 0.863 and areas with a variance of 0.145. With our individual and group variance estimates we can calculate an initial assessment of the importance of groups, which is 0.145 over the total variance. This gives us an estimate of 0.144 or just over 14% of the variability is allocated to between group differences. We can then add explanatory variables to our model in exactly the same way we do in single level regression models. So, adding a single explanatory variable weight to our model of mean heights, the regression line represents the line of best fit, that's closest to all data points simultaneously, or which minimizes the squared residuals. Now beta naught refers to the point where the line crosses the y axis, and beta one quantifies the gradient of the line or how much height increases for one unit increase in weight. And for each family, we assume the same increase in height, for each unit increase in weight, the lines are parallel, but the point where the line crosses the y axis is allowed to vary by the residual UJ; And we still assume the individual residuals EIJ, and group level residuals UJ, are approximately normal with mean 0 and variance is Sigma Squared E and Sigma Squared U.

One of the key strengths of multilevel models is that they allow us to simultaneously include group level information. This allows us to formulate interesting questions about the role of context. Group effects can be either external e.g. administrative data, or can be aggregates of included individual level variables, although the latter depends at least, a bit, on the group size. There's no need to explicitly identify these as group effects, this is captured by the group residual.

So, returning to our worked example using the crime survey for England and Wales, in model two we include three explanatory variables. The first two, age, and whether or not you've been a victim of crime in the last year, a dummy variable, are measured at the individual level. This can be seen because of the IJ subscripts. The final variable, neighbourhood crime rate, is measured at the group level with only a subscript J. Here we see generally lower levels of fear amongst older residents, and notably higher levels of fear amongst victims of crime, when compared to non-victims. We also see that residents of areas that have a higher crime rates are generally more fearful than residents in areas with lower crime rates. Having accounted for these three explanatory variables, we see a reduction in the variability at the individual and group levels. We consider these to be approximate R Squares by considering the reductions in variance in each level. So, at the individual level, the variance dropped from 0.863 to 0.85, a drop of roughly 1.5%, and at the group level, accounting for the crime rate, leaves the area variance to fall from 0.145 to 0.105, a drop of nearly 28%.

So, in this session we've introduced the variance components model and the random intercept model. The variance components model can be used to provide an initial estimate of the contribution of groups. The random intercept model allows us to include explanatory variables at the individual and group level to explain variation in our dependent variable.