# Fixed and random effects models

## Transcript video 1

Full resource: https://www.ncrm.ac.uk/resources/online/all/?id=20842

Hello. My name’s Kevin Ralston and I’m a Lecturer on Sociology and Quantitative Methods here at the University of Edinburgh. We’ve prepared a resource that provides examples of the analysis of fixed and random effects models using the software Stata. In the resource, fixed and random effects models are compared to a standard regression model and also where regression accounts for clustering in the data using Huber-White standard errors. Beyond this, we also present some models that incorporate both fixed and random effects, so the resource provides examples of the (inaudible 0:00:52) model and Allison’s hybrid model. The resource comprises of an article on methods that you are welcome to take a read of, and also a program code file that enables you to run the analysis presented and some additional analyses. This first presentation provides a general introduction to fixed and random effects models.

So, there are a number of data structures to which analysts would consider fitting random effects and/or fixed effect models. And table one here summarises some of these. Now, the focus of this resource is on the classic panel data structure, so this is longitudinal panel data where the structure is that of individuals measured at different occasions of timepoints. Now, this is the data structure of the type provided by datasets like the British Household Panel Survey or the Understanding Society datasets. For the examples we provide, the outcome variables are linear metrics. Much of what follows generalises to non-linear outcomes, but researchers should be aware that there is additional complexity that needs to be considered in modelling non-linear outcome variables in fixed effect or random effect frameworks. In the resource, when we are referring to random effects, this will mean the random intercepts model and not alternatives, such as random slopes or random intercepts, random slopes and models.

So, random and fixed effect models are also known as panel data models because they take account of the multiple measurement points of individuals measured in panel data. Table two here describes a very simple panel dataset. In this panel there are two individuals measured at three different timepoints or occasions. So, we can see person one, person two, years 2016, 2017, 2018, and we can see that one has sex quoted as 1 and one has sex quoted as 2, so we might presume that we have a man and a woman and these data. So, there are two individuals measured at three different occasions. Now, we would not wish to fit an OLF regression model to these data. If we did this, we would be violating the assumption of independence. These cases are not independent of one another; they are nested within two individuals. So, if we were to fit an OLS model, well it is likely that our standard errors would be too small because we would be assuming that there are six separate cases here when there are really only two different cases.

So, fixed and random effects models enable us to take account of this panel data structure, where there are occasions nested within individuals. And we can give this some handy notation. So, we can give the level two in the data structure the subscript i for individuals, and we can give the level one in the data the subscript t for occasions or timepoints when individuals are observed.

So, what are fixed and random effects models? So, let’s think about the fixed effects model first. So, Allison here argues that a fixed effect model treats unobserved differences between individuals as a set of fixed parameters that can either be directly estimated or partialed out of estimating equations. Now, this has some remarkable and useful properties. The fixed effect controls for all stable unobserved variables. This includes variables that have not and cannot be measured, and this is remarkable and very useful. Now, this is because each individual becomes their own control in the data, and because of this, all individual variation is accounted for in the fixed effect. All time and variant differences between individuals are contained in fixed effect, and any time variant differences can be estimated in the model, so we obtain a parameter for the time and variant differences fixed effect and we can estimate the time varying differences. Because all time and variant differences between individuals are incorporating the fixed effect, we cannot estimate time and variant parameters within a fixed effect framework. Now, this can be a little bit of a problem and it's likely to pose problems for many substantive research issues, for example, if we’re interested in sex or ethnicity, this is unlikely to vary within an individual over time so we cannot estimate parameters for these in a fixed effect framework and a fixed effect model.

So, what is the random effects model? Well, classically in a random effects model individual differences are considered as random variables drawn from a specified distribution, and it is now common to always regard the unobserved differences as random variables. Allison argues that what distinguishes the fixed effect from the random effect is defined by the structure of the associations between observed and unobserved variables. And a random effect model can account for change over time and time and variant variables. But this leads to a problem. The random effect approach assumes that unobserved variables are uncorrelated or are independent of observed variables.

We can maybe get some of a sense of this by taking a look at the algebra for each of the models. So, in the fixed effect framework top model here on the slide, all unobserved variables controlled in the fixed effect are dented by the term lambda i. A useful way of thinking about this for those familiar with OLS regression is that it would be easy with a small dataset to include a dominant variable for each individual in the dataset, with one individual acting as a reference category. Now, this would have the effect of raising or loading regression line, depending on the average individual level effects, so that’s a fixed effect. And in the random effect framework, there are two components to the error distribution. The EIT component or epsilon i t component is the familiar term for the individual at a given timepoint or occasion. The second component is an individual parameter that summarises the overall distribution of the individual respondents’ difference, e.g. the variance for this distribution. These two error components enable inclusion of time in a variant effects in the random effects model. But, as we’ve heard, an assumption of the random effects model is the unobserved characteristics are uncorrelated with the observed variables. Correlation between the observed and unobserved variables may lead to bias and random effects estimates.

So, just to summarise that introduction, some of the key issues are the fixed effect summarises patterns of change within individuals. This is a powerful method that can estimate explanatory factors with change over time so you can obtain parameters in your model for explanatory factors that change over time. It’s considered to produce optimally low standard errors because between subject variance is not part of the error, and this is sometimes described as consistent in the literature. Random effect analysis change within and between individuals. This framework can estimate explanatory factors which change over time and as well as the time-constant ones. But an assumption here is that unobserved variables are uncorrelated with the observed variables, and this approach is sometimes described in the literature as more efficient because it uses more information to estimate parameters.

In our next presentation we will go on to look at some worked examples of these approaches within this data.

National Centre for Research Methods (NCRM)
Social Sciences
Murray Building (Bldg 58)
University of Southampton
Southampton SO17 1BJ
United Kingdom

**Web** www.ncrm.ac.uk
**Email** info@ncrm.ac.uk
**Tel** +44 23 8059 4539
**Twitter** @NCRMUK