

9.0 Regression Analysis

Regression analysis is used to examine the relationship between a given variable and one or more other variables. The NIMDM results can be included within a regression analysis as an independent (explanatory) and/or dependent (explained) variable when analysing cross-sectional data. The following section describes which results can be included as the independent or dependent variables.

Including NIMDM Results in Regression Models

The Income Domain, Employment Domain, Income Deprivation Affecting Children or Income Deprivation Affecting Older People scores can be added to the regression model as the dependent variable or the independent variable in a linear regression model. The scores associated with the remaining deprivation domains and the Multiple Deprivation Measure cannot be included.

If for example the income domain score is added to the regression model as an independent variable, the interpretation of the estimated coefficient 'x' is, all other things being equal, as the percentage of income deprived people in an area increases by one, the expected change in the independent variable is 'x'.

It is recommended that the only scores appropriate for inclusion in regression analysis are the Income, Employment, Income Deprivation Affecting Children and Income Deprivation Affecting Older People scores.

The majority of the NIMDM results however are presented as ranks. As discussed previously, a rank is an example of ordinal data, as although from the deprivation ranks it can be said that an area is more or less deprived than another, it cannot be said *how much* more or *how much* less deprived. The difference in deprivation levels is not uniform across ranks, such that an increase of one rank in the lower deciles is not equivalent to an increase of one rank in the higher deciles.

In a linear regression model, the normal interpretation of the estimated coefficient of an independent variable of value 'x' is that, all other things being equal, as the independent variable increases by one unit, the expected change in the dependent variable is 'x' units. If the independent variable entered into the regression model is the multiple deprivation rank, the above interpretation is not meaningful.

To include the Multiple Deprivation Measure ranks in a linear regression model they should be converted to dummy variables for deprivation groups and added as independent variables.

It is therefore recommended that ranks are only included in a linear regression analysis when grouped and entered in the form of dummy variables.

Similarly the deprivation ranks can be grouped and used in a qualitative response models such as logit or probit models.

Deprivation ranks can also be grouped to form the dependent variable in a qualitative response model, such as a logistic regression used to examine possible factors affecting the likelihood of living in a deprived area.

Case Study - The Centre of Excellence at Queen's University Belfast, in conjunction with the Institute of Public Health is producing estimates and forecasts of chronic conditions in Northern Ireland such as hypertension, stroke, coronary heart disease and diabetes.

The Northern Ireland estimates/forecasts are produced using risk estimates derived from carrying out a logistic regression analysis of the Health and Wellbeing Survey data. One of the explanatory variables used in the regression analysis is area deprivation. This deprivation variable has five categories derived by assigning ranks for the 890 Super Output Areas and 5,022 Census Output Areas into five quintiles (quintile 1 least deprived through to quintile 5 most deprived quintile).

As the NIMDM results include seven domains of deprivation as well as the overall multiple deprivation measure, it is possible to include one or more of the domains as independent variables if for example both education deprivation and proximity to services are the interest.

The high positive correlations between some of the domains and the implications this will have on the coefficients and associated standard errors should be taken into consideration when employing this approach. It is not recommended to include both the multiple deprivation measure and one or more of the individual domains in the regression model as the domains are components of the overall multiple deprivation measure.

Similarly when including the NIMDM results in a regression model it is important to examine the component indicators. In total 52 indicators were included in the Multiple Deprivation Measure and it may be the case that one or more of the indicators is already included in the model. The extent to which this may impact on the accuracy of the estimates should be assessed by taking into account the weights of individual indicators within the domain or multiple deprivation measure.

It is recommended that the correlations between domains of deprivation and the component indicators are examined before including in a regression analysis.

Contact Point

Cathryn McBurney

Neighbourhood Statistics

NISRA

McAuley House

2-14 Castle Street

Belfast BT1 1SA

Tel: 028 90 348 112

Email: deprivation.nisra@dfpni.gov.uk

Website: www.nisra.gov.uk/deprivation.htm