

Lab 3: Regression

The variables in the Excel data set are **Lab3emp** are as follows:

Data description and measurement scales
<p>‘Wage’ is the respondent’s hourly wage rate.</p> <p>‘Age’ is the respondent’s age in years.</p> <p>‘Doc’ is the respondent’s score on a document literacy test.</p> <p>‘Quant’ is the respondent’s score on a numeracy test.</p> <p>‘Educ’ is the number of years the individual has spent in full-time education (or equivalent).</p> <p>‘Sex’ is the biological sex at birth of the respondent (1 = female; 2 = male).</p> <p>‘Region’ indicates the respondent’s region of residence where: 1 = Munster, 2 = Connaught, 3 = Leinster and 4 = Ulster</p>

1. Wage is the outcome variable of interest. Fit the ‘best’ regression model to the data using R.

In doing so consider the following:

- Are there any issues with multicollinearity?
 - Do the residuals infer that a transformation would be appropriate?
 - Use the R^2 and Adjusted R^2 to determine the model fit.
 - Use backward regression to remove any non-significant variables.
 - Interpret the findings of the regression model.
2. Use the regression model to predict a Wage for a 45 year old female with a Quant score of 400, 17 years in education and living in Leinster.
 3. Use the regression model to identify (i) the person who is paid much more than would be expected and (ii) the person who is paid much less than would be expected.