Multiple regression

Last updated: December 2014

Background

The data presented in Chapter 5 of Family Food 2013 on demographic patterns in key dietary indicators was analysed using multiple regression. Multiple regression was used to isolate the effects of one demographic characteristic while controlling for differences in others. The demographic characteristics (independent variables) analysed were region, age of HRP, ethnicity of HRP and equivalised income. The dependent variables investigated were the key dietary indicators:

- Sodium,
- saturated fatty acids,
- Non-Milk Extrinsic Sugars (NMES),
- fruit,
- vegetables and
- fibre.

Analysis

The analysis was carried out using SPSS version 18. Taking sodium (dependent variable) as an example (shown below using 2010 data) - this was analysed across all of the independent variables which were region, age of HRP, ethnicity of HRP and equivalised income. The baseline household characteristics referred to in Chapter 4 Table 4.1, are the most frequently occurring in the data because this provides the most powerful statistical comparisons.

In multiple regression, the values of one variable (the dependent variable y) are estimated from those of two or more other variables (the independent variables x_1, x_2, …, x_p). This is achieved by the construction of a linear multiple regression equation of the general form as follows:

$$\hat{y} = B_0 + B_1x_1 + B_2x_2 + B_3x_3 + B_4x_4 + B_5x_5$$

where there is an equation for each household such that:

$\hat{y}$ = predicted average sodium content of food purchases per person per day

B_0 = constant representing sodium for the baseline characteristics

B_1x_1 = corrections for region if different to the baseline region

B_2x_2 = corrections for household composition if different to the baseline

B_3x_3 = corrections for income if different to the baseline

B_4x_4 = corrections for ethnic origin if different to the baseline

B_5x_5 = corrections for age of HRP if different to the baseline

B_1 is a vector of 11 parameters where there is a separate one for each region apart from the baseline. The regression model provides estimates for each of the 11 parameters, which represent differences in sodium between the particular region and the baseline region. Similarly B_2 B_3 B_4 B_5 are vectors of parameters that are estimated by the regression model and cover the different categories of household composition, income decile, ethnic origin group, and the 6 age bands of the HRP.

To illustrate the approach we show details of the multiple regression analysis carried out on sodium. Part of the output covering demographic characteristic age of HRP taken directly from SPSS 18 was the following:

Coefficients					
	Unstandardised Coefficients		Standardised Coefficients	t	Sig.
	B	Std.	Beta		
(Constant)	3.324	.050		67.096	.000
under 30	-.524	.037	-.101	-14.078	.000
30 to 39	-.360	.032	-.084	-11.174	.000
50 to 59	.125	.033	.030	3.813	.000
60 to 69	.189	.035	.044	5.340	.000
70 to 79	-.246	.040	-.049	-6.159	.000
80 and over	-.536	.045	-.089	-11.856	.000

The ‘constant’ relates to the age of HRP in the baseline household which is the age group 40 to 49 years. Through multiple regression analysis, households in the 40 to 49 years age group, and with baseline characteristics in the other independent variables, purchased foods containing on average 3.324 grams of sodium. For an age group other than the baseline it is necessary to add the coefficient for that age group to the constant figure. So to work out the average content of sodium within food purchases for the under 30 age group you would:

Average sodium content for the 40 to 49 years =

3.324 (constant) + -0.524(coefficient) = 2.8 grams

Therefore on average households where the HRP was aged under 30, and with baseline characteristics in the other independent variables, purchased foods containing 2.8 grams of sodium per person per day.

Multiple regression analysis also informs the standard error, which tells us how confident we are with the data. Taking the 'constant' value as an example we use the standard error to calculate the upper and lower limit for error of margin. The calculation is as follows:

Upper limit = 3.324 + 0.050 = 3.4 grams of sodium

Lower limit = 3.324 – 0.050 = 3.2 grams of sodium

Therefore we are confident that the average sodium content of households where age of HRP is aged between 40 and 49 years is plus or minus 0.1 grams of sodium from the estimated average of 3.3 grams.