Module 5 - Ordinal Regression

Objectives
1. Understand the principles and theories underlying Ordinal Regression
2. Understand the assumption of Proportional Odds and how to test it
3. Be able to implement Ordinal Regression analyses using SPSS and accurately interpret the output
4. Be able to include interaction terms in your ordinal regression model and to accurately interpret the output
5. Appreciate the applications of Ordinal Regression in education research and think about how it may be useful in your own research

Start Module 5: Ordinal Regression

Using multiple variables to predict ordinal outcomes.

You can jump to specific pages using the contents list below. If you are new to this module start at the Introduction and work through section by section using the 'Next' and 'Previous' buttons at the top and bottom of each page. Be sure to tackle the exercise and the quiz to get a good understanding.

Contents

5.1 Introduction

In previous modules we have seen how we can use linear regression to model a continuous outcome measure (like age 14 test score), and also logistic regression to model a binary outcome (like achieving 5+ GCSE A*-C passes). However you will remember from the **Foundation Module** that we typically define measures at three levels: nominal, ordinal and continuous. What we have not covered therefore is this 'intermediate' level where our outcome is ordinal. You will remember that an ordinal measure includes information on rank ordering within the data. For example we might have Likert scale measures such as "How strongly do you agree that you love statistics" which may be rated on a 5 point scale ranging from strongly disagree (1) to strongly agree (5). Another example is OFSTED (Office for Standards in Education) lesson evaluations which may be graded as 'unsatisfactory', 'satisfactory', 'good' or 'outstanding'. Such examples are common in the social sciences.

There are a number of ordinal outcomes in our LSYPE dataset. One is the KS3 (age 14) English test level. In England students' performance is recorded in terms of national curriculum (NC) levels. These levels are reported on an age related scale, with the 'typical' student at age 7 expected to achieve level 2, at age 9 level 3, at age 11 level 4, and at age 14 somewhere between level 5 and level 6. These levels may be determined through teacher assessment or be expressed as summaries from continuous test marks. **Figure 5.1.1** shows the distribution of students by English level from our dataset.

Figure 5.1.1 Proportion of students at each English test level
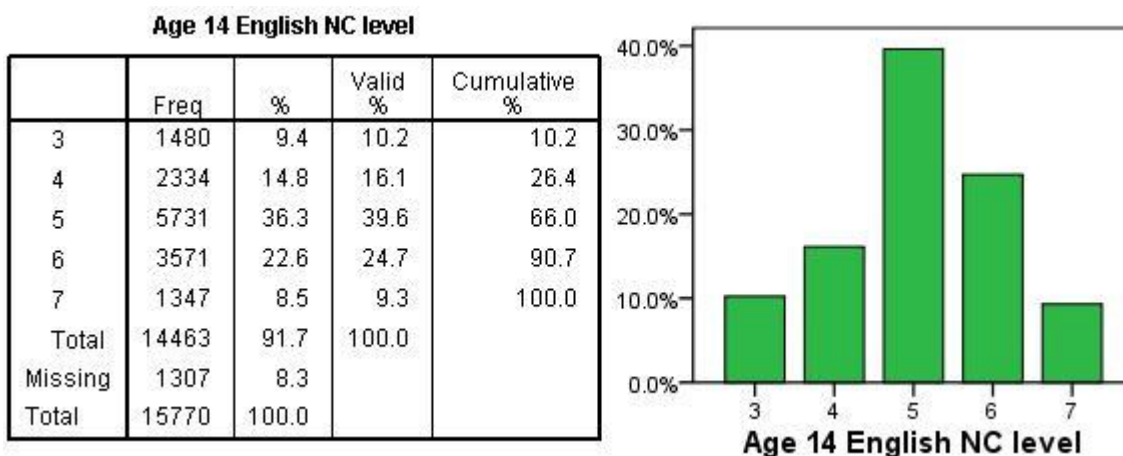
We do have access to the actual test marks in LSYPE, but often test marks are not available and NC levels might be the only data recorded. In any event, this is a good example of an

ordinal outcome which we can work with to demonstrate the particular analyses that you can apply when your outcome measure is ordinal.

The good news is that, bar a little extra work, the assumptions and concepts we need for ordinal regression have been dealt with in the **Logistic Regression Module** (Phew!). The key concepts of odds, log-odds (logits), probabilities and so on are common to both analyses. It is absolutely vital therefore that ***you do not undertake this module until you have completed the logistic regression module,*** otherwise you will come unstuck. This module assumes that you have already completed **Module 4** and are familiar with undertaking and interpreting logistic regression.

5.2 Working with ordinal outcomes

There are three general ways we can approach modelling of an ordinal outcome:

A) Treat the outcome as a continuous variable

You may look at **Figure 5.1.1** and ask why you cannot treat this as a continuous variable and use linear regression analysis. After all, there are a reasonable range of categories (five), with a fair spread of observations over all the categories and an approximately normal distribution. While this may not be unreasonable in this particular case, it does mean making assumptions about continuity in the data which are not strictly verifiable, and of course a mean level is not what we want to predict when our outcome is strictly ordinal (for example a student cannot achieve level 3.75 or level 4.63 in the National Curriculum in England - levels can only be awarded as whole numbers; 4, 5, 6 etc.). There are many other cases and examples where the linear assumption will not hold, where there are fewer than five categories or an uneven distribution across categories, or it is unreasonable to suppose an underlying continuous distribution. In such cases the choice of ordinal regression may be even clearer!

B) Treat the outcome as a series of binary logistic equations

We could treat the analysis as a series of logistic regressions by splitting or cutting the distribution at key points. This is illustrated in **Figure 5.2.1**.

Figure 5.2.1: Four different ways to split the English NC level outcome

N cases below level	English national curriuculum level achieved					N cases at or above level	% of cases at or above level
	3	4	5	6	7		
13116					Level 7+	1347	9.3%
9545				Level 6+		4918	34.0%
3814			Level 5+			10649	73.6%
1480		Level 4+				12983	89.8%
						14463	100.0%

For example, we may consider comparing those students who have achieved level 7 versus those who have not using a logistic regression. We might want to ask whether girls were more likely to achieve this level of success than boys, or whether there are ethnic or social class differences in the probability of achieving level 7. We can do the same thing for those who

achieve level 6 *or above,* compared to those who achieve below level 6. Again this is a binary logistic regression, splitting the sample into two, only this time in a different place. The same can be done to compare the probability of achieving level 5 or above, and again for the probability of achieving level 4 or above. In each case we complete a binary logistic regression to evaluate the effect of our explanatory variables on the likelihood of success at different *thresholds* (level 4+, level 5+, level 6+ and level 7). Note we do not need a category for level 3+ because this includes all (100%) of the cases in our data.

Essentially we have turned our outcome into a series of binary measures reflecting the *cumulative* outcomes at different thresholds. However estimating four separate binary logistic regression equations is wasteful of the information on ordinality in our outcome and may lead to estimating more parameters than are necessary to account for the relationships between our explanatory variables and the outcome (four sets of estimated regression coefficients rather than one set). What we want ideally is a single model of the effect of our explanatory variables on the outcome which utilises the ordinality present in the outcome variable.

C) Model the ordinality in the outcome

In ordinal regression instead of modelling the probability of an individual event, as we do in logistic regression, we are considering the probability of that event and all others above it in the ordinal ranking. We are concerned with *cumulative* probabilities rather than probabilities for discrete categories. If a single model could be used to estimate the odds of being at or above a given threshold across all cumulative splits, the model would offer far greater parsimony compared to fitting multiple (in the case of our English level example, four) separate logistic regression models corresponding to the sequential splits in the distribution as illustrated above. The goal of such a *cumulative odds model* is to simultaneously consider the effects of a set of explanatory variables across these possible consecutive cumulative splits in the outcome. To do this we make the simplifying assumption that the effects of our explanatory variables are the same across the different thresholds, the assumption of *proportional odds.* If this assumption is met there is much to gain from a single parsimonious model, as we shall see. Let us now look at this important assumption of proportional odds in more detail.

5.3 Key assumption of ordinal regression

Overview

What do we mean by the assumption of proportional odds (PO)? To explain this we need to think about the cumulative odds. **Figure 5.3.1** takes the data from **Figure 5.1.1** to show the number of students at each NC English level, the *cumulative number* of students achieving each level *or above* and the *cumulative proportion.* Remember proportions are just the % divided by 100. We can see that the proportion achieving level 7 is 0.09 (or 9%), the proportion achieving level 6 or above is 0.34 (34%) and so on.

From this we can calculate the cumulative odds of achieving each level or above (if you require a reminder on odds and exponents why not check out **Page 4.2**?). 1,347 students achieved level 7 compared to 13,116 who achieved level 6 or below. Therefore the odds of achieving level 7 are 1,347/13,116 = 0.10. Similarly the odds of being at level 6 *or above* are 4918 / 9545 = .52. The odds of achieving level 6 or above are about half that of achieving level 5 or below. If you are getting confused about the difference between odds and proportions remember that odds can be calculated directly from proportions by the formula p / (1-p). Therefore the cumulative odds of achieving level 7 are .09 / (1-.09) = 0.10. Similarly the cumulative odds of achieving level 6 or above are .34 / (1-0.34) =.52. We can do the same to find the cumulative odds of achieving level 5 or above (2.79) and level 4 or above (8.77). We do not need to calculate the cumulative odds for level 3 or above since this includes the whole sample, i.e. the cumulative proportion is 1 (or 100%). As you can see we have essentially divided our ordinal outcome variable in to four *thresholds.*

Table 5.3.1: Cumulative odds for English level

English level	3	4	5	6	7
Number students	1480	2334	5731	3571	1347
Cumulative N at each level *or above*	14463	12983	10649	4918	1347
Cumulative proportion	1.00	0.90	0.74	0.34	0.09
Cumulative odds	-	8.77	2.79	0.52	0.10
Cumulative logits	-	2.17	1.03	-0.66	-2.28

In the table we have also shown the cumulative log-odds (logits), this is just the natural log of the cumulative odds[1] which you can calculate in EXCEL or a scientific calculator. Log odds

---

[1] *If you want to use the LOG function in EXCEL to find the logit for the odds remember you need to explicitly define the base as the natural log (approx. 2.718) e.g. =LOG(odds,2.718)*

rather than odds are used in ordinal regression for the same reason as in logistic regression (i.e. they do not suffer from the ceiling and floor effects that odds do, you should remember this from **Module 4**).

The key assumption in ordinal regression is that the effects of any explanatory variables are consistent or *proportional* across the different thresholds, hence this is usually termed the *assumption of proportional odds (S*PSS calls this the assumption of *parallel lines* but it's the same thing). This assumes that the explanatory variables have the same effect on the odds regardless of the threshold. For example if a set of separate binary logistic regressions were fitted to the data, a common odds ratio for an explanatory variable would be observed across all the regressions. In ordinal regression there will be separate intercept terms at each threshold, but a single odds ratio (OR) for the effect of each explanatory variable. This is best explained by an example.

As example using gender and English NC level

As a simple example let's start by just considering gender as an explanatory variable. Before you start building your model you should always examine your 'raw' data. **Figure 5.3.2** shows the cross tabulation of English level by gender.

Figure 5.3.2: Gender by English level crosstabulation

			Age 14 English NC level					Total
			3	4	5	6	7	
Gender	0 Male	Count	967	1372	2835	1500	503	7177
		%	13.5%	19.1%	39.5%	20.9%	7.0%	100.0%
	1 Female	Count	462	904	2780	2015	828	6989
		%	6.6%	12.9%	39.8%	28.8%	11.8%	100.0%
Total		Count	1429	2276	5615	3515	1331	14166
		%	10.1%	16.1%	39.6%	24.8%	9.4%	100.0%

Clearly girls tend to achieve higher outcome levels in English than boys. What does this look like in terms of the cumulative proportions and cumulative odds? In **Figure 5.3.3** we calculate the cumulative odds separately for boys and for girls.

Figure 5.3.3: Cumulative odds for English NC level separately for boys and girls

Boys	3	4	5	6	7
Cumulative N boys	7177	6210	4838	2003	503
Cumulative proportion	1.00	0.87	0.67	0.28	0.07
Cumulative odds	-	6.42	2.07	0.39	0.08
Cumulative logits	-	1.86	0.73	-0.95	-2.59
Girls	**3**	**4**	**5**	**6**	**7**
Cumulative N girls	6987	6525	5621	2841	826
Cumulative proportion	1.00	0.93	0.80	0.41	0.12
Cumulative odds	-	14.12	4.11	0.69	0.13
Cumulative logits	-	2.65	1.41	-0.38	-2.01
Odds Ratio (Girls/Boys)	**-**	**2.20**	**1.99**	**1.77**	**1.78**
Odds Ratio (Boys/Girls)		**0.45**	**0.50**	**0.56**	**0.56**

We can calculate odds ratios by dividing the odds for girls by the odds for boys. In general the odds for girls are always higher than the odds for boys, as proportionately more girls achieve the higher levels than do boys. These odds ratios do vary slightly at the different category thresholds, but if these ratios *do not differ significantly* then we can summarise the relationship between gender and English level in a single odds ratio and therefore justify the use of an ordinal (proportional odds) regression. If we do calculate the odds ratio from an ordinal regression model (as we will do below) this gives us an OR of 0.53 (boys/girls) or equivalently 1.88 (girls/boys), which is not far from the average across the four thresholds. This assumes the odds for girls of achieving level 4+ are 1.88 greater than the odds for boys; the odds of girls achieving level 5+ are 1.88 times greater than the odds for boys, and so on for level 6+ and level 7... i.e. that the odds of success for girls are almost twice the odds of success for boys, *wherever* you split the cumulative distribution (that is to say, whatever threshold you are considering). SPSS has a statistical test to evaluate the plausibility of this assumption, which we discuss on the next page (**Page 5.4**).

5.4 Running an ordinal regression on SPSS

So let's see how to complete an ordinal regression in SPSS, using our example of NC English levels as the outcome and looking at gender as an explanatory variable.

Data preparation

Before we get started, a couple of quick notes on how the SPSS ordinal regression procedure works with the data, because it differs from logistic regression. First, for the dependent (outcome) variable, SPSS actually models the probability of achieving each level or below (rather than each level or above). This differs from our example above and what we do for logistic regression. However this makes little practical difference to the calculation, we just have to be careful how we interpret the direction of the resulting coefficients for our explanatory variables. Don't worry; this will be clear in the example. Second, for categorical (nominal or ordinal) explanatory variables, unlike logistic regression, we do not have the option to directly specify the reference category (LAST or FIRST, see **Page 4.11**) as SPSS ordinal automatically takes the LAST category as the reference category. So for our gender variable (scored boys=0, girls=1) girls will be the reference category and the coefficients will be for boys. Again this is not a huge problem because if we want to we can simply RECODE our variables to force a particular category as the reference category (e.g. if we wanted boys to be the reference category we could recode gender so girls=0 and boys=1). It is, however, slightly fiddly and annoying!

Requesting an ordinal regression

You access the menu via: **Analyses > Regression > Ordinal.** The window shown below opens. Move English level (*k3en)* to the 'Dependent' box and gender to the 'Factor(s)' box.
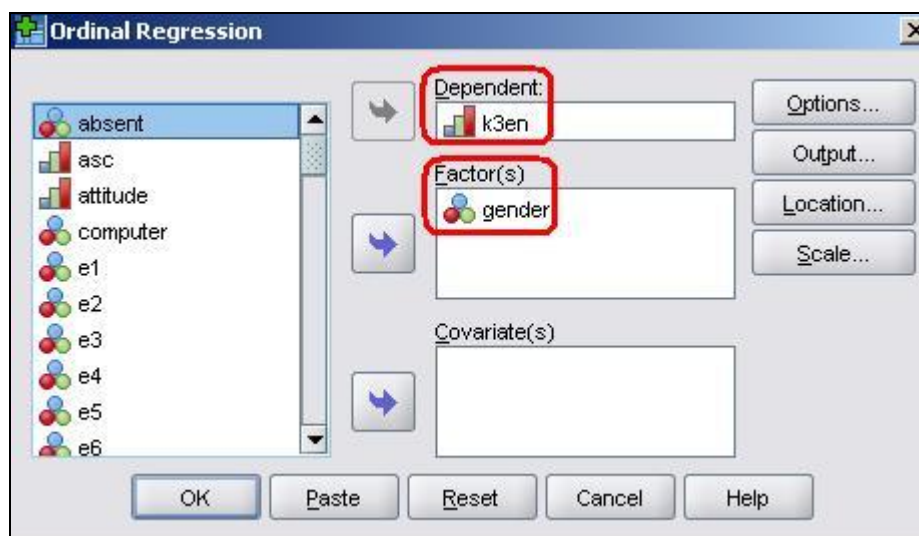
Next click on the 'Output' button. Here we can specify additional outputs. Place a tick in *Cell Information.* For relatively simple models with a few factors this can help in evaluating the model. However, this is not recommended for models with many factors or for models with continuous covariates, since such models typically result in very large tables which are often of limited value in evaluating the model because they are so extensive (they are so extensive, in fact, that they are likely to cause severe mental distress). Also place a tick in the *Test of parallel lines* box. This is essential as it will ask SPSS to perform a test of the proportional odds (or parallel lines) assumption underlying the ordinal model (see **Page 5.3**).
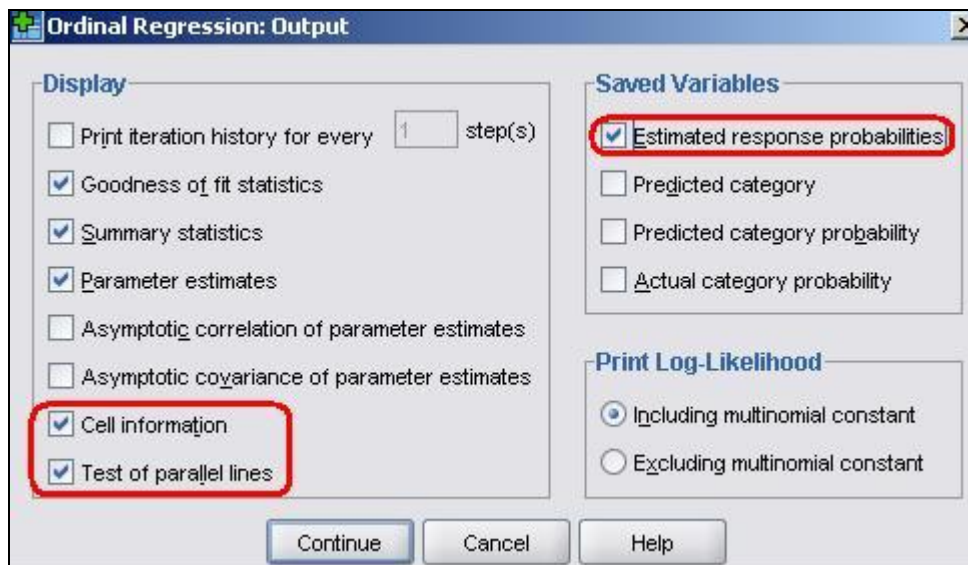
You also see here options to save new variables (see under the 'Saved Variables' heading) back to your SPSS data file. This can be particularly useful during model diagnostics. Put a tick in the *Estimated response probabilities* box. This will save, for each case in the data file, the predicted probability of achieving each outcome category, in this case the estimated probabilities of the student achieving each of the levels (3, 4, 5, 6 and 7).

That is all we need to change in this example so click *Continue* to close the submenu and then **OK** on the main menu to run the analysis...

Examining the SPSS ordinal output

Several tables of thrilling numeric output will pour forth in to the output window. Let's work through it together. **Figure 5.4.1** shows the *Case processing summary.* SPSS clearly labels the variables and their values for the variables included in the analysis. This is important to check you are analysing the variables you want to. Here I can see we are modelling KS3 English level in relation to gender (with girls coded 1).

Figure 5.4.1: Case Processing Summary

		N	Marginal Percentage
Age 14 English NC level	3	1429	10.1%
	4	2276	16.1%
	5	5615	39.6%
	6	3515	24.8%
	7	1331	9.4%
Gender	0 Male	7177	50.7%
	1 Female	6989	49.3%
Valid		14166	100.0%
Missing		1604	
Total		15770	

Figure 5.4.2 shows the *Model fitting information*. Before we start looking at the effects of each explanatory variable in the model, we need to determine whether the model improves our ability to predict the outcome. We do this by comparing a model without any explanatory variables (the baseline or 'Intercept Only' model) against the model with all the explanatory variables (the 'Final' model - this would normally have several explanatory variables but at the moment it just contains gender). We compare the final model against the baseline to see whether it has significantly improved the fit to the data. The *Model fitting Information* table gives the -2 log-likelihood (-2LL, see **Page 4.6**) values for the baseline and the final model, and SPSS performs a chi-square to test the difference between the -2LL for the two models.

Figure 5.4.2: Model Fit

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	500.992			
Final	81.101	419.891	1	.000

The significant chi-square statistic ($p<.0005$) indicates that the Final model gives a significant improvement over the baseline intercept-only model. This tells you that the model gives better predictions than if you just guessed based on the marginal probabilities for the outcome categories.

The next table in the output is the *Goodness-of-Fit table* (**Figure 5.4.3**). This table contains Pearson's chi-square statistic for the model (as well as another chi-square statistic based on the deviance). These statistics are intended to test whether the observed data are consistent with the fitted model. We start from the null hypothesis that the fit is *good*. If we do not reject this hypothesis (i.e. if the p value is large), then you conclude that the data and the model predictions are similar and that you have a good model. However if you reject the assumption

of a good fit, conventionally if *p<.05*, then the model does not fit the data well. The results for our analysis suggest the model does not fit very well (*p<.004).*

Figure 5.4.3: Goodness of fit test

	Chi-Square	df	Sig.
Pearson	13.246	3	.004
Deviance	13.346	3	.004

We need to take care not to be too dogmatic in our application of the *p<.05* rule. For example the chi-square is highly likely to be significant when your sample size is large, as it certainly is with our LSYPE sample of roughly 15,000 cases. In such circumstances we may want to set a lower p-value for rejecting the assumption of a good fit, maybe *p<.01*. More importantly, although the chi-square can be very useful for models with a small number of categorical explanatory variables, they are very sensitive to empty cells. When estimating models with a large number of categorical (nominal or ordinal) predictors or with continuous covariates, there are often many empty cells (as we shall see later). You shouldn't rely on these test statistics with such models. Other methods of indexing the goodness of fit, such as measures of association, like the pseudo R^2, are advised.

In linear regression, R^2 (the coefficient of determination) summarizes the proportion of variance in the outcome that can be accounted for by the explanatory variables, with larger R^2 values indicating that more of the variation in the outcome can be explained up to a maximum of 1 (see **Module 2** and **Module 3**). For logistic and ordinal regression models it not possible to compute the same R^2 statistic as in linear regression so three approximations are computed instead (see **Figure 5.4.4**). You will remember these from **Module 4** as they are the same as those calculated for logistic regression.

Figure 5.4.4: Pseudo R-square Statistics

Cox and Snell	.029
Nagelkerke	.031
McFadden	.010

What constitutes a "good" R^2 value depends upon the nature of the outcome and the explanatory variables. Here, the pseudo R^2 values (e.g. Nagelkerke = 3.1%) indicates that gender explains a relatively small proportion of the variation between students in their attainment. This is just as we would expect because there are numerous student, family and school characteristics that impact on student attainment, many of which will be much more important predictors of attainment than any simple association with gender. The low R^2 indicates that a model containing only gender is likely to be a poor predictor of the outcome

for any particular *individual student.* Note though that this does not negate the fact that there is a statistically significant and relatively large difference in the *average* English level achieved by girls and boys.

The Parameter estimates table (**Figure 5.4.5**) is the core of the output, telling us specifically about the relationship between our explanatory variables and the outcome.

Figure 5.4.5: Parameter Estimates Table

		Estimate	SE	Wald	df	Sig.	95% Confidence Interval Lower	95% Confidence Interval Upper
Threshold	[k3en = 3]	-2.543	.033	5783.690	1	.000	-2.608	-2.477
	[k3en = 4]	-1.376	.026	2823.372	1	.000	-1.427	-1.325
	[k3en = 5]	.354	.023	236.905	1	.000	.309	.399
	[k3en = 6]	1.988	.032	3949.668	1	.000	1.926	2.050
Location	[gender=0]	-.629	.031	414.418	1	.000	-.689	-.568
	[gender=1]	0ª	.	.	0	.	.	.

The threshold coefficients are not usually interpreted individually. They just represent the intercepts, specifically the point (in terms of a logit) where students might be predicted into the higher categories. The labelling may seem strange, but remember the odds of being *level 6 or below* (k3en=6) is just the complement of the odds of being level 7; the odds of being *level 5 or below* (k3en=5) are just the complement of the odds of being *level 6 or above*, and so on. While you do not usually have to interpret these threshold parameters directly we will explain below what is happening here so you understand how the model works. The results of our calculations are shown in **Figure 5.4.6**.

Let's start with girls. Since girls represent our base or reference category the cumulative logits for girls are simply the threshold coefficients printed in the SPSS output (k3en = 3, 4, 5, 6). We take the exponential of the logits to give the *cumulative odds (co)* for girls. Note that these do not match the cumulative logits and odds we showed in **Figure 5.3.3** because, as explained above, SPSS creates these as the odds for achieving each *level or below* as opposed to each *level or above* and because the reference category is boys not girls. However once these logits are converted to cumulative proportions/probabilities you can see they are broadly equivalent in the two tables (bar some small differences arising from the assumption of proportional odds in the ordinal model, more on which later). We calculate the predicted *cumulative probabilities* from the cumulative odds (co) simply by the formula 1/(1+co). If we want to find the predicted probability of being in a specific outcome category (e.g., at a specific English level) we can work out the *category probability* by subtraction. So if the probability of being at level 7 is 0.12 (or 12%), and the probability of being at level 6 or

above is 0.41 (or 41%), then the probability of being specifically at level 6 is .41 - .12 = .29 (or 29%). Similarly the predicted probability for being specifically at Level 5 for girls is .80 - .41 = .39 (39%) and at level 4 it is .93 - .80 = .13 (13%). Finally the probability of being at level 3 is 1 - .93 = .07 (7%).

Figure 5.4.6: Parameters from the ordinal regression of gender on English level.

	English level				
Boys	**3**	**4**	**5**	**6**	**7**
Cumulative logit	-	-1.914	-0.747	0.983	2.617
Cumulative odds [exp(*Cum.logit*)]	-	0.15	0.47	2.67	13.69
Cumulative proportion [1/(1+exp(*Cum.logit*)]	1.00	0.87	0.68	0.27	0.07
Category probability	0.13	0.19	0.41	0.20	0.07
Girls	**3**	**4**	**5**	**6**	**7**
Cumulative logit	-	-2.543	-1.376	0.354	1.988
Cumulative odds [exp(*Cum.logit*)]	-	0.08	0.25	1.42	7.30
Cumulative proportion [1/(1+exp(*Cum.logit*)]	1.00	0.93	0.80	0.41	0.12
Category probability	0.07	0.13	0.39	0.29	0.12
Odds Ratio (Girls/Boys)	**-**	**0.53**	**0.53**	**0.53**	**0.53**
Odds Ratio (Boys/Girls)		**1.88**	**1.88**	**1.88**	**1.88**

To calculate the figures for boys (gender=0) we have to combine the parameters for the thresholds with the gender parameter (-.629, see Figure 5.4.5). Usually in regression we add the coefficient for our explanatory variable to the intercept to obtain the predicted outcome (e.g. $y = a + bx$, see modules 2 & 3). However in SPSS ordinal regression the model is parameterised as $y = a - bx$. This doesn't make any difference to the predicted values, but is done so that positive coefficients tell you that higher values of the explanatory variable are associated with higher outcomes, while negative coefficients tell you that higher values of the explanatory variable are associated with lower outcomes. So for example the cumulative logit for boys at 'level 4+' is -2.543 - (-.629) = -1.914, at level 5+ it is -1.376 - (-.629) = -.747 and so on. Then, just as for girls, the *cumulative odds* (co) are the exponent of the logits, the *cumulative proportions* are calculated as 1/(1+co), and the *category probabilities* are found by subtraction in the same way as described for girls. Phew!

Odds Ratios

We can divide the odds for girls by the odds for boys at each cumulative split to give the OR (see **Figure 5.4.6**). We can see that in the proportional odds model the OR is constant (0.53) at all cumulative splits in the data (the odds of boys achieving a higher level are approximately half the odds for girls). We can express the OR the other way round by dividing the odds for

boys by the odds for girls which gives us the OR of 1.88 (the odds for girls of achieving a higher level are approximately twice the odds for boys). As we saw in **Module 4** these OR of 0.53 and 1.88 are equivalent, they just vary depending on the reference category. *(see* **Extension D** - you can convert an OR to its complement by dividing the OR into 1, e.g. 1/0.53= 1.88, equally 1/1.88=0.53*)*. The *important* thing to note here is that the gender OR is consistent at each of the cumulative splits in the distribution.

The above was completed just to demonstrate the proportional odds principle underlying the ordinal model. In fact we do not have to directly calculate the ORs at each threshold as they are summarised in the parameter for gender. This shows the estimated coefficient for gender is -.629 and we take the exponent of this to find the OR with girls as the base: exp(-.629) = 0.53. To find the complementary OR with boys as the base just reverse the sign of the coefficient before taking the exponent, exp(.629)=1.88. The interpretation of these ORs is as stated above.

Test of parallel lines

Remember that the OR is equal at each threshold because the ordinal model has constrained it to be so through the *proportional odds (PO) assumption.* We can evaluate the appropriateness of this assumption through the 'test of parallel lines'. This test compares the ordinal model which has one set of coefficients for all thresholds (labelled Null Hypothesis), to a model with a separate set of coefficients for each threshold (labelled General). If the general model gives a significantly better fit to the data than the ordinal (proportional odds) model (i.e. if *p<.05*) then we are led to reject the assumption of proportional odds. This is the conclusion we would draw for our example (see **Figure 5.5.7**), given the significant value as shown below (*p<.004*).

Figure 5.4.7: Test of Parallel Lines

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	81.101			
General	67.756	13.346	3	.004

Note: The sharp-eyed among you may have noted that the chi-square statistics given above for the Test of Parallel Lines is exactly the same as that given for the omnibus test of the 'goodness of fit' of the whole model. This is because we have only a single explanatory variable in our model, so the two tests are the same. However when we have multiple explanatory variables this will not be the case.

We can see why this is the case if we compare our OR from the ordinal regression to the separate ORs calculated at each threshold in **Figure 5.3.3**. While the odds for boys are consistently lower than the odds for girls, the OR from the ordinal regression (0.53) underestimates the extent of the gender gap at the very lowest level (Level 4+ OR = 0.45) and

slightly overestimates the actual gap at the highest level (level 7 OR =.56). We see how this results in the significant chi-square statistic in the 'test for parallel lines' if we compare the 'observed' and 'expected' values in the 'cell information' table you requested, shown below as **Figure 5.4.8**. The use of the single OR in the ordinal model leads to predicting fewer boys and more girls at level 3 than is actually the case (shown by comparing the 'expected' numbers from the model against the 'observed' numbers).

Figure 5.4.8: Output for Cell Information

Gender		Age 14 English NC level				
		3	4	5	6	7
Male	Observed	967	1372	2835	1500	503
	Expected	922.293	1384.523	2915.817	1465.918	488.449
	Pearson Residual	1.577	-.375	-1.942	.998	.682
Female	Observed	462	904	2780	2015	828
	Expected	509.559	899.923	2697.705	2040.081	841.733
	Pearson Residual	-2.188	.146	2.022	-.660	-.505

However the test of the proportional odds assumption has been described as anti-conservative, that is it nearly always results in rejection of the proportional odds assumption (O'Connell, 2006, p.29) particularly when the number of explanatory variables is large (Brant, 1990), the sample size is large (Allison, 1999; Clogg & Shihadeh, 1994) or there is a continuous explanatory variable in the model (Allison, 1999). It is advisable to examine the data using a set of separate logistic regression equations to explicitly see how the ORs vary at the different thresholds, as we have done in **Figure 5.3.3**. In this particular case it might be reasonable to conclude that the OR for gender from the PO model (0.53) - while it does underestimate the extent of the over-representation of boys at the lowest level - does not differ hugely from those of the separate logistic regressions (0.45-0.56) and so is a reasonable summary of the trend across the data. Here the statistical test that led to the rejection of the PO assumption probably reflects the large sample size in our LYPSE dataset.

Predicted probabilities

Figure 5.4.6 showed how from the model we can calculate the cumulative proportion at each threshold and, by subtraction, the predicted probability of being at any specific level. However you don't actual have to do any of these calculations to determine the predicted probabilities since we requested SPSS to save the estimated probabilities for each case. We have five possible outcomes (level 3 to level 7) so SPSS will save the predicted probabilities for each case in five new variables that by default will be labelled *EST1_1* to *EST5_1.* The first number refers to the category where 1 will indicate the lowest value for our ordinal outcome (i.e. level 3) and 5 will indicate the highest value (i.e. level 7). The second number after the underscore

(_1) indicates these are the predictions from the first model we have run. If we added some more explanatory variables and ran a second model, without first deleting the variables holding estimated probabilities from the first model, then the predictions from the second model would have the suffix _2, i.e. *EST1_2, EST2_2, EST3_2* etc. If you do intend to run multiple models it may be worth renaming these variables or labelling them carefully so that you do not lose track!

We can use these estimates to explore the predicted probabilities in relation to our explanatory variables. For example we can use the MEANS command (**Analyze>Compare Means>Means**) to report on the estimated probabilities of being at each level for boys and girls. The output is shown below (**Figure 5.4.9**):

Figure 5.4.9: Estimated probabilities for boys and girls from the ordinal regression

Gender		Estimated Cell Probability for Response Category: 3	Estimated Cell Probability for Response Category: 4	Estimated Cell Probability for Response Category: 5	Estimated Cell Probability for Response Category: 6	Estimated Cell Probability for Response Category: 7
Male	Mean	.1285	.1929	.4063	.2043	.0681
	N	7852	7852	7852	7852	7852
	SD	.00000	.00000	.00000	.00000	.00000
Female	Mean	.0729	.1288	.3860	.2919	.1204
	N	7579	7579	7579	7579	7579
	SD	.00000	.00000	.00000	.00000	.00000
Total	Mean	.1012	.1614	.3963	.2473	.0938
	N	15431	15431	15431	15431	15431
	SD	.02780	.03207	.01014	.04382	.02619

Note: the SD is zero in all cells because, with gender being the only explanatory variable in the model, all males will have the same predicted probabilities within each outcome category, and all females will also have the same predicted probabilities within each outcome category.

The ability to summarise and plot these predicted probabilities will be quite useful later on when we have several explanatory variables in our model and want to visualise their associations with the outcome.

Summary

We have seen that where we have an ordinal outcome there is value in trying to summarise the outcome in a single model, rather than completing several separate logistic regression models. However we have also seen that this can overly simplify the data and it is important to complete the separate logistic models to fully understand the nuances in our data. For example, here the ordinal (PO) model did not identify the true extent to which boys were over-

represented relative to girls at the lowest level. We should always complete separate logistic regressions if the assumption of PO is rejected. In the particular example used here it might be reasonable to conclude that the OR for gender from the ordinal (PO) model (0.53) does not differ hugely from those of the separate logistic regressions (0.45-0.56) and so is a reasonable summary of the trend across the data. However you are only in a position to conclude this if you have completed the separate logistic models, so in practice our advice is always to do the separate logistic models when the PO assumption is formally rejected. Given the anti-conservative nature of the test of the proportional odds assumption (O'Connell, 2006) this will more often than not be the case. Let us now move on to consider models which have more than one explanatory variable.

5.5 An extended example - Teacher expectations and tiering

We will now complete a more complex ordinal regression employing multiple explanatory variables. For this example we are going to use a different outcome variable, the mathematics test tier that teachers entered the student for. We give the background to this example below:

Tiering structure in national tests at age 14

As we said earlier, at the time they were recruited to LSYPE all students in England completed national tests in English, mathematics and science at the end of Year 9 (age 14). These tests awarded pupils a National Curriculum (NC) level with the 'typical' student at age 14 expected to achieve Level 5 or Level 6. The range of levels that can be awarded at age 14 range from 3-7, although a Level 8 can be awarded specifically in the mathematics test. In some subjects there are different tests to reflect different levels of ability. These different tests are organized into 'tiers'.

There are no tiering arrangements for national tests in English at age 14 - all students sit the same test papers. In Science there are only two tiers (a lower tier covering levels 3-6 and a higher tier covering levels 5-7). However for mathematics there are four test tiers covering levels 3-5, 4-6, 5-7 and 6-8 respectively. Importantly each tier has *different* test papers targeted at a restricted set of levels. This is summarised in **Figure 5.5.1**. For example, the highest outcome (Level 8) can only be achieved if the student is entered by their teacher for the highest (6-8) tier. Equally Level 7 can only be achieved if the student is entered for either of the two highest tiers (tier 5-7 or tier 6-8). Importantly which tier a student is entered for is a matter for the *professional judgment of the teacher*. This is quite a 'high-stakes' judgment since an unclassified (U) grade results if a student entered for a higher tier fails to achieve the expected level. Equally, placing the student in too low a tier can prevent them from getting the maximum grade they may be capable of. This judgement may well be influenced by other factors determining the teacher's perceptions of how students will cope with the demands made on them by the content and structure of the tests.

Table 5.5.1: Tiering structure for age 14 national tests in mathematics

Subject	Tier	Awardable NC levels					
Mathematics	Tier 3-5	U	3	4	5		
	Tier 4-6	U	3	4	5	6	
	Tier 5-7	U		4	5	6	7
	Tier 6-8	U			5	6	7	8

Notes: The numbers indicate National Curriculum (NC) levels that can be achieved through the relevant tier. U = unclassified result.

The essential point is that the outcome a student can achieve in the mathematics test will be constrained by the tier their teacher has entered them for, and there are social processes that may influence this decision. The research questions we want to address are:

- Are students from different ethnic, socio-economic and gender groups equally likely to be entered for the higher tiers by their teachers?
- If there are variation between ethnic, social and gender groups in the pattern of entry to different tiers, can these patterns be explained by other factors such as the prior attainment of the students?

Descriptive statistics

Figure 5.5.2 below shows the proportion of LSYPE students entered for each mathematics test tier. The table was generated by requesting FREQUENCIES (**Analyze>Descriptive Statistics>Frequencies**) for the variable *k3matier* and using the 'Format' button to request it to be 'Ordered by' descending values.

Figure 5.5.2: Frequency distribution in entry to different mathematics test tiers

		Freq	Percent	Valid Percent	Cumulative Percent
Valid	3 tier 6-8	2358	15.0	16.3	16.3
	2 tier 5-7	4033	25.6	27.9	44.3
	1 tier 4-6	4867	30.9	33.7	78.0
	0 tier 3-5	3181	20.2	22.0	100.0
	Total	14439	91.6	100.0	
Missing	System	1331	8.4		
Total		15770	100.0		

Overall 16.3% of students are entered for tier 6-8, 27.9% for tier 5-7, 30.9% for tier 4-6 and 22.0% for tier 3-5. As shorthand, we shall call these *top, upper, lower* and *bottom* tiers respectively. Cumulatively 16.3% are entered for the top tier, 44.3% for the upper tier *or above* and 78% for the lower tier *or above.* What we want to establish in the ordinal regression is whether there is any systematic relationship between our explanatory variables (SEC, ethnic group and gender) and the probability of being entered to a higher tier.

So do tier entry rates vary by our equity factors? The first step, as ever, is just some simple bivariate analyses (crosstabs) to see what patterns exist in the data (**Figure 5.5.3, Figure 5.5.4 and Figure 5.5.5**). The data show a strong association between ethnic group and tier of entry. Students of Mixed Heritage and Any Other ethnic group do not differ greatly from White

British, and Indian students appear more likely than White British students to be entered for the higher tiers. However, while 47% of White British students are entered for the upper and top tiers, only 35% of Bangladeshi and Black African students, 33% of Pakistani students and 27% of Black Caribbean students are entered for these two tiers.

Figure 5.5.3: Tier entry by ethnic group

| | | Age 14 maths tier | | | |
		tier 3-5	tier 4-6	tier 5-7	tier 6-8
Ethnic group	White British	20.1%	32.7%	29.5%	17.6%
	Mixed heritage	18.9%	37.6%	26.5%	17.0%
	Indian	13.9%	33.9%	30.4%	21.8%
	Pakistani	34.0%	33.0%	23.3%	9.8%
	Bangladeshi	28.0%	36.4%	24.8%	10.8%
	Black Caribbean	32.5%	40.4%	20.7%	6.4%
	Black African	28.6%	35.7%	24.1%	11.6%
	Any other group	22.1%	34.0%	23.9%	20.0%
Total		21.8%	33.7%	28.0%	16.5%

There is also a strong association with the SEC of the home. Thus while 69% of students from higher managerial & professional homes are entered for the upper and top tiers, only 24% of children from the lowest SEC homes are entered.

Figure 5.5.4: Tier entry by socio-economic classification (SEC)

| | | Age 14 maths tier | | | |
		tier 3-5	tier 4-6	tier 5-7	tier 6-8
Parent socio-economic class (SEC)	Higher managerial & professional	7.0%	23.6%	31.9%	37.5%
	Lower managerial & professional	11.5%	30.6%	34.2%	23.6%
	Intermediate occupations	16.9%	37.6%	29.7%	15.8%
	Small employers & own account	20.4%	34.0%	29.9%	15.8%
	Lower supervisory & technical	23.5%	37.2%	28.7%	10.6%
	Semi-routine occupations	26.8%	38.3%	25.4%	9.4%
	Routine occupations	34.0%	36.8%	22.4%	6.8%
	Never worked/long term unemployed	40.2%	36.0%	16.7%	7.1%
Total		20.8%	33.7%	28.5%	17.0%

The gender difference is small relative to the SEC and ethnic group differences. The pattern of gender entry seems to vary by tier, so boys are both more likely than girls to be entered for the lowest tier (22.9% vs 20.7%) and also more likely to be entered for the highest tier (17.3% vs. 15.6%).

Figure 5.5.5: Tier entry by gender.

		Age 14 maths tier			
		tier 3-5	tier 4-6	tier 5-7	tier 6-8
Gender	Male	22.9%	32.6%	27.2%	17.3%
	Female	20.7%	34.9%	28.8%	15.6%
Total		21.8%	33.7%	28.0%	16.4%

What we do not know from these simple bivariate relationships is how our variables interrelate when considered together, to do this we need to build a model. Our primary focus in the analysis is to explore the ethnic group differences in tier entry rates, to try and understand why ethnic minority students are so much less likely to be entered for the higher tiers, and to see whether the association seen above can be explained by other variables in the dataset.

5.6 Running an ordinal regression for mathematics tier of entry

Data preparation

Remember that we said that SPSS ordinal would automatically make the reference category for a nominal explanatory variable the *last* category. We want the reference group for ethnicity to be White British as this is the majority ethnic group, but White British students are currently coded 0 in the variable *ethnic*. We therefore need to RECODE (we discuss recoding variables in the **Foundation Module**) the *ethnic* variable so that White British are coded 8, and hence the last category. If you do this and call your new variable *ethnic2* then a frequency table should give you the output shown below in **Figure 5.6.1**.

Figure 5.6.1: Frequency Table for recoded Ethnicity Variable

		Freq	Percent	Valid Percent	Cumulative Percent
Valid	Mixed heritage	795	5.0	5.2	5.2
	Indian	1013	6.4	6.6	11.7
	Pakistani	940	6.0	6.1	17.8
	Bangladeshi	722	4.6	4.7	22.5
	Black Caribbean	576	3.7	3.7	26.2
	Black African	613	3.9	4.0	30.2
	Any other ethnic group	654	4.1	4.2	34.5
	White British	10103	64.1	65.5	100.0
	Total	15416	97.8	100.0	
Missing	-9	354	2.2		
Total		15770	100.0		

We are also going to take advantage of the fine-grained hierarchical nature of the SEC variable to treat this as a continuous variable. It is helpful in all forms of regression if the zero value of a continuous explanatory variable (or covariate) has an interpretable meaning, and this is particularly the case in logistic and ordinal regression. We have therefore recoded SEC so zero is a meaningful category (the highest level of SEC, higher managerial and professional homes) and have called this recoded variable *Sec2*. (If you are unsure how to do this we explain the process in the text surrounding **Figure 5.7.4**) The frequencies for *Sec2* are displayed in **Figure 5.6.2**.

Figure 5.6.2: Frequency Table for recoded SEC Variable

		Freq	Percent	Valid Percent	Cumulative Percent
Valid	0 Higher Managerial and professional occupations	1567	9.9	12.2	12.2
	1 Lower managerial and professional occupations	3083	19.5	24.0	36.2
	2 Intermediate occupations	932	5.9	7.3	43.5
	3 Small employers and own account workers	1672	10.6	13.0	56.5
	4 Lower supervisory and technical occupations	1454	9.2	11.3	67.9
	5 Semi-routine occupations	1637	10.4	12.8	80.6
	6 Routine occupations	1409	8.9	11.0	91.6
	7 Never worked/long term unemployed	1075	6.8	8.4	100.0
	Total	12829	81.4	100.0	
Missing	System	2941	18.6		
Total		15770	100.0		

Specifying the ordinal regression

We request an ordinal regression as we did earlier, through **Analyse > Regression > Ordinal**. We enter *k3matier* as our Dependent variable. Categorical (nominal or ordinal) explanatory variables are entered to the Factor(s) box, so this is where we enter *ethnic2* and *gender*. Continuous explanatory variables (in this case *sec2*) are entered as covariates.
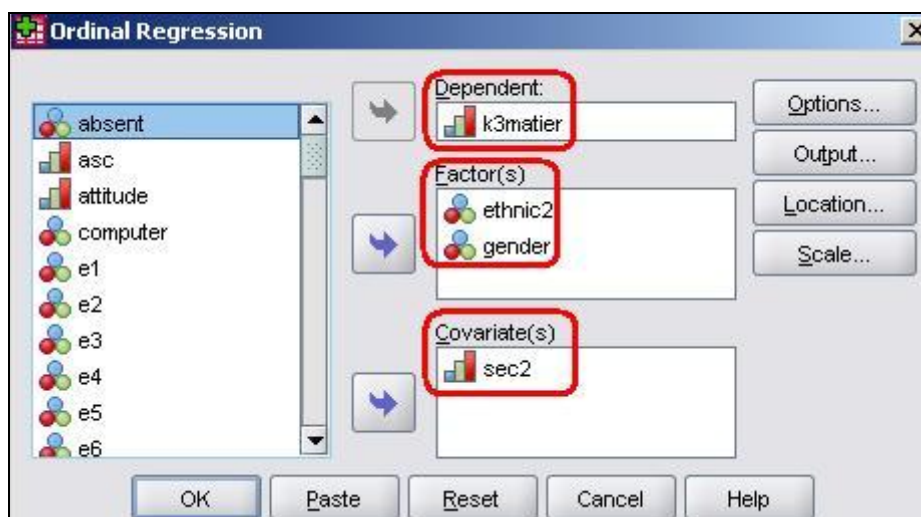
Interpreting the results

The case processing summary is a reminder of the variables we have entered (**Figure 5.6.3**), note though that it does not display any covariates in the model. Note that SPSS also issues

a warning that there are "20 (3.9%) cells with zero frequencies". This just means that even with 11,601 cases there are some combinations of variables that are not represented (for example there are no Bangladeshi, boys, from higher managerial and professional homes entered for the bottom tier). This does not mean we cannot complete an analysis; there may be sound empirical reasons for low counts in some cells. However we should consider the proportion of empty cells when we come to evaluating the goodness-of-fit test.

Figure 5.6.3: Case Processing Summary

		N	Marginal Percentage
Age 14 maths tier	tier 3-5	2393	20.6%
	tier 4-6	3910	33.7%
	tier 5-7	3317	28.6%
	tier 6-8	1981	17.1%
Ethnic recoded (WBRI=8)	Mixed heritage	612	5.3%
	Indian	764	6.6%
	Pakistani	686	5.9%
	Bangladeshi	474	4.1%
	Black Caribbean	445	3.8%
	Black African	450	3.9%
	Any other ethnic group	446	3.8%
	White British	7724	66.6%
Gender	Male	5946	51.3%
	Female	5655	48.7%
Valid		11601	100.0%
Missing		4169	
Total		15770	

Let's look at the *Parameter Estimates table* (**Figure 5.6.4**). The model confirms there are systematic effects in tier of entry related to SEC and to ethnic group. In relation to *Sec2* the coefficient is -.278. We take the exponent of this coefficient to get the OR: exp(-.278)=.78, which indicates that odds of being entered for a higher tier decrease by 0.78 for each unit increase in *Sec2* score (remember higher values of SEC indicate lower SEC homes). Conversely we could say the odds of being entered to a higher tier increase by (1/0.78)=1.33 for ever unit change in *Sec2* score. To calculate the predicted OR at any point on the SEC scale we can multiply the coefficient by the relevant *Sec2* score and take the exponent of the result. For example the OR for students from the lowest social class category (7=long term unemployed) is exp(-0.78 * 7) = 0.143. Students from homes where the main parent is long term unemployed are only .14 as likely to be entered for a higher tier compared to the odds for a student from the higher managerial and professional home. To express this as the OR in favour of high SEC (1/.143=7.00) indicating students from the highest SEC are 7 times more likely than students from the lowest SEC to be entered for a higher tier.

Figure 5.6.4: Parameter Estimates for the Model

		Estimate	SE	Wald	df	Sig.	95% CI Lower	95% CI Upper
Threshold	[k3matier = 0]	-2.388	.043	3136.466	1	.000	-2.472	-2.305
	[k3matier = 1]	-.729	.037	397.985	1	.000	-.800	-.657
	[k3matier = 2]	.790	.038	443.448	1	.000	.717	.864
Location	sec2	-.278	.008	1142.832	1	.000	-.294	-.262
	[ethnic2=1]	-.073	.076	.902	1	.342	-.222	.077
	[ethnic2=2]	.410	.069	35.089	1	.000	.274	.545
	[ethnic2=3]	-.218	.074	8.740	1	.003	-.363	-.074
	[ethnic2=4]	.104	.089	1.385	1	.239	-.069	.278
	[ethnic2=5]	-.853	.090	89.036	1	.000	-1.030	-.676
	[ethnic2=6]	-.351	.089	15.432	1	.000	-.526	-.176
	[ethnic2=7]	.088	.089	.978	1	.323	-.086	.262
	[ethnic2=8]	0[a]	.	.	0	.	.	.
	[gender=0]	-.039	.034	1.362	1	.243	-.106	.027
	[gender=1]	0[a]	.	.	0	.	.	.

There is also a strong association between ethnic group and tier entry, even after SEC has been controlled. We can see significant and negative coefficients for Black Caribbean, Black African and Pakistani students. Again taking the exponent of the logits will give the OR, so for example the odds of Black Caribbean (ethnic=5) students being entered for a higher tier are exp(-.853)= 0.43 the odds for White British students, or less than half. Conversely there is a significant positive coefficient for Indian students (.410) indicating they are exp(.410)=1.51 times more likely than White British students to be entered for a higher tier, even after controlling for SEC.

Evaluating the model and the assumption of proportional odds

Of course we should not take the above at face value without examining the adequacy of the model and particularly the PO assumption. Looking at the model fit (**Figure 5.6.5**) we can see a highly significant reduction in the chi-square statistics (*p<.005*) so the model is clearly a significant improvement over the baseline or intercept only model. The Nagelkerke R^2 indicates the model can account for 12.4% of the variance in tier of entry.

Figure 5.6.5: Model-fit Data

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	3244.175			
Final	1819.460	1424.716	9	.000

However the goodness-of-fit statistics (**Figure 5.6.6**) suggest the model does not fit the data well.

Figure 5.6.6: Goodness of Fit for Model

	Chi-Square	df	Sig.
Pearson	515.591	372	.000
Deviance	523.350	372	.000

And the test of parallel lines (**Figure 5.6.7**) also rejects the null hypothesis of the assumption of PO.

Figure 5.6.7: Test of Parallel Lines

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	1819.460			
General	1744.167	75.292	18	.000

These statistics suggest the model does not fit the data well. I said earlier that these statistics can be unreliable in certain circumstances. For example the large sample size here means that even very small departures from the PO assumption may be found to be statistically significant. We will explore what to do in these circumstances in more detail later (section 5.9). But first it might be prudent to test whether this is due to the absence of any interaction terms in the model. We saw previously that there were significant interactions between ethnic group and SEC in their impact on attainment. As we have seen before in earlier modules, it is important to test for interaction effects, particularly where these are suggested by theory or prior research evidence. We should include the interaction terms and will show you how to do so on the next page.

# 5.7 Evaluating interaction effects in ordinal regression

Requesting a model with interaction terms

First ask for an ordinal regression through selecting **Analyse>Regression>Ordinal** as we did on **Page 5.6**. To specify interaction terms in SPSS ordinal we use the 'Location' submenu, so click on the 'Location' button. What we want to do now is specify not a 'Main Effects' but a 'Custom' model, so place a tick in the 'Custom' button as shown in the image below.
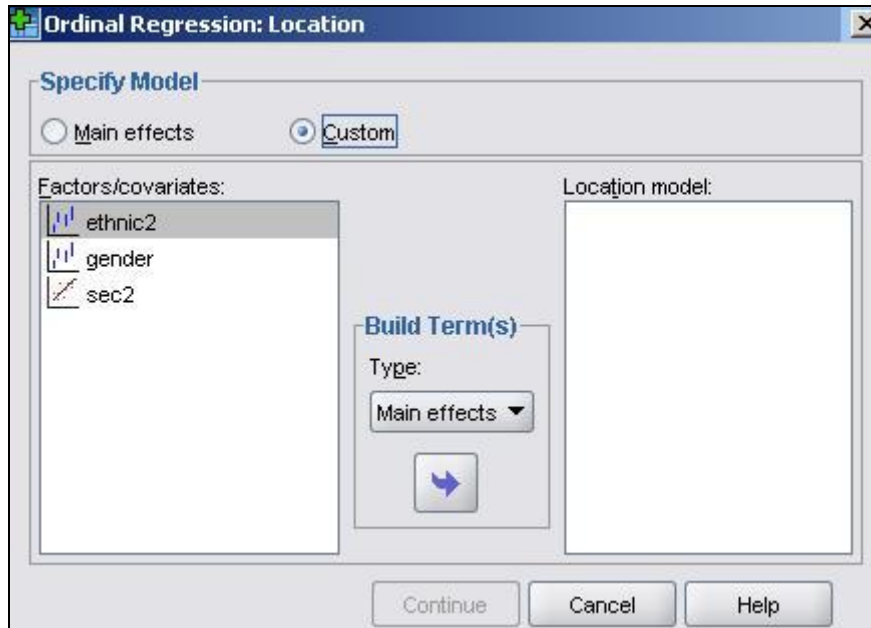
We now need to build our model. The logic of the approach to testing interactions is as we have described earlier in **Module 3** (linear regression) and **Module 4** (logistic regression). We start by specifying a full model that includes all the main effects, plus all 2-way interactions plus the three-way interaction. We then run this model and if the highest order interaction (here the 3-way interaction between ethnic2 * Sec2 * gender) is not significant it can be removed. We then run the model with the main effects and all the 2-way interactions, again subsequently eliminating any non-significant 2-way interaction terms.

To build this full model we hold down the CTRL key and click on *ethnic2*, *gender* and *Sec2* in the *'Factors/covariates'* box so all three variables are highlighted, then in the 'Build terms' box click 'main effects', and then drag (or click on the arrow) to move these to the 'Location model' box. Then do the same but click on 'All 2-way' in the 'build terms' box, and lastly again with 'All 3-way' in the build terms box as shown below.
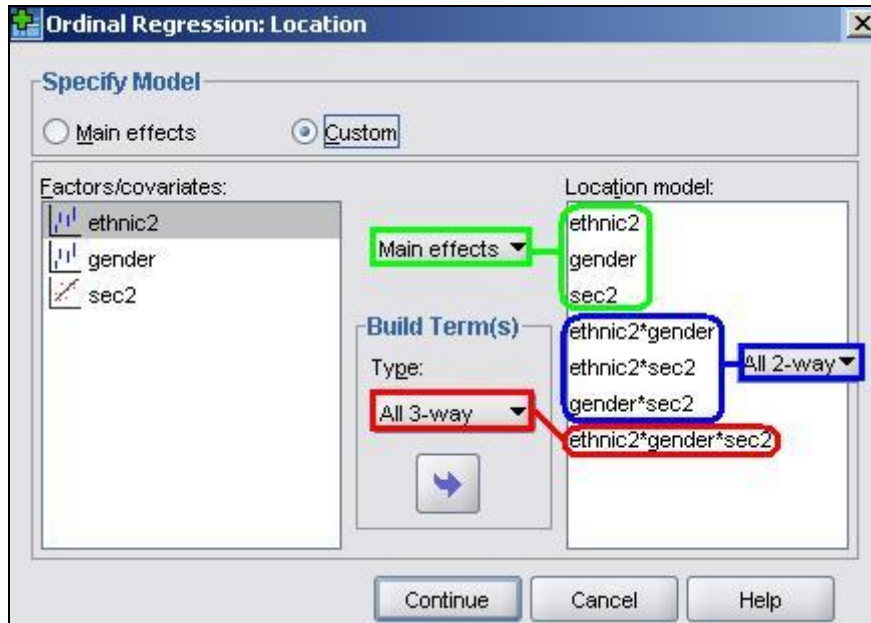
We have decided not to bombard you with the output but running this analysis on the data indicated no statistically significant 3-way interaction so this term was removed. Running a subsequent model with all 2-way interactions revealed no significant 2-way interactions between *ethnic2*gender* or *Sec2*gender* but a highly significant *ethnic2*Sec2* interaction. Rather than repeating this analysis here, we will just request the model including the main effects and the significant *ethnic2*sec2* interaction. You can remove the *ethnic*gender*sec*, *gender*sec2* and *ethnic2*gender* terms by clicking on them and dragging (or using the reverse arrow) to move them back to the Factors/covariates box, leaving just the main effects and the *ethnic2*sec2* interaction in the *Location* model. This is what the final version of the *Location* submenu will look like:
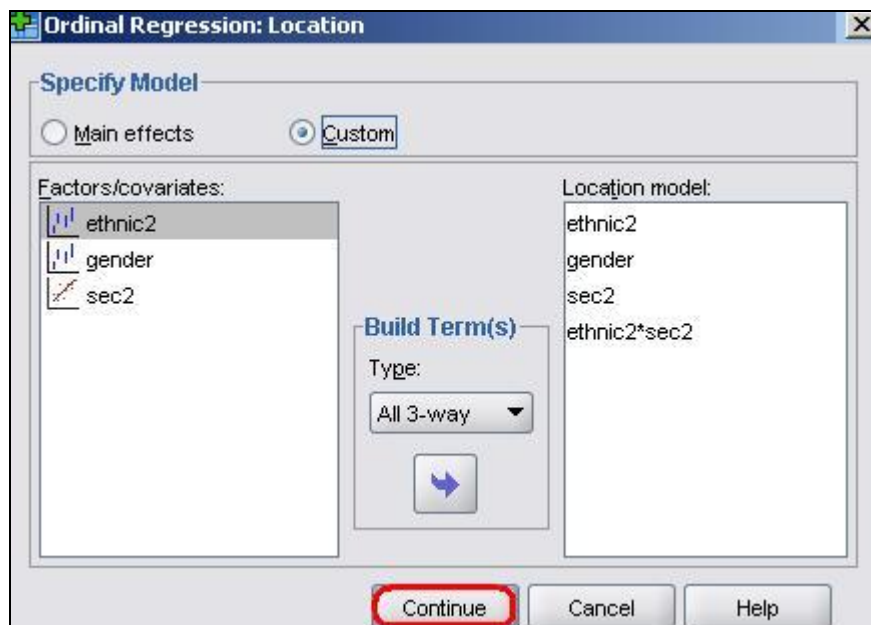
Click 'Continue' and then **OK** to run the model.

Interpreting the results with interactions

Examining the linear part of the model (the logits)

The parameter estimates are shown below (**Figure 5.7.1**). We can see that the interaction terms are all highly significant, particularly so for Indian, Pakistani and Bangladeshi students.

Figure 5.7.1: Parameter Estimates for Model with Interaction Terms

		Estimate	SE	Wald	df	Sig.	95% CI Lower	95% CI Upper
Threshold	[k3matier = 0]	-2.483	.046	2899.658	1	.000	-2.573	-2.392
	[k3matier = 1]	-.821	.040	413.555	1	.000	-.900	-.742
	[k3matier = 2]	.703	.041	297.571	1	.000	.623	.783
Location	[ethnic2=1]	-.270	.125	4.643	1	.031	-.515	-.024
	[ethnic2=2]	.138	.130	1.127	1	.288	-.117	.393
	[ethnic2=3]	-.657	.157	17.585	1	.000	-.964	-.350
	[ethnic2=4]	-.790	.251	9.930	1	.002	-1.281	-.299
	[ethnic2=5]	-1.138	.153	55.543	1	.000	-1.438	-.839
	[ethnic2=6]	-.612	.156	15.357	1	.000	-.917	-.306
	[ethnic2=7]	-.043	.152	.079	1	.779	-.341	.256
	[ethnic2=8]	0ª	.	.	0	.	.	.
	[gender=0]	-.036	.034	1.152	1	.283	-.103	.030
	[gender=1]	0ª	.	.	0	.	.	.
	sec2	-.311	.010	927.255	1	.000	-.331	-.291
	[ethnic2=1] * sec2	.068	.034	3.972	1	.046	.001	.135
	[ethnic2=2] * sec2	.085	.033	6.744	1	.009	.021	.149
	[ethnic2=3] * sec2	.112	.033	11.678	1	.001	.048	.177
	[ethnic2=4] * sec2	.184	.045	16.495	1	.000	.095	.273
	[ethnic2=5] * sec2	.099	.042	5.610	1	.018	.017	.181
	[ethnic2=6] * sec2	.077	.035	4.947	1	.026	.009	.145
	[ethnic2=7] * sec2	.044	.037	1.428	1	.232	-.028	.116
	[ethnic2=8] * sec2	0ª	.	.	0	.	.	.

The significant interaction terms indicates the slope of the assumed linear relationship between SEC and entry to a higher tier varies significantly between ethnic groups. Just as we did when looking at interactions on **Page 3.11** and **Page 4.13** we can understand the interaction best by calculating the predicted values for each ethnic group by SEC combination and plotting these graphically. If we work with the linear part of the model (the logits) this will most clearly display the form of the interaction.

Just as with linear regression, we can think of the line representing the relationship between SEC and the predicted logit for entry to higher tiers as having the formula:

$Y = a + b_1x_1 + b_2x_2 + b_3x_3$ ...etc.

Because the effect of gender is a constant (it does not interact significantly with either *ethnic2* or *Sec2*) then we only need to be concerned with three parameters: a= the intercept (the coefficients for ethnic group when *Sec2*=0); b_1x_1 representing the coefficient and value respectively for *Sec2*, and b_2x_2 represents the coefficient and values respectively for each of the *ethnic2*Sec2* interactions.

The intercepts are just the ethnic group coefficients when *Sec2*=0. Because White British are the reference group for ethnicity their logit is represented by zero and the coefficients for each ethnic group are contrasts against White British. These values can be read directly from the SPSS output and are highlighted in red on **Figure 5.7.2**.

We then need to calculate the change in the logit for different levels of *Sec2*. The printed value of *Sec2* in the SPSS output (-.311) is the unit change in logits associated with a one unit increase in the value of *Sec2* for the reference group, i.e. White British students. So to calculate the predicted logits at each level of *Sec2* for White British students we simply multiply -.311 by the respective value of *Sec2*. So for White British students from SEC=5 (semi-routine homes) the predicted logit is: 0 + (-.311 * 5) = -1.55. These predicted values are highlighted by the blue box on **Figure 5.7.2**.

For ethnic minority students the slope for *Sec2* is moderated by the *ethnic2*sec2* interaction. For example for Black Caribbean students from SEC=5 (semi-routine homes) the predicted logit is: -1.14 + (-.311 * 5) + (.099 * 5) = -2.20. It gives exactly the same result, but is slightly computational easier, to calculate the unit change in logits for each unit change in SEC for Black Caribbean (-.311 + .099 = -.211). We can then just multiply -.211 by the respective values of *Sec2* and do this for all levels of *Sec2*.

We follow the same process as above for each minority ethnic group. These predicted values are shown in the rest of **Figure 5.7.2**. We have used EXCEL to calculate the predicted logits shown in **Figure 5.7.2** and to plot the relationship graphically **Figure 5.7.3**.

Figure 5.7.2: Predicted logits for each ethnic group and SEC combination from the *ethnic2*sec2* interaction model

Ethnic group	High 0	1	2	SEC2 3	4	5	6	Low 7
White British	0.00	-0.31	-0.62	-0.93	-1.24	-1.55	-1.86	-2.17
Mixed	-0.27	-0.51	-0.75	-1.00	-1.24	-1.48	-1.72	-1.97
Indian	0.14	-0.09	-0.31	-0.54	-0.77	-0.99	-1.22	-1.44
Pakistani	-0.66	-0.86	-1.05	-1.25	-1.45	-1.65	-1.85	-2.04
Bangladeshi	-0.79	-0.92	-1.04	-1.17	-1.30	-1.42	-1.55	-1.68
Black Caribbean	-1.14	-1.35	-1.56	-1.77	-1.98	-2.20	-2.41	-2.62
Black African	-0.61	-0.85	-1.08	-1.31	-1.55	-1.78	-2.01	-2.25
Any Other group	-0.04	-0.31	-0.58	-0.84	-1.11	-1.38	-1.64	-1.91

Figure 5.7.3: Predicted logits for each ethnic group and SEC combination from the ethnic2*sec2 interaction model
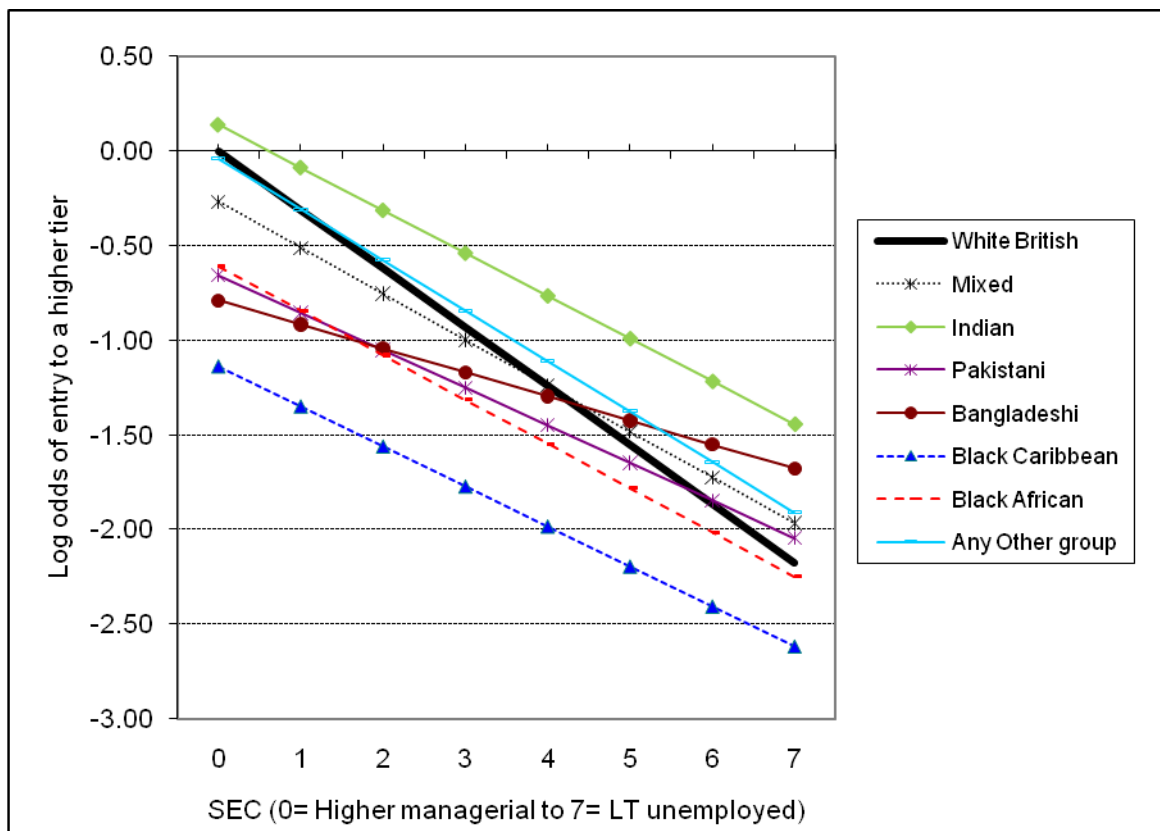
Examining the ORs

The above figure shows the relationship in terms of log odds (logits). What does this mean in terms of ORs? The OR for each ethnic group when *sec2*=0 (higher managerial and

professional occupations) can be found directly by exponentiating the ethnic coefficients. For example the OR for Black Caribbean students is exp(-1.138)= 0.32, so for Black Caribbean students from homes in the highest SEC category the odds of being entered for a higher tier are only one third (0.32) the odds for White British students, or conversely the odds for White British students from the highest SEC category being entered for a higher tier are over three times (1/.32=3.12) the odds for Black Caribbean student being entered.

What about the ethnic group ORs at other levels of SEC? Exponentiating the values in the body of **Figure 5.7.2** will give the ORs for each SEC and ethnic combination relative to the overall reference group (White British students from higher managerial & professional homes). This is fine, but what we really want are the ORs that compare each minority group to their White British peers *within* each SEC category.

The easiest way to calculate the ethnic group ORs at different values of SEC is simply to **RECODE SEC to *a new variable (Sec2) where the zero value represents the reference category of interest.*** This means the coefficients (and the associated standard errors and p values) for each ethnic group will give the contrast with White British students for the reference SEC category. This can easily be done by using the COMPUTE command to recode the original SEC variable. For example we used the following syntax to create the new variable *sec2* taking one away from the value of SEC for each case.

- *COMPUTE Sec2= SEC - 1.*

We also adjusted the 'Missing' setting in the *Variable View* such that '-1' (and not 0) was the missing value. This gives *Sec2* the values we have used in the above analysis where higher managerial and professional homes is the reference category (0). You can see the recoded values for *sec2* in the column labelled 'Sec-1' of **Figure 5.7.4** below.

Figure 5.7.4: Varying the reference category for SEC using the COMPUTE command

| SEC label | Sec | Values of *Sec2* after the COMPUTE command | | | | | | | |
		Sec-1	Sec-2	Sec-3	Sec-4	Sec-5	Sec-6	Sec-7	Sec-8
Higher managerial & prof.	1	0	-1	-2	-3	-4	-5	-6	-7
Lower managerial & prof.	2	1	0	-1	-2	-3	-4	-5	-6
Intermediate	3	2	1	0	-1	-2	-3	-4	-5
Small employers/SE	4	3	2	1	0	-1	-2	-3	-4
Lower supervisory & tech.	5	4	3	2	1	0	-1	-2	-3
Semi-routine	6	5	4	3	2	1	0	-1	-2
Routine	7	6	5	4	3	2	1	0	-1
Long term unemployed	8	7	6	5	4	3	2	1	0

Suppose we want long term unemployed to be the reference category for SEC. We can do this by subtracting 8 from every value of SEC to create a temporary SEC variable (SECtemp). This is the SPSS syntax:

- *COMPUTE SECtemp= SEC - 8.*

This gives *SECtemp* the values in the last column of **Figure 5.7.4** labelled 'Sec-8'. Now long term unemployed are the reference (0) category. If you run the regression again with this new coding of SEC (remembering to change the 'Missing' setting to '-9' and the 'measure' column to *ordinal*) you will get the regression output shown below (**Figure 5.7.5**).

Figure 5.7.5: Parameter estimates with long term unemployed as the reference group

		Estimate	SE	Wald	df	Sig.	95% CI Lower	95% CI Upper
Threshold	[k3matier = 0]	-.308	.050	37.382	1	.000	-.406	-.209
	[k3matier = 1]	1.354	.052	685.540	1	.000	1.253	1.455
	[k3matier = 2]	2.878	.057	2549.408	1	.000	2.766	2.990
Location	[ethnic2=1]	.208	.160	1.691	1	.193	-.105	.520
	[ethnic2=2]	.732	.137	28.416	1	.000	.463	1.001
	[ethnic2=3]	.131	.118	1.215	1	.270	-.102	.363
	[ethnic2=4]	.498	.122	16.686	1	.000	.259	.738
	[ethnic2=5]	-.444	.192	5.326	1	.021	-.820	-.067
	[ethnic2=6]	-.074	.145	.259	1	.611	-.357	.210
	[ethnic2=7]	.265	.160	2.727	1	.099	-.050	.579
	[ethnic2=8]	0ª	.	.	0	.	.	.
	[gender=0]	-.036	.034	1.152	1	.283	-.103	.030
	[gender=1]	0ª	.	.	0	.	.	.
	SECtemp	-.311	.010	927.255	1	.000	-.331	-.291
	[ethnic2=1] * SECtemp	.068	.034	3.972	1	.046	.001	.135
	[ethnic2=2] * SECtemp	.085	.033	6.744	1	.009	.021	.149
	[ethnic2=3] * SECtemp	.112	.033	11.678	1	.001	.048	.177
	[ethnic2=4] * SECtemp	.184	.045	16.495	1	.000	.095	.273
	[ethnic2=5] * SECtemp	.099	.042	5.610	1	.018	.017	.181
	[ethnic2=6] * SECtemp	.077	.035	4.947	1	.026	.009	.145
	[ethnic2=7] * SECtemp	.044	.037	1.428	1	.232	-.028	.116
	[ethnic2=8] * SECtemp	0ª	.	.	0	.	.	.

Now the coefficients (and standard errors and p values) for each ethnic group represent the contrasts when the SEC reference category is long term unemployed. As you would expect, these are substantially different for some ethnic groups from the ORs among students from higher managerial & professional homes. For example among the lowest SEC homes Bangladeshi pupils are significantly *more* likely than White British pupils to be entered for higher tiers (OR= exp(.498)= 1.65, *p<.005*). Black Caribbean students are still less likely than White British to be entered for a higher tier (OR = exp(-.444) = 0.64, *p<.02*), but this is a smaller degree of under-representation than we saw among students from the highest SEC homes where the OR was 0.32.

Evaluating model fit

The interactions of *ethnic2*sec2* are highly statistically significant, and, as we can see from **Figure 5.7.6**, including them has increased the Nagelkerke R^2 from 12.4% to 12.7%. However the goodness-of-fit test still indicates a less than adequate fit, and the test of parallel lines still formally rejects the proportional odds assumption. We therefore need to consider further refinements to our model.

Figure 5.7.6: Statistics for Evaluating the Model

Pseudo R-Square

Cox and Snell	.118
Nagelkerke	.127
McFadden	.047

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	474.270	365	.000
Deviance	486.433	365	.000

Test of Parallel Lines[a]

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	1782.542			
General	1689.383	93.159	32	.000

5.8 Including a control for prior attainment

A key variable so far missing from our model is information on students' prior attainment. This is important since we would expect that students with lower attainment would be entered for the lower tiers and those with higher attainment would be entered for higher tiers. Fortunately we have a measure of students' prior attainment in the form of their average score in national tests at age 11, so we can add this to the model. As we said earlier, to aid the interpretation of a continuous explanatory variable we need to make the zero value meaningful. For this reason the (by now quite familiar!) standardised measure of attainment at age 11 (*ks2stand)* is most useful because zero indicates the mean age 11 score, so our model coefficients will be calculated for students of average prior attainment. We also know the SD for *ks2stand* is 10, which is a nice round number should we want to calculate the coefficient for students 1 SD below or 1 SD above the mean. So now run the previous model (with *Ethnic2*, *gender* and *sec2*) but also add *ks2stand.* This should be added to the *covariates* box because it is a continuous explanatory variable.

Goodness of fit and Pseudo-R^2

Let's take a look at **Figure 5.8.1**. The goodness of fit test is now no longer significant, which suggests the model fits the data well. However as we said earlier we should probably not pay too much attention to this because with a continuous predictor (age 11 standard score) there are many empty cells. In fact whereas before we had 20 (3.9%) empty cells now SPSS issues a warning message telling us we have 7086 (62.9%) empty cells. This warning does not mean we cannot proceed with the analysis, but it does mean we should be wary of the accuracy of the goodness-of-fit test.

Figure 5.8.1: Statistics for Evaluating the Prior Attainment Model

Goodness-of-Fit					Pseudo R-Square	
	Chi-Square	df	Sig.		Cox and Snell	.652
Pearson	7725.867	8425	1.000		Nagelkerke	.699
Deviance	5234.206	8425	1.000		McFadden	.391

We can inspect the Pseudo- R^2 to get an alternative indication of the predictive power of the model. Our Pseudo-R^2 value has increased hugely (from 12.7% to 69.9%) so the model is much better able to predict the outcome for individual cases. This improvement in R^2 is in line with that we have seen previously when we add prior attainment to our prediction of achievement (see Module 3.13).

Parameter estimates table

Looking at the parameter estimates table below (**Figure 5.8.2**), none of the terms for the interaction between *ethnic2*sec2* remain statistically significant in our new model, so the interaction terms can be dropped to simplify the model.

Figure 5.8.2: Parameter Estimates for Prior Attainment Model with interaction term

		Estimate	SE	Wald	df	Sig.	95% CI Lower	95% CI Upper
Threshold	[k3matier = 0]	-2.956	.060	2455.233	1	.000	-3.073	-2.839
	[k3matier = 1]	.284	.050	31.622	1	.000	.185	.383
	[k3matier = 2]	3.206	.061	2757.341	1	.000	3.086	3.325
Location	[ethnic2=1]	-.095	.151	.396	1	.529	-.391	.201
	[ethnic2=2]	.645	.155	17.266	1	.000	.341	.949
	[ethnic2=3]	.472	.190	6.163	1	.013	.099	.844
	[ethnic2=4]	.252	.301	.702	1	.402	-.338	.843
	[ethnic2=5]	-.512	.180	8.107	1	.004	-.864	-.159
	[ethnic2=6]	.523	.202	6.720	1	.010	.128	.918
	[ethnic2=7]	.217	.194	1.245	1	.264	-.164	.598
	[ethnic2=8]	0[a]	.	.	0	.	.	.
	[gender=0]	.110	.040	7.573	1	.006	.032	.188
	[gender=1]	0[a]	.	.	0	.	.	.
	sec2	-.129	.012	116.388	1	.000	-.153	-.106
	ks2stand	.291	.004	5858.585	1	.000	.284	.299
	[ethnic2=1] * sec2	.042	.041	1.073	1	.300	-.038	.123
	[ethnic2=2] * sec2	.050	.039	1.654	1	.198	-.026	.125
	[ethnic2=3] * sec2	.039	.040	.941	1	.332	-.040	.117
	[ethnic2=4] * sec2	.035	.054	.409	1	.522	-.072	.141
	[ethnic2=5] * sec2	.036	.051	.504	1	.478	-.064	.136
	[ethnic2=6] * sec2	.003	.046	.004	1	.950	-.088	.094
	[ethnic2=7] * sec2	.084	.047	3.231	1	.072	-.008	.177
	[ethnic2=8] * sec2	0[a]	.	.	0	.	.	.

Below is the parameter estimates table from a model run after we have dropped the *ethnic2*sec2* interaction terms (**Figure 5.8.3**).

Figure 5.8.3: Parameter estimates after dropping the *ethnic2*sec2* interactions.

		Estimate	SE	Wald	df	Sig.	95% CI Lower	95% CI Upper
Threshold	[k3matier = 0]	-2.915	.056	2688.072	1	.000	-3.025	-2.805
	[k3matier = 1]	.323	.047	48.182	1	.000	.232	.415
	[k3matier = 2]	3.244	.058	3104.820	1	.000	3.129	3.358
Location	[ethnic2=1]	.028	.090	.096	1	.757	-.149	.205
	[ethnic2=2]	.808	.082	98.151	1	.000	.648	.968
	[ethnic2=3]	.617	.089	48.437	1	.000	.443	.791
	[ethnic2=4]	.401	.107	14.166	1	.000	.192	.610
	[ethnic2=5]	-.410	.108	14.438	1	.000	-.622	-.199
	[ethnic2=6]	.524	.120	19.193	1	.000	.289	.758
	[ethnic2=7]	.495	.110	20.143	1	.000	.279	.711
	[ethnic2=8]	0[a]	.	.	0	.	.	.
	[gender=0]	.108	.040	7.319	1	.007	.030	.186
	[gender=1]	0[a]	.	.	0	.	.	.
	sec2	-.115	.010	138.094	1	.000	-.134	-.096
	ks2stand	.291	.004	5868.388	1	.000	.284	.299

Clearly prior attainment has accounted for a large proportion of the variance in tier entry. A unit change in KS2 standard score increase the odds of entry to the higher tier by exp(.291) = 1.34. The SD of *ks2stand* is 10 and Exp(.291*10)= 18.4, so a student scoring 1 SD above the mean score at age 11 is 18 times more likely to be entered for a higher tier than a student with the average (mean) age 11 score. Equally a student with an age 11 score 1 SD below the mean is 18 times less likely to be entered for a higher tier than a student with the mean age 11 score. However there are still significant association with ethnic group, SEC and gender.

Boys (gender=0) are significantly more likely to be entered for the higher tier, even after prior attainment, SEC and ethnicity are controlled: Exp(.108)=1.11, *p<.007*. Students from lower SEC are still less likely to be entered for the higher tiers, even after control for prior attainment, ethnic group and gender: Exp(-.115)=0.89, *p<.005*. In relation to ethnic group there are mixed results. Students from Indian (*ethnic2= 2*), Pakistani (*ethnic2= 3*), Bangladeshi (*ethnic2= 4*), Black African (*ethnic2= 6*) and Any other ethnic group (*ethnic2= 7*) are significantly more likely to be entered to higher tiers than comparable White British students. For example Pakistani students are Exp(.617)=1.85 times as likely as similar White British students to be entered to higher tiers. However Black Caribbean (*ethnic2= 5*) students are distinctive as the only ethnic group to be less likely to be entered to higher tiers than White British students (Exp(-.410)=0.66, *p<.005*). Even after control for prior attainment, SEC and gender, the odds of Black Caribbean students being entered for the higher tiers are about

two-thirds of the odds for White British students. Put the other way round the odds for White British students of being entered are (1/.66)=1.5 times or 50% greater than the odds for Black Caribbean students of the same prior attainment, SEC and gender.

Test of parallel lines

However the test of parallel lines (**Figure 5.8.4**) still indicates that the PO assumption may not hold for the data, with the '*p<.000*' indicating rejection of the null hypothesis of PO.

Figure 5.8.4: Test of Parallel Lines after dropping the *ethnic2*sec2* interactions.

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	8102.674			
General	8051.247	51.427	20	.000

This indicates that there may be some explanatory variables for which the ORs are not stable across different cumulative thresholds in relation to tier of entry. We therefore need to delve into this a bit deeper.

5.9 What to do if the assumption of proportional odds is not met?

As we said earlier (on **Page 5.4** while 'Examining the SPSS output') the test of the PO assumption has been described as "*anti-conservative, that is it nearly always results in rejection of the proportional odds assumption, particularly when the number of explanatory variables is large (Brant, 1990), the sample size is large (Allison, 1999; Clogg and Shihadeh, 1994) or there is a continuous explanatory variable in the model (Allison, 1999).*" (O'Connell, 2006, p29). It is important to examine the data using a set of separate logistic regression equations to explicitly see how the ORs for our explanatory variables vary at the different thresholds.

In **Figure 5.9.1** below we have dichotomised the ordinal tier variable at three cut-off points: *top tier, upper tier or above, and lower tier or above* (there is no need for a model for bottom tier or above since this includes all students). You can find these three variables in the dataset called *TierTop, TierUpper* and *TierLower* respectively. We have then completed three separate binary logistic regression models (see **Module 4**!), one for each dichotomised response.

Figure 5.9.1: Results of three binary logistic regressions for entry to top tier, upper tier or above and lower tier or above

| | B coefficents | | | Odds Ratios | | | test of parallel lines p value |
| | Tier entered for | | | Tier entered for | | | |
	Lower or above	Upper or above	Top	Lower or above	Upper or above	Top	
Intercept	3.02	-.26	-2.77	-	-	-	
Mixed	.284	-.230	.062	1.33	.79	1.06	.027
Indian	.981***	.765***	.744***	2.67	2.15	2.10	.015
Pakistani	.643***	.684***	.576***	1.90	1.98	1.78	.103
Bangladeshi	.400*	.342*	.492*	1.49	1.41	1.64	.477
Black Caribbean	-.304*	-.490**	-.584*	.74	.61	.56	.008
Black African	.741***	.442**	.391	2.10	1.56	1.48	.945
Any Other ethnic group	.662***	.210	.707***	1.94	1.23	2.03	.018
Girls	.028	-.097	-.300***	1.03	.91	.74	.000
SEC	-.120***	-.097***	-.140***	.89	.91	.87	.000
Age 11 score	.305***	.309***	.268***	1.36	1.36	1.31	.000
R^2 (Nagelkerke)	61.1	65.9	56.2				
Model Chi-squared	5405***	7564***	4600***				
Hosmer & Lemeshow test	7.6 ns	12.1 ns	42.1 p<.005				

*Notes: *=p<.05, **=p<.01; ***=p<.001.*

We can now examine the ORs for each explanatory variable across the separate splits in the data to determine how consistent they are. As an additional check on the proportional odds assumption, the test of parallel lines was completed for each explanatory variable separately, that is separate ordinal models were completed each containing only a single explanatory variable. For ethnic group this was done using the dummy variables *e1* (mixed heritage) to *e7* (Any Other group), entering each in a separate model. Because of the large sample size a *p<.01* level was used to guide decisions regarding non-proportionality. The p values are presented in the final column of **Figure 5.9.1**.

The PO assumption appears to be rejected for both *Sec2* and *KS2stand* using the separate tests of parallel lines (*p<.000*), but as explained earlier these are continuous variables and are likely to result in a high proportion of empty cells. Looking at the separate ORs across the three splits the difference in ORs appears negligible (.87 to .91 for *Sec2,* and 1.31 to 1.36 for *ks2stand),* so a common OR for each of these variables is a very plausible assumption.

In relation to ethnic group, the PO assumption is upheld (*p>.01*) for all ethnic groups except for Black Caribbean. Looking across the separate ORs this seems plausible. While the ORs for Indian and Black African students are particularly high for the lower tier or above relative to the other two splits, the coefficients are all in the same direction. For Black Caribbean students the p value (*p<.008*) is just below our selected critical level *p<.01*. The cumulative OR of 0.66 does slightly underestimate the under-representation of Black Caribbean students in the top tier (0.56), and slightly overestimate their under-representation in the lower tier or above (0.74). However the coefficient for Black Caribbean students is broadly consistent in direction and magnitude across all the splits in the data. Black Caribbean students are the only ethnic group that is under-represented relative to White British students, the result is statistically significant at all splits, with broadly similar ORs which average close to the ordinal OR (.74+.61+.56) / 3 = 0.64. The OR for Black Caribbean students from the ordinal PO model therefore represents a reasonable summary of the overall pattern for Black Caribbean students. Importantly this under-representation persists after we control for prior attainment and for social class and so cannot be accounted for by these factors.

The variable most out of line with the PO assumption is gender. The PO assumption is strongly rejected for this variable (*p<.000*). In the ordinal PO model the OR for boys was 1.11, which is the same as an OR for girls of (1/1.11) = 0.90. This suggests girls are under-represented relative to boys in entry to higher tiers. However this over-simplifies the data which shows girls are significantly under-represented in the top tier (0.74) but not significantly so for upper or above (.91) or for lower or above (1.03). This finding was obscured in the

single cumulative OR for gender and summarising this relationship in a single OR misses this observation. Thus gender may well be the major factor underlying the overall rejection of the PO assumption.

There are partial proportional odds (PPO) models that allow the assumption of PO to be relaxed for one or a small subset of explanatory variables, but retained for the majority of explanatory variables. In the present case it might be apposite to run such a model, relaxing the PO assumption for the gender variable. However this requires restructuring of the dataset which is beyond the scope of this web resource. The reader is referred to O'Connell (2006) p48-54 for further detail on PPO models. For the current analysis, inspection of the separate ORs for a series of binary logistic regressions suggests it is reasonable to conclude the ordinal PO model is a fair summary of the patterns in the data in relation to tier of entry, but noting that the under-representation of girls is specific to the top tier and should not be assumed to generalise across all cumulative splits in the data.

5.10 Reporting the results of ordinal regression

Perhaps the most noteworthy outcome from the analysis of the example completed above is the finding that Black Caribbean students are under-represented in entry to the higher mathematics test tiers relative to White British students in the ratio 0.66:1, i.e. the odds of being entered to the higher tiers for Black Caribbean students are about two-thirds the odds for White British students. As we saw in Module 4 (page 4.8) we can express this in % terms by subtracting 1 from the OR and multiplying by 100: (0.66-1)*100 = a 34% decrease in the odds for Black Caribbean students. Put the other way round the odds for White British students of being entered for higher tiers are (1/.66)=1.5 times or 50% greater than the odds for Black Caribbean students. Importantly our models show that this finding cannot be explained in terms of the prior attainment of the students or by differences in social class composition. If you are interested to see a full ordinal analysis of the tiering data and how it is reported then you can find this in the following journal article:

- Strand, S. (2011). The White British-Black Caribbean achievement gap: Tests, tiers and teacher expectations. *British Educational Research Journal, iFirst at* _Link to journal article_

This shows how an ordinal model was built hierarchically over a series of steps, looking first at prior attainment then progressively adding further explanatory variables. Note that the results from an analysis of the summary LSYPE dataset used here will not agree precisely with the analyses presented in the paper because slightly different variables are used, and the data in the paper have been weighted to account for selection and non-response issues and for clustering within schools, but the overall pattern of findings will be similar. The paper also discusses in depth the interpretation of the model findings, and what these mean in relation to policy and practice. Statistical models are only useful if the results and their implications can be communicated clearly to the intended audiences.

5.11 Conclusions

In this module we have looked at regression models that can be applied when our outcome is represented by an ordinal variable. Several authors (e.g., Ananth & Kleinbaum, 1997) argue that ordinal models are under-used in social sciences, with researchers often reducing ordinal data to a series of binary logistic regressions. In contrast ordinal regression models take advantage of the ordinality in the outcome by summarising the relationships between explanatory variables and the outcome in a single parsimonious model. However the most popular approach to ordinal regression - the proportional odds (PO) model - makes an substantial simplifying assumption that the ORs associated with each explanatory variable are the same over the cumulative splits in the outcome. This assumption has to be tested if ordinal models are to be applied appropriately. We have noted that the statistical test of the PO assumption is over-conservative and the PO assumption is frequently rejected particularly when the number of explanatory variables is large (Brant, 1990), the sample size is large (Allison, 1999; Clogg & Shihadeh, 1994) or there is a continuous explanatory variable in the model (Allison, 1999). Therefore we have seen how to evaluate the ordinal PO model by completing a series of binary logistic regressions at each of the cumulative splits in the data, and how this can allow us to directly evaluate the consistency in ORs across an ordinal outcome. Where the PO assumption is justified ordinal regression models can be a powerful means of summarising relationships that utilises all the information present in the ordinal outcome. As you can tell, there is some debate by statisticians about when to use Ordinal Regression. A sure sign that you have reached the end of a tricky Module... Well done!

5.12 Other types of categorical models

Other ordinal models

The PO model used in SPSS is by far the most common form of ordinal regression. However there are other approaches to ordinal regression, for example the continuation ratio model and the adjacent categories model. Both involve restructuring your dataset, sometimes using link functions other than the logit and sometimes statistical packages other than SPSS. They are therefore not core to our coverage in this module. The interested reader can find excellent treatment of such models in O'Connell (2006).

Log-linear analysis

You might also come across a range of models called log-linear models as you read statistics texts (for light reading on a Sunday morning in bed perhaps). This class of models is used where we are interested in the relationships between a set of variables *but have no clear outcome or dependent variable*. However where we have a variable that clearly represents an outcome (such as a test score, NC level, tier of entry etc.) it makes sense to analyse this in a regression framework asking how various explanatory variables impact on the outcome variable. In such cases a linear, logistic or ordinal regression model is more appropriate than log-linear analysis. For an introduction to log-linear analysis including a guide to implementation in SPSS we recommend Acton & Miller (2009) Chapter 11.

Multinomial logistic regression

Another type of model is multinomial logistic regression. In such models there is a clear outcome variable, but there is usually no ordinality in the outcome. An example in education might be in identifying the characteristics of students who select different kinds of secondary schools (e.g. Independent schools, Church schools or Local Authority schools). In this case there are therefore multiple outcomes but no ordinality in the outcome. Some authors recommend undertaking multinomial regression even with an ordinal outcome if the PO assumption is not met (e.g. Norusis, 2005). However we can see no advantage to this, indeed if there is ordinality in your outcome and the PO assumption is not supported it is definitely better to complete a series of logistic regressions on cumulative splits in the outcome as demonstrated in this module. For good coverage on multinomial logistic regression we recommend Field (2009) Chapter 8.

References

Acton, C., & Miller, R. (2009). *SPSS for social scientists (second edition).* Basingstoke: Palgrave Macmillan.

Allison, P. D. (1999). *Logistic regression using the SAS system: Theory and application.* Cary, NC.: SAS Institute.

Ananth, C. V., & Kleinbaum, D. G. (1997). Regression models for ordinal responses: a review of methods and applications. *International Journal of Epidemiology, 26*(6), 1323-1333.

Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics, 46*, 1171-1178.

Clogg, C., & Shihadeh, E. S. (1994). *Statistical models for ordinal variables.* Thousand Oaks, California: Sage publications.

Norusis, M. (2005). *SPSS 14.0: Advanced statistical companion.* New Jersey: Prentice Hall Inc.

O'Connell, A. (2006). *Logistic regression models for ordinal response variables.* Thousand Oaks: Sage publications.

Module 5 Exercise

We have seen that Black Caribbean students are systematically under-represented relative to White British students in entry to higher tiers of the age 14 national mathematics test. This difference remains significant even after controlling for prior attainment, socio-economic class of the home and gender. However are there other variables in the LSYPE dataset that may account for their under-representation in entry to the higher tiers?

Use the **LSYPE 15,000 dataset** to work through each of the following questions. Answer them in full sentences with supporting tables or graphs where appropriate as this will help when you to better understand how you may apply these techniques to your own research. The answers are on the next page.

Note: *The variable names as they appear in the SPSS dataset are listed in italics. We have also included some hints in italics.*

1. Explore the relationship between ethnic group (*ethnic2*) and a) having an identified Special Educational Need (*sen*), b) whether the student reported truanting at any time during year 9 (*truancy)* and c) whether the student has been excluded from school at any point during Years 7 and 9 (*exclude*). What are the differences between Black Caribbean and White British students on these variables? Are any differences statistically significant?

    Use crosstabs and chi-square analyses.

2. Complete an ordinal regression similar to the one we have used as an example throughout this module but entering the variables *SEN, truancy* and *exclude* along with the factors *(ethnic2, gender and sec2)* and the covariate (*ks2stand*). Are *SEN, truancy* and *exclude* related to the likelihood of a student being entered to a higher mathematics tier? What are the odds ratios for these three new variables?

    Use the Parameter Estimates Table and calculate the odds ratios.

3. Does the addition of these new variables change the ethnic Odds Ratios substantially compared to **Figure 5.8.3 (Page 5.3)**? Are Black Caribbean students still under-represented relative to White British students in entry to the higher maths tiers?

    Compare the two Parameter Estimates Tables.

4. Is the assumption of 'Proportional Odds' met for this final version of the model?

    Request a 'Test of Parallel Lines' from the 'Output' submenu when running the regression analysis.

Answers

1. Explore the relationship between ethnic group (*ethnic2*) and a) having an identified Special Educational Need (*sen*), b) whether the student reported truanting at any time during year 9 (*truancy*) and c) whether the student has been excluded from school at any point during Years 7 and 9 (*exclude*). What are the differences between Black Caribbean and White British students on these variables? Are any differences statistically significant?

The following cross-tabulations were created using the **Analyze>Descriptive Statistics>Crosstabs** option on SPSS. We have also included the *row percentages* and *Pearson's Chi-square* for each crosstab (the results of which are reported above each table). If you are unsure about how to do this we describe the process on **Page 2.2**. Note that we have re-sized each table and added the numeric codes for each category. This is to make them easier to interpret here so don't worry if your output looks slightly different (though the actual figures should be the same!).

Ethnicity by SEN crosstab

Chi-square = 48.03, df = 7, *p* <.0005.

			SEN statemented or School Action Plus		
			0 No	1 Yes	Total
Ethnic group	0 White British	Count	8803	730	9533
		%	92.3%	7.7%	100.0%
	1 Mixed heritage	Count	709	53	762
		%	93.0%	7.0%	100.0%
	2 Indian	Count	952	41	993
		%	95.9%	4.1%	100.0%
	3 Pakistani	Count	854	66	920
		%	92.8%	7.2%	100.0%
	4 Bangladeshi	Count	654	52	706
		%	92.6%	7.4%	100.0%
	5 Black Caribbean	Count	490	70	560
		%	87.5%	12.5%	100.0%
	6 Black African	Count	543	33	576
		%	94.3%	5.7%	100.0%
	7 Any other group	Count	573	27	600
		%	95.5%	4.5%	100.0%
Total		Count	13578	1072	14650
		%	92.7%	7.3%	100.0%

Ethnicity by Truancy crosstab

Chi-square = 66.56, df = 7, *p* <.0005.

| | | | Truant in last 12 months | | |
			0 No	1 Yes	Total
Ethnic group	0 White British	Count	8041	1514	9555
		%	84.2%	15.8%	100.0%
	1 Mixed heritage	Count	586	154	740
		%	79.2%	20.8%	100.0%
	2 Indian	Count	855	84	939
		%	91.1%	8.9%	100.0%
	3 Pakistani	Count	725	125	850
		%	85.3%	14.7%	100.0%
	4 Bangladeshi	Count	521	93	614
		%	84.9%	15.1%	100.0%
	5 Black Caribbean	Count	435	111	546
		%	79.7%	20.3%	100.0%
	6 Black African	Count	479	65	544
		%	88.1%	11.9%	100.0%
	7 Any other group	Count	525	77	602
		%	87.2%	12.8%	100.0%
Total		Count	12167	2223	14390
		%	84.6%	15.4%	100.0%

Ethnicity by Exclusion crosstab

Chi-square = 121.85, df = 7, *p* <.0005.

| | | | One or more exclusions from school age 11-14 | | |
			0 No	1 Yes	Total
Ethnic group	0 White British	Count	8783	1015	9798
		%	89.6%	10.4%	100.0%
	1 Mixed heritage	Count	636	117	753
		%	84.5%	15.5%	100.0%
	2 Indian	Count	796	37	833
		%	95.6%	4.4%	100.0%
	3 Pakistani	Count	504	43	547
		%	92.1%	7.9%	100.0%
	4 Bangladeshi	Count	300	16	316
		%	94.9%	5.1%	100.0%
	5 Black Caribbean	Count	431	108	539
		%	80.0%	20.0%	100.0%
	6 Black African	Count	449	50	499
		%	90.0%	10.0%	100.0%
	7 Any other group	Count	481	47	528
		%	91.1%	8.9%	100.0%
Total		Count	12380	1433	13813
		%	89.6%	10.4%	100.0%

The results indicate that there are statistically significant associations between ethnicity and each of SEN, truancy and exclusion – in short that there are significant differences between ethnic groups with regard to these three variables. Let us explore the specific comparison

between White British and Black Caribbean students more closely. We can see that Black Caribbean students are more likely to be identified as having special educational needs (12.5%) than White British students (7.7%). They are also more likely to report truanting (20.3% of students) compared to White British students (15.8%). Finally they are nearly twice as likely to be excluded as White British students with 20% having been excluded compared to only 10.4% of White British students. There seem to be grounds to include all three of these explanatory variables in any model which attempts to account for ethnic differences in entry to different test tiers.

2. Complete an ordinal regression similar to the one we have used as an example throughout this module but entering the variables *SEN, truancy* and *exclude* along with the factors *(ethnic2, gender and sec2)* and the covariate (*ks2stand*). Are *SEN, truancy* and *exclude* related to the likelihood of a student being entered to a higher mathematics tier? What are the odds ratios for these three new variables?

An Ordinal Regression was completed using the model as shown in **Figure 5.8.3** but with the addition of the SEN, truant and exclusion variable. The key output from the parameter encoding table is shown below:

Parameter Estimates Table

		Estimate	Std. Error	Wald	df	Sig.	95% CI Lower	95% CI Upper
Threshold	[k3matier = 0]	-.936	.173	29.096	1	.000	-1.275	-.596
	[k3matier = 1]	2.378	.176	182.137	1	.000	2.033	2.723
	[k3matier = 2]	5.340	.182	858.016	1	.000	4.983	5.697
Location	ks2stand	.290	.004	4.902E3	1	.000	.282	.299
	[gender=0]	.189	.044	18.826	1	.000	.104	.274
	[gender=1]	0ª	.	.	0	.	.	.
	[sec2=0]	.690	.124	30.762	1	.000	.446	.933
	[ethnic2=7]	.359	.123	8.456	1	.004	.117	.601
	[ethnic2=8]	0ª	.	.	0	.	.	.
	[sen=0]	.603	.122	24.460	1	.000	.364	.843
	[sen=1]	0ª	.	.	0	.	.	.
	[truancy=0]	.273	.064	17.926	1	.000	.146	.399
	[truancy=1]	0ª	.	.	0	.	.	.
	[exclude=0]	.604	.082	53.848	1	.000	.443	.765
	[exclude=1]	0ª	.	.	0	.	.	.

Let's look at the coefficients for our three new explanatory variables (*sen, truancy* and *exclude*). To make sense of these logits we will need to take the exponentials to turn them into ORs. This is where EXCEL or a graphics calculator comes in handy! So:

SEN	Exp (.603) = 1.83
Exclude	Exp (.604) = 1.83
Truancy	Exp (.273) = 1.31

In SPSS ordinal regression the reference category for nominal explanatory variables is indicated by the value 1, and for all three variable the value '1' indicates 'yes'. The fact that the ORs are greater than 1 tells us that students who do not have SEN, who do not report truanting and who have never been excluded from school are *more* likely to be entered for the higher tiers. Those who do not have an identified SEN are nearly twice (1.83 times) as likely

to be entered for a higher tier as those with SEN. Similarly the odds of students who have never been excluded from school being entered to a higher tier are 1.83 higher than the odds for students who have been excluded. Finally there is a less pronounced but still highly significant association with truancy: students who did not report truanting were 1.31 times more likely to be entered for a higher tier than those who did report truanting. Remember that these associations are significant even after differences in prior attainment, ethnicity, gender and socio-economic class of the home are accounted for.

3. Does the addition of these new variables change the ethnic Odds Ratios substantially compared to Figure 5.8.3 (Page 5.3)? Are Black Caribbean students still under-represented relative to White British students in entry to the higher maths tiers?

In the original model - before we entered *SEN, exclude and truancy* (see **Figure 5.8.3**) - the logit for Black Caribbean students was -.410, and the exponential of this indicates an odds ratio of 0.66. In this original model then Black Caribbean students were roughly two-thirds as likely as White British students to be entered for the highest tier.

In our new model (see table below) we can see that the logit for Black Caribbean students (ethnic2=5) is -.348 which is an odds ratio of Exp (-.189) = 0.71. Thus the under-representation of Black Caribbean students is very slightly reduced, but not by much. The statistical significance of the coefficient for Black Caribbean is still very high (*p<.0005*). So even after accounting for the higher likelihood of Black Caribbean students to have SEN, to truant or to be excluded from school, Black Caribbean students are still roughly two-thirds as likely as comparable White British students to be entered for a higher tier.

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% CI Lower	95% CI Upper
Threshold	[k3matier = 0]	-.936	.173	29.096	1	.000	-1.275	-.596
	[k3matier = 1]	2.378	.176	182.137	1	.000	2.033	2.723
	[k3matier = 2]	5.340	.182	858.016	1	.000	4.983	5.697
Location	ks2stand	.290	.004	4.902E3	1	.000	.282	.299
	[gender=0]	.189	.044	18.826	1	.000	.104	.274
	[gender=1]	0ª	.	.	0	.	.	.
	[sec2=0]	.690	.124	30.762	1	.000	.446	.933
	[sec2=7]	0ª	.	.	0	.	.	.
	[ethnic2=1]	.118	.096	1.518	1	.218	-.070	.305
	[ethnic2=2]	.750	.091	67.789	1	.000	.571	.928
	[ethnic2=3]	.527	.113	21.618	1	.000	.305	.749
	[ethnic2=4]	.422	.158	7.142	1	.008	.113	.732
	[ethnic2=5]	-.348	.115	9.220	1	.002	-.573	-.123
	[ethnic2=6]	.482	.134	12.991	1	.000	.220	.744
	[ethnic2=7]	.359	.123	8.456	1	.004	.117	.601
	[ethnic2=8]	0ª	.	.	0	.	.	.
	[sen=0]	.603	.122	24.460	1	.000	.364	.843
	[sen=1]	0ª	.	.	0	.	.	.
	[truancy=0]	.273	.064	17.926	1	.000	.146	.399
	[truancy=1]	0ª	.	.	0	.	.	.
	[exclude=0]	.604	.082	53.848	1	.000	.443	.765
	[exclude=1]	0ª	.	.	0	.	.	.

4. Is the assumption of 'Proportional Odds' met for this final model?

It is worth noting that the test of parallel lines for the model is significant, so strictly speaking the proportional odds assumption is rejected, although we should remember the anti-conservative nature of this test.

Test of Parallel Lines

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	8709.137			
General	8631.969	77.169	38	.000

For completeness you *might* want to undertake separate binary logistic regressions for *TierTop, TierUpper* and *TierLower* as was demonstrated on **Page 5.9**, this time also including the three new variables in the model. The results for the Black Caribbean variable are reported below.

Coefficient for Black Caribbean in three separate binary logistic regressions for a model including SEN, truancy and school exclusion as well as prior attainment, ethnicity, gender and SEC

Explanatory Variable		lower tier or above	Upper tier or above	Top tier	Average (a)	Ordinal model
Black Caribbean	logit	-0.179	-0.518***	-0.552*	-	-0.35
	OR	0.84	0.60	0.58	0.67	0.71

*Notes. (a). The average value across the three logistic models. *=p<.05, **=p<.005.*

You will see that with the inclusion of *SEN, truancy* and *exclude* the under-representation of Black Caribbean students in entry to the lower tier or above compared to the bottom tier (0.84) is no longer statistically significant. However the under-representation of Black Caribbean students in entry to the upper tier and above (0.60), and to the top tier (0.58), are large and statistically significant. As a result the cumulative OR from the ordinal model (0.71) is a somewhat less robust summary than was previously the case (see discussion on **Page 5.9**). In this example it would probably be wise to report both the separate logistic as well as the ordinal regression results, so that readers are aware of the variation in the OR across the separate thresholds.