

Ordinal Regression

Many variables of interest are ordinal. That is, you can rank the values, but the real distance between categories is unknown. Diseases are graded on scales from *least severe* to *most severe*. Survey respondents choose answers on scales from *strongly agree* to *strongly disagree*. Students are graded on scales from *A* to *F*.

You can use ordinal categorical variables as predictors, or factors, in many statistical procedures, such as linear regression. However, you have to make difficult decisions. Should you forget the ordering of the values and treat your categorical variables as if they are nominal? Should you substitute some sort of scale (for example, numbers 1 to 5) and pretend the variables are interval? Should you use some other transformation of the values hoping to capture some of that extra information in the ordinal scale?

When your dependent variable is ordinal you also face a quandary. You can forget about the ordering and fit a multinomial logit model that ignores any ordering of the values of the dependent variable. You fit the same model if your groups are defined by color of car driven or severity of a disease. You estimate coefficients that capture differences between all possible pairs of groups. Or you can apply a model that incorporates the ordinal nature of the dependent variable.

The SPSS Ordinal Regression procedure, or PLUM (**P**olytomous **U**niversal **M**odel), is an extension of the general linear model to ordinal categorical data. You can specify five link functions as well as scaling parameters. The procedure can be used to fit heteroscedastic probit and logit models.

Fitting an Ordinal Logit Model

Before delving into the formulation of ordinal regression models as specialized cases of the general linear model, let's consider a simple example. To fit a binary logistic regression model, you estimate a set of regression coefficients that predict the probability of the outcome of interest. The same logistic model can be written in different ways. The version that shows what function of the probabilities results in a linear combination of parameters is

$$\ln\left(\frac{\text{prob}(\text{event})}{1 - \text{prob}(\text{event})}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

The quantity to the left of the equal sign is called a **logit**. It's the log of the odds that an event occurs. (The odds that an event occurs is the ratio of the number of people who experience the event to the number of people who do not. This is what you get when you divide the probability that the event occurs by the probability that the event does not occur, since both probabilities have the same denominator and it cancels, leaving the number of events divided by the number of non-events.) The coefficients in the logistic regression model tell you how much the logit changes based on the values of the predictor variables.

When you have more than two events, you can extend the binary logistic regression model, as described in Chapter 3. For ordinal categorical variables, the drawback of the multinomial regression model is that the ordering of the categories is ignored.

Modeling Cumulative Counts

You can modify the binary logistic regression model to incorporate the ordinal nature of a dependent variable by defining the probabilities differently. Instead of considering the probability of an individual event, you consider the probability of that event and all events that are ordered before it.

Consider the following example. A random sample of Vermont voters was asked to rate their satisfaction with the criminal justice system in the state (Doble, 1999). They rated judges on the scale: *Poor* (1), *Only fair* (2), *Good* (3), and *Excellent* (4). They also indicated whether they or anyone in their family was a crime victim in the last three years. You want to model the relationship between their rating and having a crime victim in the household.

Defining the Event

In ordinal logistic regression, the event of interest is observing a particular score *or less*. For the rating of judges, you model the following odds:

$$\theta_1 = \text{prob}(\text{score of } 1) / \text{prob}(\text{score greater than } 1)$$

$$\theta_2 = \text{prob}(\text{score of } 1 \text{ or } 2) / \text{prob}(\text{score greater than } 2)$$

$$\theta_3 = \text{prob}(\text{score of } 1, 2, \text{ or } 3) / \text{prob}(\text{score greater than } 3)$$

The last category doesn't have an odds associated with it since the probability of scoring up to and including the last score is 1.

All of the odds are of the form:

$$\theta_j = \text{prob}(\text{score} \leq j) / \text{prob}(\text{score} > j)$$

You can also write the equation as

$$\theta_j = \text{prob}(\text{score} \leq j) / (1 - \text{prob}(\text{score} \leq j)),$$

since the probability of a score greater than j is $1 - \text{probability of a score less than or equal to } j$.

Ordinal Model

The ordinal logistic model for a single independent variable is then

$$\ln(\theta_j) = \alpha_j - \beta X$$

where j goes from 1 to the number of categories minus 1.

It is not a typo that there is a minus sign before the coefficients for the predictor variables, instead of the customary plus sign. That is done so that larger coefficients indicate an association with larger scores. When you see a positive coefficient for a dichotomous factor, you know that higher scores are more likely for the first category. A negative coefficient tells you that lower scores are more likely. For a continuous variable, a positive coefficient tells you that as the values of the variable increase, the likelihood of larger scores increases. An association with higher scores means smaller cumulative probabilities for lower scores, since they are less likely to occur.

Each logit has its own α_j term but the same coefficient β . That means that the effect of the independent variable is the same for different logit functions. That's an assumption you have to check. That's also the reason the model is also called the proportional odds model. The α_j terms, called the threshold values, often aren't of much interest. Their

values do not depend on the values of the independent variable for a particular case. They are like the intercept in a linear regression, except that each logit has its own. They're used in the calculations of predicted values. From the previous equations, you also see that combining adjacent scores into a single category won't change the results for the groups that aren't involved in the merge. That's a desirable feature.

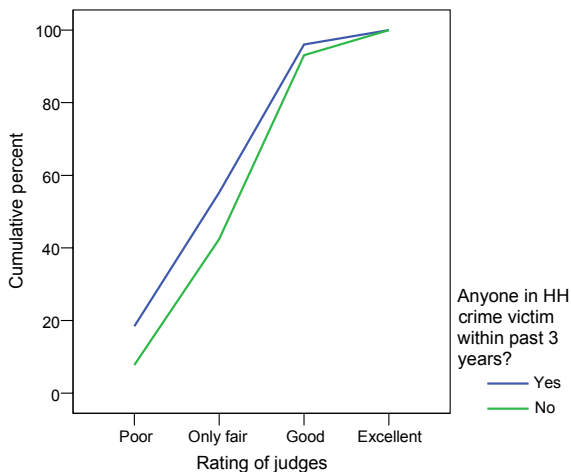
Examining Observed Cumulative Counts

Before you start building any model, you should examine the data. Figure 4-1 is a cumulative percentage plot of the ratings, with separate curves for those whose households experienced crime and those who didn't. The lines for those who experienced crime are above the lines for those who didn't. Figure 4-1 also helps you visualize the ordinal regression model. It models a function of those two curves.

Consider the rating *Poor*. A larger percentage of crime victims than non-victims chose this response. (Because it is the first response, the cumulative percentage is just the observed percentage for the response.) As additional percentages are added (the cumulative percentage for *Only fair* is the sum of *Poor* and *Only fair*), the cumulative percentages for the crime victim households remain larger than for those without crime. It's only at the end, when both groups must reach 100%, that they must join. Because the victims assign lower scores, you expect to see a negative coefficient for the predictor variable, *hhcrime* (household crime experience).

Figure 4-1

Plot of observed cumulative percentages



Specifying the Analysis

To fit the cumulative logit model, open the file *vermontcrime.sav* and from the menus choose:

```
Analyze
  Regression
    Ordinal...
  ▶ Dependent: rating
  ▶ Factors: hhcrime

Options...
Link: Logit

Output...
  Display
   Goodness of fit statistics
   Summary statistics
   Parameter estimates
   Cell information
   Test of Parallel Lines
  Saved Variables
   Estimated response probabilities
```

Parameter Estimates

Figure 4-2 contains the estimated coefficients for the model. The estimates labeled *Threshold* are the α_j 's, the intercept equivalent terms. The estimates labeled *Location* are the ones you're interested in. They are the coefficients for the predictor variables. The coefficient for *hhcrime* (coded 1 = *yes*, 2 = *no*), the independent variable in the model, is -0.633 . As is always the case with categorical predictors in models with intercepts, the number of coefficients displayed is one less than the number of categories of the variable. In this case, the coefficient is for the value of 1. Category 2 is the reference category and has a coefficient of 0.

The coefficient for those whose household experienced crime in the past three years is negative, as you expected from Figure 4-1. That means it's associated with poorer scores on the rankings of judges. If you calculate $e^{-\beta}$, that's the ratio of the odds for lower to higher scores for those experiencing crime and those not experiencing crime. In this example, $\exp(0.633) = 1.88$. This ratio stays the same over all of the ratings.

The Wald statistic is the square of the ratio of the coefficient to its standard error. Based on the small observed significance level, you can reject the null hypothesis that

it is zero. There appears to be a relationship between household crime and ratings of judges. For any rating level, people who experience crime score judges lower than those who don't experience crime.

Figure 4-2
Parameter estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[rating = 1]	-2.392	.152	248.443	1	.000	-2.690	-2.095
	[rating = 2]	-.317	.091	12.146	1	.000	-.495	-.139
	[rating = 3]	2.593	.172	228.287	1	.000	2.257	2.930
Location	[hhcrime=1]	-.633	.232	7.445	1	.006	-1.088	-.178
	[hhcrime=2]	0 ¹	.	.	0	.	.	.

Link function: Logit.

1. This parameter is set to zero because it is redundant.

Testing Parallel Lines

When you fit an ordinal regression you assume that the relationships between the independent variables and the logits are the same for all the logits. That means that the results are a set of parallel lines or planes—one for each category of the outcome variable. You can check this assumption by allowing the coefficients to vary, estimating them, and then testing whether they are all equal.

The result of the test of parallelism is in Figure 4-3. The row labeled *Null Hypothesis* contains -2 log-likelihood for the constrained model, the model that assumes the lines are parallel. The row labeled *General* is for the model with separate lines or planes. You want to know whether the general model results in a sizeable improvement in fit from the null hypothesis model.

The entry labeled *Chi-Square* is the difference between the two -2 log-likelihood values. If the lines or planes are parallel, the observed significance level for the change should be large, since the general model doesn't improve the fit very much. The parallel model is adequate. You don't want to reject the null hypothesis that the lines are parallel. From Figure 4-3, you see that the assumption is plausible for this problem. If you do reject the null hypothesis, it is possible that the link function selected is incorrect for the data or that the relationships between the independent variables and logits are not the same for all logits.

Figure 4-3
Test of parallel lines

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	30.793			
General	28.906	1.887	2	.389

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

1. Link function: Logit.

Does the Model Fit?

A standard statistical maneuver for testing whether a model fits is to compare observed and expected values. That is what's done here as well.

Calculating Expected Values

You can use the coefficients in Figure 4-2 to calculate cumulative predicted probabilities from the logistic model for each case:

$$\text{prob}(\text{event } j) = 1 / (1 + e^{-(\alpha_j - \beta x)})$$

Remember that the events in an ordinal logistic model are not individual scores but cumulative scores. First, calculate the predicted probabilities for those who didn't experience household crime. That means that β is 0, and all you have to worry about are the intercept terms.

$$\text{prob}(\text{score } 1) = 1 / (1 + e^{2.392}) = 0.0838$$

$$\text{prob}(\text{score } 1 \text{ or } 2) = 1 / (1 + e^{0.317}) = 0.4214$$

$$\text{prob}(\text{score } 1 \text{ or } 2 \text{ or } 3) = 1 / (1 + e^{-2.59}) = 0.9302$$

$$\text{prob}(\text{score } 1 \text{ or } 2 \text{ or } 3 \text{ or } 4) = 1$$

From the estimated cumulative probabilities, you can easily calculate the estimated probabilities of the individual scores for those whose households did not experience crime. You calculate the probabilities for the individual scores by subtraction, using the formula:

$$\text{prob}(\text{score} = j) = \text{prob}(\text{score less than or equal to } j) - \text{prob}(\text{score less than } j).$$

The probability for score 1 doesn't require any modifications. For the remaining scores, you calculate the differences between cumulative probabilities:

$$\text{prob}(\text{score} = 2) = \text{prob}(\text{score} = 1 \text{ or } 2) - \text{prob}(\text{score} = 1) = 0.3376$$

$$\text{prob}(\text{score} = 3) = \text{prob}(\text{score } 1, 2, 3) - \text{prob}(\text{score } 1, 2) = 0.5088$$

$$\text{prob}(\text{score} = 4) = 1 - \text{prob}(\text{score } 1, 2, 3) = 0.0698$$

You calculate the probabilities for those whose households experienced crime in the same way. The only difference is that you have to include the value of β in the equation. That is,

$$\text{prob}(\text{score} = 1) = 1 / (1 + e^{(2.392 - 0.633)}) = 0.1469$$

$$\text{prob}(\text{score} = 1 \text{ or } 2) = 1 / (1 + e^{(0.317 - 0.633)}) = 0.5783$$

$$\text{prob}(\text{score} = 1, 2, \text{ or } 3) = 1 / (1 + e^{(-2.593 - 0.633)}) = 0.9618$$

$$\text{prob}(\text{score} = 1, 2, 3, \text{ or } 4) = 1$$

Of course, you don't have to do any of the actual calculations, since SPSS will do them for you. In the Options dialog box, you can ask that the predicted probabilities for each score be saved.

Figure 4-4 gives the predicted probabilities for each cell. The output is from the Means procedure with the saved predicted probabilities (*EST1_1*, *EST2_1*, *EST3_1*, and *EST4_1*) as the dependent variables and *hhcrime* as the factor variable. All cases with the same value of *hhcrime* have the same predicted probabilities for all of the response categories. That's why the standard deviation in each cell is 0. For each rating, the estimated probabilities for everybody combined are the same as the observed marginals for the rating variable.

Figure 4-4

Estimated response probabilities

Anyone in HH crime victim within past 3 years?		Estimated Cell Probability for Response Category: 1	Estimated Cell Probability for Response Category: 2	Estimated Cell Probability for Response Category: 3	Estimated Cell Probability for Response Category: 4
Yes	Mean	.1469	.4316	.3833	.0382
	N	76	76	76	76
	Std. Deviation	.00000	.00000	.00000	.00000
No	Mean	.0838	.3378	.5089	.0696
	N	490	490	490	490
	Std. Deviation	.00000	.00000	.00000	.00000
Total	Mean	.0922	.3504	.4920	.0654
	N	566	566	566	566
	Std. Deviation	.02154	.03202	.04284	.01071

For each rating, the estimated odds of the cumulative ratings for those who experience crime divided by the estimated odds of the cumulative ratings for those who didn't experience crime is $e^{-\beta} = 1.88$. For the first response, the odds ratio is

$$\frac{0.1469/(1 - 0.1469)}{0.0838/(1 - 0.0838)} = 1.88$$

For the cumulative probability of the second response, the odds ratio is

$$\frac{(0.1469 + 0.4316)/(1 - 0.1469 - 0.4316)}{(0.0838 + 0.3378)/(1 - 0.0838 - 0.3378)} = 1.88$$

Comparing Observed and Expected Counts

You can use the previously estimated probabilities to calculate the number of cases you expect in each of the cells of a two-way crosstabulation of rating and crime in the household. You multiply the expected probabilities for those without a history by 490, the number of respondents who didn't report a history. The expected probabilities for those with a history are multiplied by 76, the number of people reporting a household history of crime. These are the numbers you see in Figure 4-5 in the row labeled *Expected*. The row labeled *Observed* is the actual count.

The Pearson residual is a standardized difference between the observed and predicted values:

$$\text{Pearson residual} = \frac{O_{ij} - E_{ij}}{\sqrt{n_i \hat{p}_{ij}(1 - \hat{p}_{ij})}}$$

Figure 4-5
Cell information

Anyone in HH crime victim within past 3 years?		Rating of judges			
		Poor	Only fair	Good	Excellent
Yes	Observed	14	28	31	3
	Expected	11.16	32.800	29.134	2.903
	Pearson Residual	.919	-1.112	.440	.058
No	Observed	38	170	248	34
	Expected	41.05	165.501	249.4	34.094
	Pearson Residual	-.497	.430	-.123	-.017

Link function: Logit.

Goodness-of-Fit Measures

From the observed and expected frequencies, you can compute the usual Pearson and Deviance goodness-of-fit measures. The Pearson goodness-of-fit statistic is

$$\chi^2 = \sum \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

The deviance measure is

$$D = 2 \sum \sum O_{ij} \ln \left(\frac{O_{ij}}{E_{ij}} \right)$$

Both of the goodness-of-fit statistics should be used only for models that have reasonably large expected values in each cell. If you have a continuous independent variable or many categorical predictors or some predictors with many values, you may have many cells with small expected values. SPSS warns you about the number of empty cells in your design. In this situation, neither statistic provides a dependable goodness-of-fit test.

If your model fits well, the observed and expected cell counts are similar, the value of each statistic is small, and the observed significance level is large. You reject the null hypothesis that the model fits if the observed significance level for the goodness-of-fit statistics is small. Good models have large observed significance levels. In Figure 4-6, you see that the goodness-of-fit measures have large observed significance levels, so it appears that the model fits.

Figure 4-6
Goodness-of-fit statistics

	Chi-Square	df	Sig.
Pearson	1.902	2	.386
Deviance	1.887	2	.389

Link function: Logit.

Including Additional Predictor Variables

A single predictor variable example makes explaining the basics easier, but real problems almost always involve more than one predictor. Consider what happens when

additional factor variables—such as *sex*, *age2* (two categories), and *educ5* (five categories)—are included as well.

Recall the Ordinal Regression dialog box and select:

- ▶ Dependent: rating
- ▶ Factors: hcrime, sex, age2, educ5

Options...

Link: Logit

Output...

Display

- Goodness of fit statistics
- Summary statistics
- Parameter estimates
- Test of Parallel Lines

Saved Variables

- Predicted category

The dimensions of the problem have quickly escalated. You've gone from eight cells, defined by the four ranks and two crime categories, to 160 cells. The number of cases with valid values for all of the variables is 536, so cells with small observed and predicted frequencies will be a problem for the tests that evaluate the goodness of fit of the model. That's why the warning in Figure 4-7 appears.

Figure 4-7

Warning for empty cells

There are 44 (29.7%) cells (i.e., dependent variable levels by combinations of predictor variable values) with zero frequencies.

Overall Model Test

Before proceeding to examine the individual coefficients, you want to look at an overall test of the null hypothesis that the location coefficients for all of the variables in the model are 0. You can base this on the change in $-2 \log$ -likelihood when the variables are added to a model that contains only the intercept. The change in likelihood function has a chi-square distribution even when there are cells with small observed and predicted counts.

From Figure 4-8, you see that the difference between the two log-likelihoods—the chi square—has an observed significance level of less than 0.0005. This means that

you can reject the null hypothesis that the model without predictors is as good as the model with the predictors.

Figure 4-8

Model-fitting information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	322.784			
Final	288.600	34.183	7	.000

Link function: Logit.

You also want to test the assumption that the regression coefficients are the same for all four categories. If you reject the assumption of parallelism, you should consider using multinomial regression, which estimates separate coefficients for each category. Since the observed significance level in Figure 4-9 is large, you don't have sufficient evidence to reject the parallelism hypothesis.

Figure 4-9

Test of parallelism

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	288.600			
General	276.115	12.485	14	.567

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

Examining the Coefficients

From the observed significance levels in Figure 4-10, you see that sex, education, and household history of crime are all related to the ratings. They all have negative coefficients. Men (code 1) are less likely to assign higher ratings than women, people with less education are less likely to assign higher ratings than people with graduate education (code 5), and persons whose households have been victims of crime are less likely to assign higher ratings than those in crime-free households. Age doesn't appear to be related to the rating.

Figure 4-10
Parameter estimates for the model

		Estimate	Std. Error	Wald	df	Sig.
Threshold	[rating = 1]	-3.630	.335	117.579	1	.000
	[rating = 2]	-1.486	.302	24.265	1	.000
	[rating = 3]	1.533	.311	24.378	1	.000
Location	[hcrime=1]	-.643	.238	7.318	1	.007
	[hcrime=2]	0 ¹	.	.	0	.
	[sex=1]	-.424	.163	6.758	1	.009
	[sex=2]	0 ¹	.	.	0	.
	[age2=0]	.076	.176	.186	1	.666
	[age2=1]	0 ¹	.	.	0	.
	[educ5=1]	-1.518	.389	15.198	1	.000
	[educ5=2]	-1.256	.288	19.004	1	.000
	[educ5=3]	-.941	.310	9.188	1	.002
	[educ5=4]	-.907	.302	9.015	1	.003
	[educ5=5]	0 ¹	.	.	0	.

Link function: Logit.

1. This parameter is set to zero because it is redundant.

Measuring Strength of Association

There are several R^2 -like statistics that can be used to measure the strength of the association between the dependent variable and the predictor variables. They are not as useful as the R^2 statistic in regression, since their interpretation is not straightforward. Three commonly used statistics are:

- Cox and Snell R^2

$$R^2_{CS} = 1 - \left(\frac{L(\mathbf{B}^{(0)})}{L(\hat{\mathbf{B}})} \right)^{\frac{2}{n}}$$

- Nagelkerke's R^2

$$R^2_N = \frac{R^2_{CS}}{1 - L(\mathbf{B}^{(0)})^{2/n}}$$

■ McFadden's R^2

$$R^2_M = 1 - \left(\frac{L(\hat{\mathbf{B}})}{L(\mathbf{B}^{(0)})} \right)$$

where $L(\hat{\mathbf{B}})$ is the log-likelihood function for the model with the estimated parameters and $L(\mathbf{B}^{(0)})$ is the log-likelihood with just the thresholds, and n is the number of cases (sum of all weights). For this example, the values of all of the pseudo R -square statistics are small.

Figure 4-11
Pseudo R -square

Cox and Snell	.059
Nagelkerke	.066
McFadden	.027

Link function: Logit.

Classifying Cases

You can use the predicted probability of each response category to assign cases to categories. A case is assigned to the response category for which it has the largest predicted probability. Figure 4-12 is the classification table, which is obtained by crossstabulating *rating* by *pre_1*. (This is sometimes called the confusion matrix.)

Figure 4-12
Classification table

Rating of judges	Predicted Response Category		Total
	Only fair	Good	
Poor	15	36	51
Only fair	42	156	198
Good	33	246	279
Excellent	2	35	37
Total	92	473	565

Of the 198 people who selected the response *Only fair*, only 42 are correctly assigned to the category using the predicted probability. Of the 279 who selected *Good*, 246 are correctly assigned. None of the respondents who selected *Poor* or *Excellent* are correctly assigned. If the goal of your analysis is to study the association between the grouping variable and the predictor variables, the poor classification should not concern you. If your goal is to target marketing or collections efforts, the correct classification rate may be more important.

Generalized Linear Models

The ordinal logistic model is one of many models subsumed under the rubric of generalized linear models for ordinal data. The model is based on the assumption that there is a latent continuous outcome variable and that the observed ordinal outcome arises from discretizing the underlying continuum into j -ordered groups. The thresholds estimate these cutoff values.

The basic form of the generalized linear model is

$$\text{link}(\gamma_j) = \frac{\theta_j - [\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k]}{\exp(\tau_1 z_1 + \tau_2 z_2 + \dots + \tau_m z_m)}$$

where γ_j is the cumulative probability for the j th category, θ_j is the threshold for the j th category, $\beta_1 \dots \beta_k$ are the regression coefficients, $x_1 \dots x_k$ are the predictor variables, and k is the number of predictors.

The numerator on the right side determines the **location** of the model. The denominator of the equation specifies the scale. The $\tau_1 \dots \tau_m$ are coefficients for the scale component and $z_1 \dots z_m$ are m predictor variables for the scale component (chosen from the same set of variables as the x 's).

The **scale component** accounts for differences in variability for different values of the predictor variables. For example, if certain groups have more variability than others in their ratings, using a scale component to account for this may improve your model.

Link Function

The link function is the function of the probabilities that results in a linear model in the parameters. It defines what goes on the left side of the equation. It's the link between the random component on the left side of the equation and the systematic component

on the right. In the criminal rating example, the link function is the logit function, since the log of the odds results is equal to the linear combination of the parameters. That is,

$$\ln\left(\frac{\text{prob}(\text{event})}{(1 - \text{prob}(\text{event}))}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

Five different link functions are available in the Ordinal Regression procedure in SPSS. They are summarized in the following table. The symbol γ represents the probability that the event occurs. Remember that in ordinal regression, the probability of an event is redefined in terms of cumulative probabilities.

Function	Form	Typical application
Logit	$\ln\left(\frac{\gamma}{1 - \gamma}\right)$	Evenly distributed categories
Complementary log-log	$\ln(-\ln(1 - \gamma))$	Higher categories more probable
Negative log-log	$-\ln(-\ln(\gamma))$	Lower categories more probable
Probit	$\Phi^{-1}(\gamma)$	Analyses with explicit normally distributed latent variable
Cauchit (inverse Cauchy)	$\tan(\pi(\gamma - 0.5))$	Outcome with many extreme values

If you select the probit link function, you fit the model described in Chapter 5. The observed probabilities are replaced with the value of the standard normal curve below which the observed proportion of the area is found.

Probit and logit models are reasonable choices when the changes in the cumulative probabilities are gradual. If there are abrupt changes, other link functions should be used. The complementary log-log link may be a good model when the cumulative probabilities increase from 0 fairly slowly and then rapidly approach 1. If the opposite is true, namely that the cumulative probability for lower scores is high and the approach to 1 is slow, the negative log-log link may describe the data. If the complementary log-log model describes the probability of an event occurring, the log-log model describes the probability of the event not occurring.

Fitting a Heteroscedastic Probit Model

Probit models are useful for analyzing signal detection data. Signal detection describes the process of detecting an event in the face of uncertainty or “noise.” You must decide whether a signal is present or absent. For example, a radiologist has to decide whether a tumor is present or not based on inspecting images. You can model the uncertainty in the decision-making process by asking subjects to report how confident they are in their decision.

You postulate the existence of two normal distributions: one for the probability of detecting a signal when only noise is present and one for detecting the signal when both the signal and the noise are present. The difference between the means of the two distributions is called d , a measure of the sensitivity of the person to the signal.

The general probit model is

$$p(Y \leq k|X) = \Phi\left(\frac{c_k - d_n X}{\sigma_s^x}\right)$$

where Y is the dependent variable, such as a confidence rating, with values from 1 to K , X is a 0–1 variable that indicates whether the signal was present or absent, c_k are ordered distances from the noise distribution, d_n is the scaled distance parameter, and σ_s is the standard deviation of the signal distribution. The model can be rewritten as

$$\Phi^{-1}[p(Y \leq k|X)] = \frac{c_k - d_n X}{e^{ax}}$$

where Φ^{-1} is the inverse of the cumulative normal distribution and a is the natural log of σ_s . The numerator models the location; the denominator, the scale.

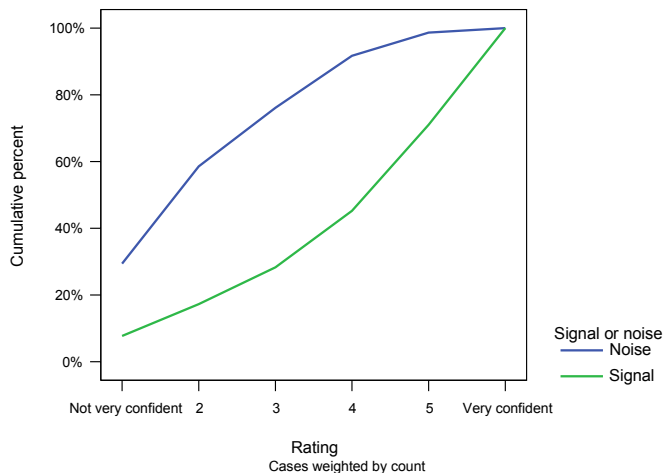
If the noise and signal distributions have different variances, you must include this information in the model. Otherwise, the parameter estimates are biased and inconsistent. Even large sample sizes won’t set things right.

Modeling Signal Detection

Consider data reported from a light detection study by Swets, et al. (1961) and discussed by DeCarlo (2003). Data are for a single individual who rated his confidence that a signal was present in 591 trials when the signal was absent and 597 trials when the signal was present.

In Figure 4-13, you see the cumulative distribution of the ratings under the two conditions (signal absent and signal present). The noise curve is above the signal curve, indicating that the low confidence ratings were more frequent when a signal was not present.

Figure 4-13
Plot of cumulative confidence ratings



Fitting a Location-Only Model

If you assume that the variance of the noise and signal distributions are equal, you can fit the usual probit model. Open the file *swets.sav*. The data are aggregated. For each possible combination of *signal* and *response*, there is a *count* of the number of times that each response was chosen.

You must weight the data file before proceeding. From the menus choose:

- Data
- Weight Cases...
 - Weight cases by
 - ▶ count

- Analyze
 - Regression
 - Ordinal...
- ▶ Dependent: response
- ▶ Covariate(s): signal
- Options...
 - Link: Probit
- Output...
 - Display
 - Parameter Estimates
 - Goodness of fit statistics

Examining the Goodness of Fit

Since the model has only 12 cells and none of them have zero frequencies, you can examine the goodness-of-fit statistics without concern that the expected counts are too small for the chi-square approximation to be valid. From Figure 4-14, you see that the model does not fit well. The observed significance level is less than 0.0005.

Figure 4-14
Goodness-of-fit statistics

	Chi-Square	df	Sig.
Pearson	33.728	4	.000
Deviance	32.972	4	.000

Link function: Probit.

One of the reasons the model may fit poorly is because the variance of the two distributions of responses may be different. You need a model that allows the variance of the underlying variable to vary as a function of one or more of the independent variables.

You can select a model for the standard deviation such that

$$\sigma_i = e^{Z_i \gamma}$$

where Z_i is a vector of covariates selected from the predictor variables.

Fitting a Scale Parameter

To fit a model that allows for different variances in the two groups, you must specify a model for the scale parameters, the denominator in the previous equation. To fit a model with different variances in the two groups, recall the dialog box and select:

Scale...

Scale model: signal

Because you have only one independent variable, *signal*, separate variances are estimated for each of the two categories of *signal*. If you have several predictor variables, you can specify a separate model for the scale component.

The goodness-of-fit statistics in Figure 4-15 indicate that the model fits much better than the location-only model. The variability of the distributions are an important consideration in this problem.

Figure 4-15

Goodness of fit with scale model

	Chi-Square	df	Sig.
Pearson	1.497	3	.683
Deviance	1.482	3	.687

Link function: Probit.

Parameter Estimates

When you fit a model with scale parameters as well as location parameters, parameter estimates for both are displayed.

Figure 4-16

Parameter estimates for model with location and scale parameters

		Parameter Estimates						
		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[response = 1]	-.533	.054	98.809	1	.000	-.638	-.428
	[response = 2]	.204	.050	16.979	1	.000	.107	.301
	[response = 3]	.710	.053	182.311	1	.000	.607	.813
	[response = 4]	1.366	.067	414.418	1	.000	1.235	1.498
	[response = 5]	2.294	.113	409.475	1	.000	2.072	2.516
Location	signal	1.519	.096	250.110	1	.000	1.331	1.707
Scale	signal	.348	.063	30.711	1	.000	.225	.472

Link function: Probit.

The threshold values are distances of the response criteria from the mean of the noise distribution. The location parameter estimate is the estimate of the detection parameter, d_n . To convert the scale parameter to an estimate of the ratio of the noise to signal standard deviations, you must compute $e^{-0.348}$, which is 0.71.

Model-Fitting Information

The overall test of the model is shown in Figure 4-17. When there is a scale parameter, the null hypothesis is that *both* the location parameters and the scale parameters are 0. A scale parameter of 0 means that the variances are equal. Based on the small observed significance level, you can reject this composite null hypothesis. Consult DeCarlo (2003) for further discussion of this example and for other examples of using the Ordinal Regression procedure in signal detection.

Figure 4-17
Model-fitting information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	461.699			
Final	59.261	402.437	2	.000

Link function: Probit.

