Chapter 1

Ordinal Outcomes Regression

1.1 Introduction

This is my best effort to succinctly explain the theory behind the ordinal logistic regression model (with apologies to the probit model).

The main takeaway point is supposed to be this:

> The same data leads to different estimates from different programs. That happens because the ordinal model can be written down in several different ways. None of them are wrong, but they are different, and as a result the user must be cautious.

Estimates obtained from four different programs are offered in Table 1.1. If we line these up side by side, we see that estimates from one of the routines for R matches Stata (after chopping off the small differences in the decimals), while SAS appears to provide the "wrong sign" for the first row and the second procedure for R seems to provide the "wrong signs" for the second and third rows.
None of these are actually wrong, they are all correct *given the model they specified.* This the point at which the student may be tempted to give up. Please don't. I've worked very hard to clear this up in the following sections.

Table 1.1: Ordinal Regression Results

	R: polr	R: lrm	SAS	Stata
$\hat{b}_1$	-0.28	-0.28	0.28	-0.28
$\hat{\zeta}_1$	-4.24	4.24	-4.24	-4.24
$\hat{\zeta}_2$	-2.32	2.32	-2.32	-2.32

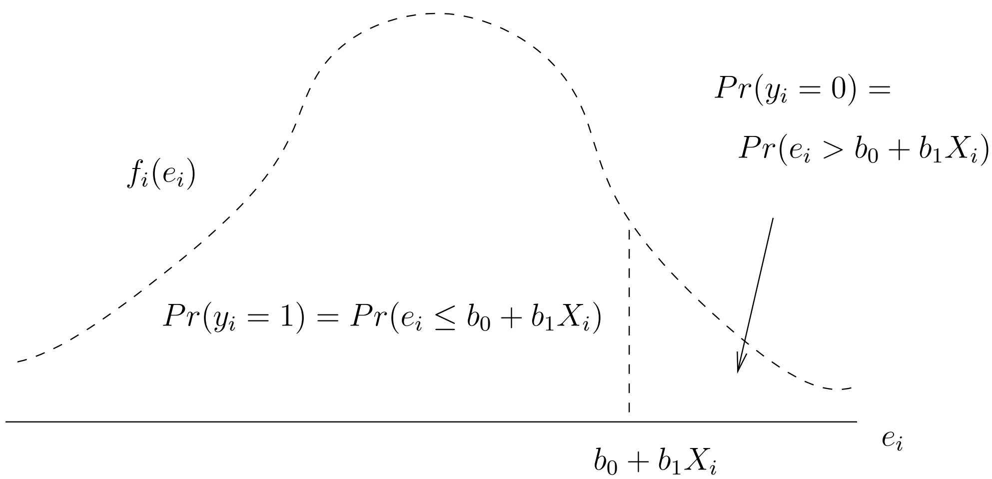

Figure 1.1: Dichotomous Outcome Variable

1.2 Extending the Logit Model to deal with Ordinal Dependent Variables

The easiest way to understand regression with ordinal dependent variables is to extend the "cumulative probability interpretation" of the two category model (**?**).

In the two category model, y_i is 1 with probability

$$F(b_0+b_1X_i) = \int_{-\infty}^{b_0+b_1X_i} f(e_i)de_i \tag{1.1}$$

And, of course, the probability that y_i is 0 will be $1-F(b_0+b_1X_i)$. The formula F is a "cumulative distribution function" (CDF), it represents the probability that a random variable e_i will be as small or smaller than $b_0+b_1X_i$. The function f is a "probability density function" (PDF), which represents the probability that e_i is equal to some particular value. This is illustrated in Figure 1.1. The "probability density function" f is defined from left to right and the possible outcomes are divided into two sets by the line drawn at $e_i = b_0+b_1X_i$. The area under the curve on the left side is the probability of getting a "yes" (or 1). The area on the right is the chance of a "no" (0).

Suppose y_i can have 3 values, 0, 1, and 2. (Keep in mind that this model can be written down in several ways. We tackle my favorite first, and then consider the others.) Leave the predictive part of the model ($b_0+b_1X_i$) the same, but we now introduce two new positive constants (Π_0 and Π_1) that divide the space. Considering Figure 1.2, it should be easy to see why some people call these new parameters "thresholds".

To summarize the effect of these new thresholds, we write down 1 equation for each possible outcome. My tendency is to write the thresholds as positive values like so:

$$y_i = \begin{cases} 2 & if\, b_0+b_1X_i-e_i \geq \Pi_1 \\ 1 & if\, \Pi_0 \leq b_0+b_1X_i-e_i < \Pi_1 \\ 0 & if\, b_0+b_1X_i-e_i < \Pi_0 \end{cases} \tag{1.2}$$

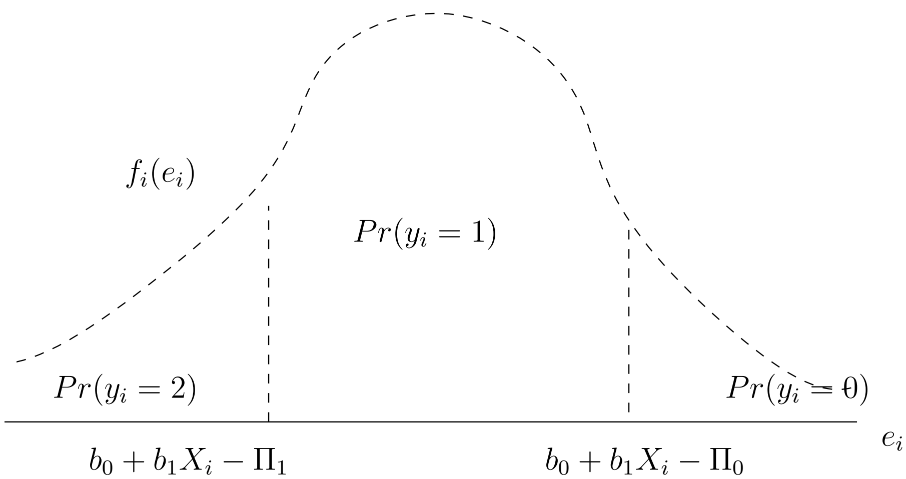

Figure 1.2: Ordinal Logit

Note we don't really need 3 equations. If we have two, say $Pr(y_i = 0)$ and $Pr(y_i = 1)$, then the chance of ending up in the other category is $1 - Pr(y_i = 0) - Pr(y_i = 1)$.

In order to translate this into a model involving the cumulative probability distribution, re-arrange so that the random variable e_i is by itself.

$$y_i = \begin{cases} 2 & if\, e_i \le b_0 + b_1X_i - \Pi_1 \\ 1 & if\, b_0 + b_1X_i - \Pi_1 < e_i \le b_0 + b_1X_i - \Pi_0 \\ 0 & if\, b_0 + b_1X_i - \Pi_0 < e_i \end{cases} \tag{1.3}$$

As in the dichotomous case, the probabilities of the various outcomes are calculated by use of cumulative probability. Rearrange 1.2 to convert these into probabilities of the individual outcomes.

$$\begin{array}{lll} Pr(y_i = 2) & = Pr(e_i \le b_0 + b_1X_i - \Pi_1) & = F(b_0 + b_1X_i - \Pi_1) \\ Pr(y_i = 1) & = Pr(b_0 + b_1X_i - \Pi_1 \le e_i < b_0 + b_1X_i - \Pi_0) & \\ & = 1 - F(b_0 + b_1X_i - \Pi_0) - F(b_0 + b_1X_i - \Pi_1) & \\ Pr(y_i = 0) & = Pr(b_0 + b_1X_i - \Pi_0 < e_i) & = 1 - F(b_0 + b_1X_i - \Pi_0) \end{array} \tag{1.4}$$

Note that any one category can be thought of as a "residual" category after the others have been assigned their shares. The middle category, $y_i = 1$, is left over if we "chop off" the outcomes on the left ($y_i = 2$) and the right ($y_i = 0$). We are left with the chance of ending up in the middle. In that sense, the probability of landing in the middle is equal to 1.0 minus the chance of a very small amount of random noise ($e_i \le b_0 + b_1X_i - \Pi_1$) and minus the chance of having a very large random noise ($b_0 + b_1X_i - \Pi_0 < e_i$). Similarly, the chances of being in the top category equal 1 minus the chance of ending up in the lower categories.

Any probability distribution can be used for the random error e_i, the two most common being Logistic and Normal. If the Normal is chosen, it is customary to call this a "probit" model and the symbol for the cumulative distribution is usually $\Phi()$.

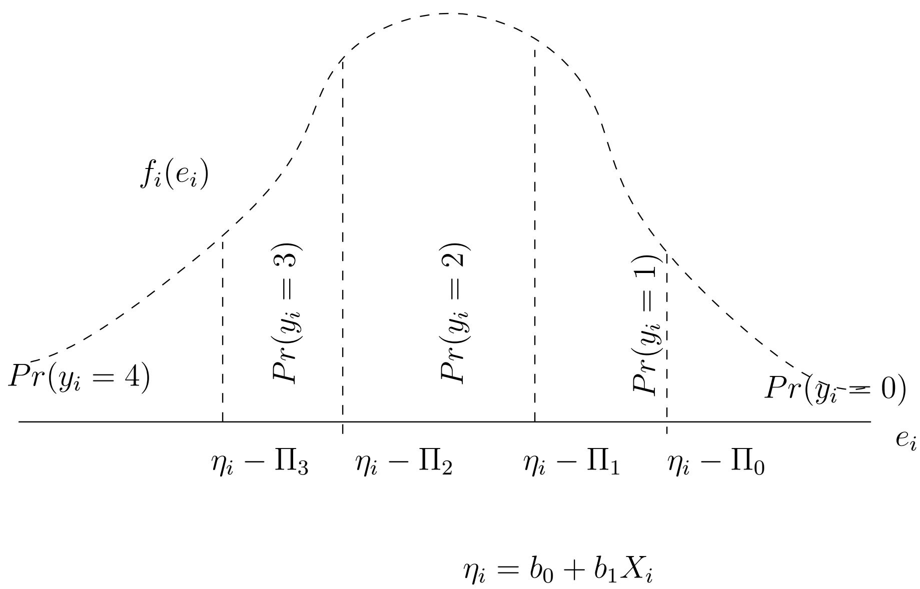

Figure 1.3: Ordinal Model with More Categories

What if your dependent variable have more categories? Add more thresholds! See the example in Figure 1.3.