Limited Dependent Variable

Aadi Dev S
Department of Economics
Deva Matha College, Kuravilangad

- **Till now we studied regression models for dichotomous response variables; however, many discrete response variables have three or more categories**

- **Examples ?**

An Interesting example would be to model student's intended career where the possibility can consist of any number of career (say 20)

Possible explanatory variables include gender, achievement test scores, and other variables in the data set.

Multinomial vs. Binary logistic regression

Many of the concepts used in binary logistic regression, such as the interpretation of parameters in terms of odds ratios and modeling probabilities, carry over to multi-category logistic regression models.

However, a major modifications are needed to deal with multiple categories of the response variable

- One difference is that with three or more levels of the response variable, there are multiple ways to dichotomize the response variable.

- If *J equals the number of categories* of the response variable, then *J(J –1)/2 different* ways exist to dichotomize the categories.
  For example let there be 3 categories A,B and C the dichotomized pairs would be AB, AC and BC

A second modification to extend binary logistic regression to the polytomous case is the
need for a more complex distribution for the response variable

In the binary case, the distribution of the response is assumed to be binomial; however, with multi category responses, the natural choice is the multinomial distribution,

How the response variable is dichotomized depends on

on the nature of the variable – If there is a baseline or control category, then the
analysis could focus on comparing each of the other categories to the baseline.

Odds Ratio

For a binary response variable, there is only one kind of odds that we may consider

$$\frac{\pi}{1-\pi}.$$

For a multi-category response variable with J > 2 categories and category probabilities($\pi_1, \pi_2, \ldots\ldots, \pi_j$); we may consider various kinds of odds, though some of them are more interpretable than others:

- odds between two categories: $\pi i / \pi j$
- odds between a group of categories vs another group of Categories

$$\frac{\pi_1 + \pi_3}{\pi_2 + \pi_4 + \pi_5}.$$

Odds Example for multi nominal :

E.g., if Y = source of meat (in a broad sense) with 5 categories
beef, pork, chicken, turkey, fish
We may consider the odds of

- beef vs. chicken: $\pi_{beef}/\pi_{chicken}$
- red meat vs. white meat:

$$\frac{\pi_{\text{beef}} + \pi_{\text{pork}}}{\pi_{\text{chicken}} + \pi_{\text{turkey}} + \pi_{\text{fish}}}$$

- red meat vs. poultry:

$$\frac{\pi_{\text{beef}} + \pi_{\text{pork}}}{\pi_{\text{chicken}} + \pi_{\text{turkey}}}$$

Odds for ordinal variables

If Y is ordinal with ordered categories:
1<2<3…..<J

we may consider the odds of Y ≤ J:

$$\frac{P(Y \le j)}{P(Y > j)} = \frac{\pi_1 + \pi_2 + \cdots + \pi_j}{\pi_{j+1} + \cdots + \pi_J}$$

e.g., Y = political ideology, with 5 levels
very liberal < slightly liberal < moderate < slightly conservative < very conservative

we may consider the odds

$$\frac{P(\text{very or slightly liberal''})}{P(\text{moderate or conservative})} = \frac{\pi_{\text{vlib}} + \pi_{\text{slib}}}{\pi_{\text{mod}} + \pi_{\text{scon}} + \pi_{\text{vcon}}}$$

Multi Nominal Logistic Regression

Baseline Model :

- Consider a high school program types data. There are three possible program types: academic, general, and vocational.
  Let *P(Yi = academic),*
  P(Yi = general),
  and P(Yi = vocational) be the probabilities of each of the program types for individual *I*

- *There is no natural or pre mentioned baseline or reference, so lets consider general program as our reference.*

- *Dichotomizing the categories we can make 3 pairs (General Academic, General Vocational and Academic Vocational)*

- only two of the three possible pairs of program types are needed because the third can be found by taking the product of the other two.

Choosing the general program as the reference, the odds of academic versus general and the odds of vocational versus general equal

$$\frac{P(Y_i = \text{academic})}{P(Y_i = \text{general})}$$

$$\frac{P(Y_i = \text{vocational})}{P(Y_i = \text{general})}.$$

The third odds, academic versus vocational, equals the product of these two odds

$$\frac{P(Y_i = \text{academic})}{P(Y_i = \text{vocational})} = \frac{P(Y_i = \text{academic})/P(Y_i = \text{general})}{P(Y_i = \text{vocational})/P(Y_i = \text{general})}.$$

- More generally, let *J equal the number of categories* or levels of the response variable. Of the *J(J* – 1)/2 possible pairs of categories, only (*J – 1) of* them are needed.

- If the same category is used in the denominator of the (*J – 1) odds*, and all other possible odds can be formed from this set

As a model for Odds

- Continuing our example, where the general program is chosen as the baseline category, consider the model containing a single explanatory variable, the mean of five achievement test scores for each student (i.e., math, science, reading, writing, and civics).

- The baseline model is simply two binary logistic regression models applied to each pair of program types; that is,

$$\frac{P(Yi=academic|xi)}{P(Yi=general|xi)} = e^{[\alpha 1+ \beta 1xi]}$$

&

$$\frac{P(Yi=vocational\ |xi)}{P(Yi=general\ |xi)} = e^{[\alpha 2+ \beta 2xi]}$$

where *P(Yi = academic|xi), P(Yi = general|xi),* and *P(Yi = vocational|xi) are the probabilities* for each program type given mean achievement test score *xi for student i, the* α*js are intercepts, and the* β*js are regression* coefficients.

- **The odds of academic versus vocational are found by taking the ratio of :**

$$\frac{P(Yi=academic|xi)}{P(Yi=vocational|xi)} = \frac{e^{[\alpha1+\beta1xi]}}{e^{[\alpha2+\beta2xi]}}$$

$$= e^{[(\alpha1-\alpha2)+(\beta1-\beta2)xi}$$

$$= e^{[\alpha3+\beta3xi]}$$

where $\alpha3 = (\alpha1 - \alpha2)$ and $\beta3 = (\beta1 - \beta2)$.

- **For generality, let *j = 1, . . . , J represent categories* of the response variable. The probability that individual *i is in category j given a value of xi on the explanatory* variable is represented by *P(Yi = j|xi)***

When fitting the baseline model to data, the binary logistic regressions for the *(J – 1)* odds must be estimated simultaneously to ensure that intercepts and coefficients for all other odds equal the differences of the corresponding intercepts and coefficients (e.g., $\alpha 3 = (\alpha 1 - \alpha 2)$ and $\beta 3 = (\beta 1 - \beta 2)$)

To demonstrate this, three separate binary logistic regression models were fit to the High School and Beyond data, as well as the baseline regression model, which simultaneously estimates the models for all the odds. The estimated parameters and their standard errors are reported in the Table

Table 26.1 Estimated Parameters (and Standard Errors) From Separate Binary Logistic Regressions and From the Simultaneously Estimated Baseline Model

		Separate Models		*Baseline Model*	
Odds	*Parameter*	*Estimate*	*SE*	*Estimate*	*SE*
$\frac{P(Y_i = \text{academic}\|x_i)}{P(Y_i = \text{general}\|x_i)}$	α_1	–5.2159	0.8139	–5.0391	0.7835
	β_1	0.1133	0.0156	0.1099	0.0150
$\frac{P(Y_i = \text{vocational}\|x_i)}{P(Y_i = \text{general}\|x_i)}$	α_2	2.9651	0.8342	2.8996	0.8156
	β_2	–0.0613	0.0172	–0.0599	0.0168
$\frac{P(Y_i = \text{academic}\|x_i)}{P(Y_i = \text{vocational}\|x_i)}$	α_3	–7.5331	0.8572	–7.9387	0.8439
	β_3	0.1618	0.0170	0.1698	0.0168

- Although the parameters for the separate and simultaneous cases are quite similar, the logical relationships between the parameters when the models are fit separately are not met (e.g., $\hat{\beta}_1 - \hat{\beta}_2 = 0.1133 + 0.0163 = 0.1746 \neq 0.1618$); however, the relationships hold for simultaneous estimation (e.g., $\hat{\beta}_1 - \hat{\beta}_2 = 0.1099 + 0.0599 = 0.1698$).

- A second advantage of simultaneous estimation is that it is a more efficient use of the data, which in turn leads to more powerful statistical hypothesis tests and more precise estimates of parameters. Notice that the parameter estimates in Table from the baseline model have smaller standard errors than those in the estimation of separate regressions.

Interpreting the regression coefficients

- Using the parameter estimates of the baseline model (column 5 of Table), the estimated odds that a student is from an academic program versus a general program given achievement score *x equals*

$$\frac{\hat{P}(Y_i=academic|x)}{\hat{P}(Y_i=general|x)} = e^{[-5.0391+0.1099x]}$$

and the estimated odds of an academic versus a general program for a student with achievement score *x + 1 equals*

$$\frac{\hat{P}(Y_i=academic|x+1)}{\hat{P}(Y_i=general|x+1)} = e^{[-5.0391+0.1099(x+1)]}$$

- The ratio of these two odds :

$$\frac{\hat{P}(Y_i = \text{academic}|x+1)\ \hat{P}(Y_i = \text{academic}|x)}{\hat{P}(Y_i = \text{general}|x+1)\ \hat{P}(Y_i = \text{general}|x)}$$

$$= \frac{\exp[-5.0391 + 0.1099(x+1)]}{\exp[-5.0391 + 0.1099x]}$$

$$= \exp(0.1099) = 1.12.$$

- This odds ratio is interpreted as follows: For a one-unit increase in achievement, the odds of a student attending an academic versus a general program are 1.12 times larger.

- For example, the odds of a student with *x = 50* attending an academic program versus a general one is 1.12 times the odds for a student with *x = 49.*

- Given the scale of the achievement variable (i.e., mean(*x)* = 51.99, *s = 8.09, min = 32.94,* and max = 70.00), it may be advantageous to report the odds ratio for an increase of one standard deviation of the explanatory variable rather than a one-unit increase.

- Generally, speaking, $e^{(\beta c)}$ *where c is a constant, equals* the odds ratio for an increase of *c units*

- For example, for an increase of one standard deviation in mean achievement, the odds ratio for academic versus general equals exp(0.1099(8.09)) = 2.42. Likewise, for a one standard deviation increase in achievement, the odds of an academic versus a vocational program are exp(0.1698(8.09)) = 3.95 times larger, but the odds of a vocational program versus a general program are only exp(–0.0599(8.09)) =0.62 times as large.

As a Model of Probabilities

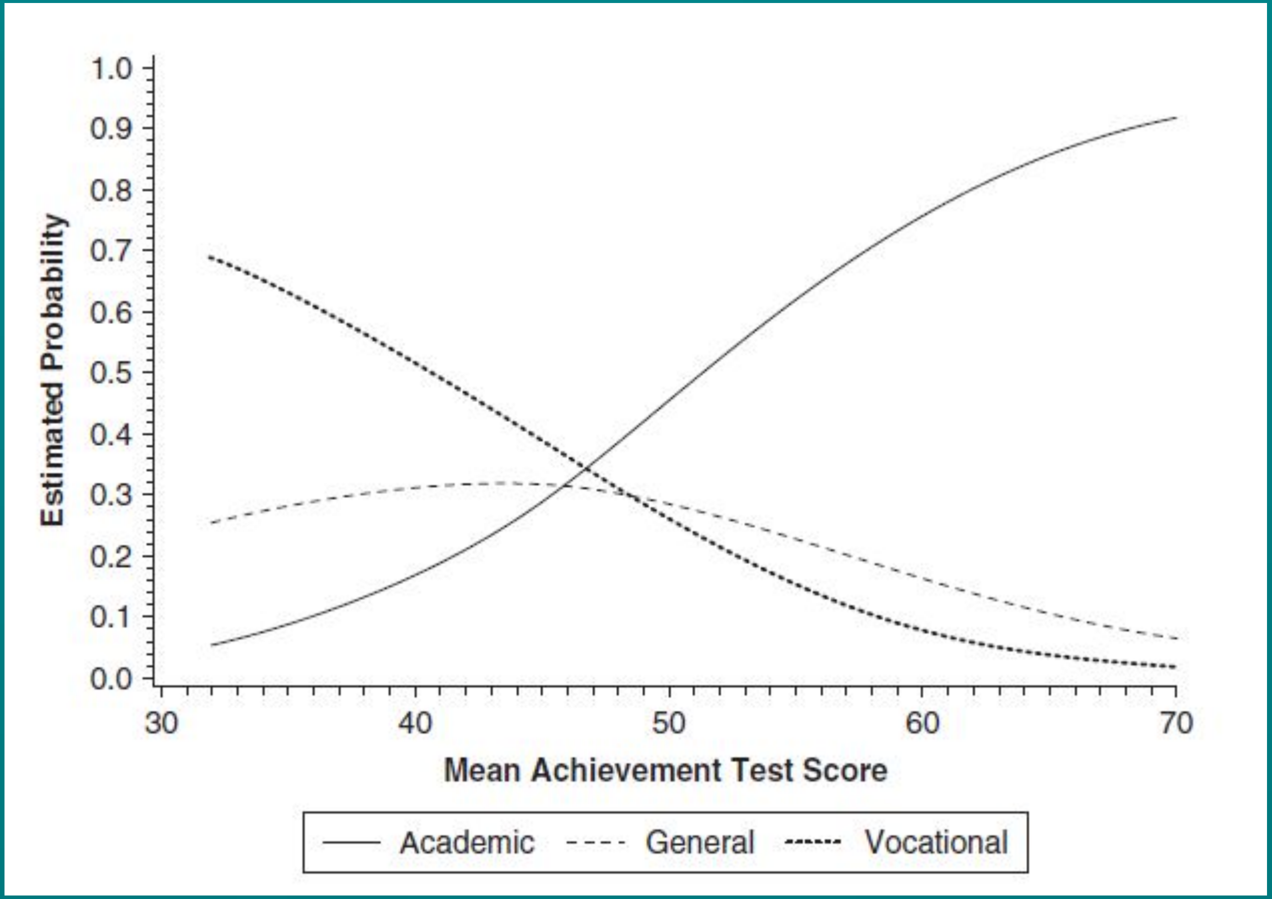
- Probabilities are generally a more intuitively understood concept than odds and odds ratios

- The model for probabilities is :

$$P(Y_i = j \mid x_i) = \frac{e^{[\alpha j + \beta j x i]}}{\sum_{h=1}^{J} e^{[\alpha h + \beta h x i]}}$$

where $j = 1, \ldots, J$.

The estimated probabilities are plotted in Figure :



The baseline model will always have one curve that monotonically decreases (e.g., *P(Yi = vocational|xi)) and one that monotonically* increases (e.g., *P(Yi = academic|xi)). All others* will increase and at some point start to decrease (e.g., *P(Yi = general|xi)). At any point along the* horizontal axis, the sum of the three probabilities equals 1.

What to do when there are multiple independent variables ?

Multiple Independent variable model

- Models with multiple explanatory variables are illustrated here by adding to our model a nominal (i.e.,whether the school a student attends is public or private) and an ordinal variable (i.e., socioeconomic status reported as low, middle, or high).

- Discrete variables are added using either dummy or effect coding. For example, school type could be coded either as a dummy variable (Equation 11a) or as an effect code (Equation 11b

$$p_i = \begin{cases} 1 & \text{if public} \\ 0 & \text{if private} \end{cases} \qquad (11a)$$

or

$$p_i = \begin{cases} 1 & \text{if public} \\ -1 & \text{if private} \end{cases} \qquad (11b)$$

- The model presented and developed here has 3 independent variable achievement, school type, and socioeconomic status (SES)

- The effects codes used to add SES, which has three levels, to the model are as follows:

$$s_{1i} = \begin{cases} 1 & \text{for low SES} \\ 0 & \text{for middle SES} \\ -1 & \text{for high SES} \end{cases}$$

and

$$s_{2i} = \begin{cases} 0 & \text{for low SES} \\ 1 & \text{for middle SES} \\ -1 & \text{for high SES} \end{cases}$$

- Defining *j = 1 for academic, j = 2 for vocational,* and *j = 3 = J for general program, the first* model with multiple explanatory variables examined here is

$$\frac{P(Yi=j|xi,pi,s1i,s2i)}{P(Yi=J|xi,pi,s1i,s2i)}=e^{[\alpha j+\ \beta jixi+\ \beta j2pi+\ \beta j3s1i+\ \beta j4s2i]}$$

- The same model expressed in terms of probabilities is

$$P\ (Y_i = j|\ x_i,\ p_i,\ s_{1i},\ s_{2i}) = \frac{e^{[\alpha j+\ \beta j1xi+\ \beta j2pi+\ \beta j3s1i+\ \beta j4s2i]}}{\sum_{h=1}^{J} e^{[\alpha h+\ \beta h1xi+\ \beta h2pi+\ \beta h2s1i+\ \beta h3s2i]}}$$

Table 26.2 Estimated Parameters, Standard Errors, and Wald Test Statistics for All Main Effects Model

Odds	*Effect*	*Parameter*	*Estimate*	*SE*	*exp(β)*	*Wald*	*p Value*
$\frac{P(Y_i = \text{academic})}{P(Y_i = \text{general})}$	Intercept	α_1	–3.92	0.83		22.06	< .01
	Achievement	β_{11}	0.10	0.02	1.10	37.80	< .01
	School type (public)	β_{12}	–0.61	0.18	0.54	12.01	< .01
	School type (private)	$-\beta_{12}$	0.61		1.84		
	SES (low)	β_{13}	–0.46	0.18	0.63	6.83	.01
	SES (middle)	β_{14}	–0.07	0.15	0.94	0.19	.66
	SES (high)	$-(\beta_{13} + \beta_{14})$	0.53		1.70		
$\frac{P(Y_i = \text{vocational})}{P(Y_i = \text{general})}$	Intercept	α_2	2.88	0.88		10.61	< .01
	Achievement	β_{13}	–0.06	0.02	0.94	13.28	< .01
	School type (public)	β_{22}	0.13	0.24	1.94	0.27	.60
	School type (private)	$-\beta_{22}$	–0.13		0.88		
	SES (low)	β_{23}	–0.23	0.19	0.80	1.45	.23
	SES (middle)	β_{24}	0.24	0.17	1.28	2.16	.14
	SES (high)	$-(\beta_{23} + \beta_{24})$	–0.02		0.98		

NOTE: SES is treated as a nominal variable.

- The interpretation in terms of odds ratios is the same as binary logistic regression

- Using the parameters reported in Table 26.2, for a one unit increase in mean achievement, the odds of an academic versus a general program are 1.10 times larger, and for a one standard deviation increase, the odds are exp(0.10(0.809)) = 2.25 times larger.

- With ordinal explanatory variables such as SES, one way to use the ordinal information is by assigning scores or numbers to the categories and treating the variables as numerical variables in the model

- Often, equally spaced integers are used, which amounts to putting equality restrictions on the βs for the variable. In our example, suppose we assign 1 to low SES, 2 to middle SES, and 3 to high SES and refit the model.

- Now SES can be denoted using one variable only

- Placing the restrictions on the βs for ordinal variables is often a good way to reduce the complexity of a model. For example, the estimated odds ratio of academic versus general for middle versus low SES equals $e^{\hat{\beta}_{13}(2-1)} = e^{\hat{\beta}_{13}} = 1.70$, which is the same as the odds ratio of high versus middle SES, $\exp(\hat{\beta}_{13}(3-2)) = 1.70$
  That is for two adjacent levels the odds ratio is the same and can be easily found out (refer table in the next slide for data)

Table 26.3 Estimated Parameters, Standard Errors, and Wald Statistics for All Main Effects Model

Odds	*Effect*	*Parameter*	*Estimate*	*SE*	*exp*(β)	*Wald*	*p Value*
$\frac{P(Y_i = \text{academic})}{P(Y_i = \text{general})}$	Intercept	α_1	–4.97	0.83	—	35.73	< .01
	Achievement	β_{11}	0.10	0.02	1.10	37.48	< .01
	School type	β_{12}	–0.61	0.18	0.55	11.80	< .01
	SES	β_{13}	0.53	0.18	1.70	11.80	< .01
$\frac{P(Y_i = \text{vocational})}{P(Y_i = \text{general})}$	Intercept	α_2	2.57	0.87	—	8.78	< .01
	Achievement	β_{13}	–0.06	0.02	0.95	12.96	< .01
	School type	β_{22}	0.12	0.24	1.13	0.26	.61
	SES	β_{23}	0.17	0.19	1.19	0.92	.34

NOTE: SES is treated as a numerical variable with scores of 1 = low, 2 = middle, and 3 = high.

But…

- In our example, putting in equally spaced scores for SES is not warranted and is misleading

- The order of the SES levels for the odds of academic (versus general) schools is in the expected order (i.e., the odds of an academic program are larger the higher the student's SES level) (Table 26.2)

- On the other hand, the parameter estimates of SES for odds of vocational schools do not follow the natural ordering of low to high, are relatively close together, and are not significantly different from zero.

- The numerical scores could be used for the SES effect on the odds of academic programs but the scores are inappropriate for the odds of vocational programs. There may not even be a difference between vocational and general programs in terms of SES. Furthermore, there may not be a difference between students who attended vocational and general programs with respect to school type