Multinomial and Ordinal Logistic Regression

ME104: Linear Regression Analysis
Kenneth Benoit

August 22, 2012

Regression with categorical dependent variables

When the dependent variable is categorical, with > 2 categories

Example: Which party did you vote for?

- Data from the European Social Survey (2002/2003), British sample
- Question: For which party did you vote in 2001? (Here we only consider the Conservatives, Labour, and the Liberal Democrats)

party

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Conservative	386	18.8	31.4	31.4
	Labour	624	30.4	50.8	82.2
	Liberal Democrat	218	10.6	17.8	100.0
	Total	1228	59.8	100.0	
Missing	other party/no answer	824	40.2		
Total		2052	100.0		

The multinomial logistic regression model

- We have data for n sets of observations ($i = 1, 2, \dots n$)
- Y is a categorical (polytomous) response variable with C categories, taking on values $0, 1, \dots, C-1$
- We have k explanatory variables $X_1, X_2, \dots, X_k$
- The multinomial logistic regression model is defined by the following assumptions:
  - Observations Y_i are statistically independent of each other
  - Observations Y_i are a random sample from a population where Y_i has a *multinomial* distribution with probability parameters: $\pi_i^{(0)}, \pi_i^{(1)}, \dots, \pi_i^{(C-1)}$
  - As with binomial logistic regression, we have to set aside one category for a *base* category (hence the $C-1$ parameters π)

The multinomial logistic regression model

The logit for each non-reference category $j = 1, \ldots, C-1$ against the *reference category* 0 depends on the values of the explanatory variables through:

$$\log\left(\frac{\pi_i^{(j)}}{\pi_i^{(0)}}\right) = \alpha^{(j)} + \beta_1^{(j)} X_{1i} + \cdots + \beta_k^{(j)} X_{ki}$$

for each $j = 1, \ldots, C-1$ where $\alpha^{(j)}$ and $\beta_1^{(j)}, \ldots, \beta_k^{(j)}$ are unknown population parameters

Multinomial distribution

$$\Pr(Y_1 = y_1, \ldots, Y_k = y_k) = \begin{cases} \frac{n!}{y_j! \cdots y_k!} \pi_1^{(0)} \cdots \pi_k^{(C-1)} & \text{when } \sum_{j=1}^{k} y_j = n \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} \mathrm{E}(y_{ij}) &= n\pi_j \\ \mathrm{Var}(y_{ij}) &= n\pi_j(1 - \pi_j) \end{aligned}$$

Example: vote choice

Response variable: Party voted for in 2001

- Labour is the reference category, $j = 0$
- Conservatives are the $j = 1$ category
- Liberal Democrats will be $j = 2$

(note that this coding is arbitrary)

Explanatory variables:

- Age: continuous in years X_1
- Educational level (categorical)
  - lower secondary or less (omitted reference category)
  - upper secondary ($X_2 = 1$)
  - post-secondary ($X_3 = 1$)

If we were to fit binary logistic models

One model for the log odds of voting Convervative v. Labour:

$$\log\left(\frac{\pi_i^{(1)}}{\pi_i^{(0)}}\right) = \alpha^{(1)} + \beta_1^{(1)} X_{1i} + \beta_2^{(1)} X_{2i} + \beta_3^{(1)} X_{3i}$$

A second model for the log odds of voting Lib Dem v. Labour:

$$\log\left(\frac{\pi_i^{(2)}}{\pi_i^{(0)}}\right) = \alpha^{(2)} + \beta_1^{(2)} X_{1i} + \beta_2^{(2)} X_{2i} + \beta_3^{(2)} X_{3i}$$

$$\log\left(\frac{\pi_i^{(LD)}}{\pi_i^{(Lab)}}\right) = \alpha^{(LD)} + \beta_1^{(LD)} \text{age}_i + \beta_2^{(LD)} \text{upper\_sec}_i + \beta_3^{(LD)} \text{post\_sec}_i$$

Estimates of this model

Parameter Estimates

party[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B) Lower Bound	95% Confidence Interval for Exp(B) Upper Bound
Conservative	Intercept	-1.861	.265	49.147	1	.000			
	age	.021	.004	24.804	1	.000	1.021	1.013	1.029
	[educ=post_sec]	.638	.151	17.807	1	.000	1.892	1.407	2.544
	[educ=upper_sec]	.474	.225	4.422	1	.035	1.606	1.033	2.499
	[educ=lower_sec]	0[b]	.	.	0	.	.	.	.
Liberal Democrat	Intercept	-1.809	.316	32.796	1	.000			
	age	.005	.005	1.044	1	.307	1.005	.995	1.015
	[educ=post_sec]	1.026	.181	32.031	1	.000	2.791	1.956	3.982
	[educ=upper_sec]	.746	.263	8.068	1	.005	2.108	1.260	3.527
	[educ=lower_sec]	0[b]	.	.	0	.	.	.	.

a. The reference category is: Labour.

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

Interpreting $\hat{\beta}$: continuous X

Parameter Estimates

party		B	Exp(B)
Conservative	Intercept	-1.861	
	age	.021	1.021
	[educ=post_sec]	.638	1.892
	[educ=upper_sec]	.474	1.606
	[educ=lower_sec]	0	.
Liberal Democrat	Intercept	-1.809	
	age	.005	1.005
	[educ=post_sec]	1.026	2.791
	[educ=upper_sec]	.746	2.108
	[educ=lower_sec]	0	.

- Holding education level constant, a one-year increase in age multiplies the odds of voting Conservative rather than Labour by 1.021, i.e. increases them by 2.1%
- A five-year increase in age (controlling for education level) multiplies the odds of voting Conservative rather than Labour by $\exp(5 * 0.021) = 1.021^5 = 1.11$, i.e. increases them by 11%
- Holding education level constant, a one-year increase in age multiplies the odds of voting Lib Dem rather than Labour by 1.005, i.e. increases them by 0.5%

Interpreting $\hat{\beta}$: categorical X

Parameter Estimates

party		B	Exp(B)
Conservative	Intercept	-1.861	
	age	.021	1.021
	[educ=post_sec]	.638	1.892
	[educ=upper_sec]	.474	1.606
	[educ=lower_sec]	0	.
Liberal Democrat	Intercept	-1.809	
	age	.005	1.005
	[educ=post_sec]	1.026	2.791
	[educ=upper_sec]	.746	2.108
	[educ=lower_sec]	0	.

- Holding age constant, the odds for someone with post-secondary education of voting Conservative rather than Labour are 1.892 times (89.2% higher than) the odds for someone with lower secondary or less education
- Holding age constant, the odds for someone with upper secondary education of voting Conservative rather than Labour are 1.606 times (60.6% higher than) the odds for someone with lower secondary or less education

Interpreting $\hat{\beta}$: categorical X

Parameter Estimates

party		B	Exp(B)
Conservative	Intercept	-1.861	
	age	.021	1.021
	[educ=post_sec]	.638	1.892
	[educ=upper_sec]	.474	1.606
	[educ=lower_sec]	0	.
Liberal Democrat	Intercept	-1.809	
	age	.005	1.005
	[educ=post_sec]	1.026	2.791
	[educ=upper_sec]	.746	2.108
	[educ=lower_sec]	0	.

- Holding age constant, the odds for someone with post-secondary education of voting Lib Dem rather than Labour are 2.791 times (179.1% higher than) the odds for someone with lower secondary or less education
- Holding age constant, the odds for someone with upper secondary education of voting Lib Dem rather than Labour are 2.108 times (110.8% higher than) the odds for someone with lower secondary or less education

Interpreting $\hat{\beta}$ between non-reference categories of X

Parameter Estimates

party		B	Exp(B)
Conservative	Intercept	-1.861	
	age	.021	1.021
	[educ=post_sec]	.638	1.892
	[educ=upper_sec]	.474	1.606
	[educ=lower_sec]	0	.
Liberal Democrat	Intercept	-1.809	
	age	.005	1.005
	[educ=post_sec]	1.026	2.791
	[educ=upper_sec]	.746	2.108
	[educ=lower_sec]	0	.

- Holding age constant, the odds for someone with post-secondary education of voting Conservative rather than Labour are 1.178 times (17.8% higher than) the odds for someone with upper secondary education
  Calculation: $\exp(0.638 - 0.474) = \exp(0.164)$ or $1.892/1.606$
- Holding age constant, the odds for someone with upper secondary education of voting Lib Dem rather than Labour are 0.756 times (24.4% lower than) the odds for someone with post-secondary education
  Calculation: $\exp(0.746 - 1.026) = \exp(-0.28)$ or $2.108/2.791$

Interpreting $\hat{\beta}$ between non-reference categories of Y

Parameter Estimates

party		B	Exp(B)
Conservative	Intercept	-1.861	
	age	.021	1.021
	[educ=post_sec]	.638	1.892
	[educ=upper_sec]	.474	1.606
	[educ=lower_sec]	0	.
Liberal Democrat	Intercept	-1.809	
	age	.005	1.005
	[educ=post_sec]	1.026	2.791
	[educ=upper_sec]	.746	2.108
	[educ=lower_sec]	0	.

$$\log\left(\frac{\pi_i^{(j)}}{\pi_i^{(1)}}\right) = (\alpha^{(1)} - \alpha^{(j)}) + (\beta_1^{(j)}\beta_1^{(1)})X_{1i} + \cdots + (\beta_k^{(j)} - \beta_k^{(1)})X_{ki}$$

for each $j = 2, \ldots, C-1$

Holding age constant, the odds for someone with post-secondary education of voting Lib Dem rather than Conservative are 1.47 times (47.4% higher than) the odds for someone with lower secondary education

Calculation: $\exp(1.026 - 0.638) = \exp(0.388)$ or $2.791/1.892$

Computing fitted probabilities

- We fit a logit model for each non-reference category j
- Let $L(j) = \log(\pi_i(j)/|pi_i(0))$ — the log odds of a response in category j rather than the reference category 0
- Probability of response in category j can be calculated as

$$\pi^{(j)} = P(Y = j) = \frac{\exp(L^{(j)})}{1 + \exp(L^{(1)}) + \cdots + \exp(L^{(C-1)})}$$

- Probability of response in category 0 can be calculated as

$$\pi^{(0)} = P(Y = 0) = \frac{1}{1 + \exp(L^{(1)}) + \cdots + \exp(L^{(C-1)})}$$

Fitted probabilities from the example

(with voting Labour as the reference category)

- Logit for voting Conservative rather than Labour:

$$\begin{aligned} L^{(\text{Cons})} &= \log(\pi_i^{(\text{Cons})}/\pi_i^{(\text{Lab})}) \\ &= -1.861 + 0.021 * \text{age} + 0.474 * \texttt{upper\_sec} + 0.638 * \texttt{post\_se} \end{aligned}$$

- Logit for voting Liberal Democrat rather than Labour:

$$\begin{aligned} L^{(\text{Lib})} &= \log(\pi_i^{(\text{Lib})}/\pi_i^{(\text{Lab})}) \\ &= -1.809 + 0.005 * \text{age} + 0.746 \text{upper\_sec} + 1.026 * \texttt{post\_sec} \end{aligned}$$

- Estimated logits for (for example), a 55-year old with upper secondary education:

$$\begin{aligned} L(\text{Cons}) &= -1.861 + 0.021 * (55) + 0.474 * (1) + 0.638 * (0) = -0.232 \\ L(\text{Lib}) &= -1.809 + 0.005 * (55) + 0.746 * (1) + 1.026 * (0) = -0.788 \end{aligned}$$

More fitted probabilities from the example

- Probability of 55 year old with upper secondary education voting Conservative:

$$\hat{\pi}^{(\text{Cons})} = \frac{\exp(-0.232)}{1 + \exp(-0.232) + \exp(-0.788)} = \frac{0.793}{2.248} = 0.35$$

- Probability of 55 year old with upper secondary education voting Liberal Democrat:

$$\hat{\pi}^{(\text{Lib})} = \frac{\exp(-0.788)}{1 + \exp(-0.232) + \exp(-0.788)} = \frac{0.455}{2.248} = 0.20$$

- Probability of 55 year old with upper secondary education voting Labour:

$$\hat{\pi}^{(\text{Lab})} = \frac{1}{1 + \exp(-0.232) + \exp(-0.788)} = \frac{1}{2.248} = 0.44$$

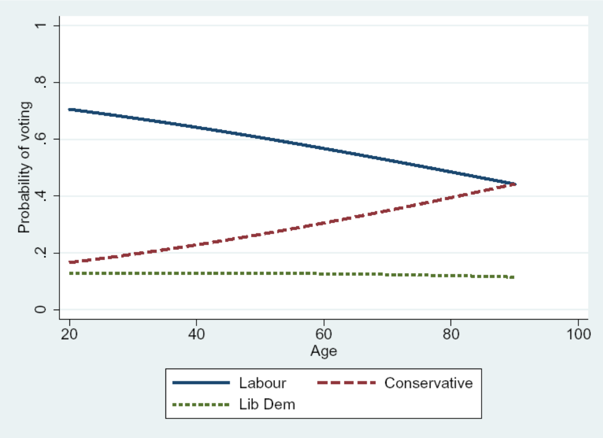
for a categorical explanatory variable

Fitted probabilities of party choice given education, with age fixed at 55 years:

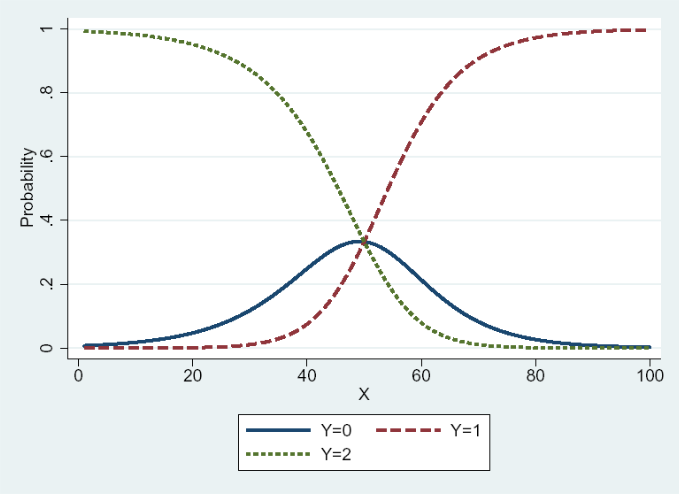
	Lower secondary	Upper secondary	Post-secondary
Conservative	0.29	0.35	0.36
Lib Dem	0.13	0.20	0.24
Labour	0.59	0.44	0.39

for a continuous explanatory variable

Fitted probabilities of party choice given age, with education fixed at lower secondary or less:



for all three response categories



Confidence intervals for $\hat{\beta}$

- These are calculated just as for binomial logits, using $\pm 1.96\hat{\sigma}_{\hat{\beta}}$
- So the 95% confidence interval for an estimated coefficient is:

$$\hat{\beta}^{(j)} \pm 1.96\,\hat{\text{se}}(\hat{\beta}^{(j)})$$

- and the 95% confidence interval for an odds ratio is:

$$\left(\exp[\hat{\beta}^{(j)} - 1.96\,\hat{\text{se}}(\hat{\beta}^{(j)})];\ \exp[\hat{\beta}^{(j)} + 1.96\,\hat{\text{se}}(\hat{\beta}^{(j)})]\right)$$

Wald tests for $\hat{\beta}$

- Wald tests are provided in the SPSS output by default
- Here, testing H_0: $\beta^{(j)} = 0$
  i.e. null hypothesis that a given X has no effect on odds of $Y = j$ versus $Y = 0$
- But we often want to test the null hypothesis that a given X has no effect on odds of any category of the response variable, e.g.
  $H_0 : \beta^{(1)} = \beta^{(2)} = \cdots = \beta^{(C-1)} = 0$
- We can use likelihood ratio comparison test, in the usual way, to test several coefficients at once

Likelihood ratio comparison tests

- Reminder: we compare two models:
  - Model 1 is the simpler, restricted, model, with likelihood L_1
  - Model 2 is the more complex, full, model, with likelihood L_2
  - Must be nested: so Model 2 is Model 1 with some extra parameters
- H_0: more complex model is no better than simpler one; then L_1 and L_2 will be similar, i.e. difference between them will be small
- Likelihood ratio test statistic:

$$D = 2(\log L_2 - \log L_1) = (-2\log L_1) - (-2\log L_2)$$

- Obtain p-value for tests statistic from χ^2 distribution with degrees of freedom equal to the difference in the degrees of freedom in the two models (i.e. the number of extra parameters in the larger model)

Likelihood ratio comparison tests: Example

We can test a more restricted model excluding age.

$$H_0 : \beta_{\text{age}}^{(\text{Cons})} = \beta_{\text{age}}^{(\text{Lib})} = 0$$

- -2 log likelihood of Model 1, without age = 977.718
- -2 log likelihood of Model 2, including age = 951.778
- Difference in -2 log likelihoods = 25.940
- Difference in degrees of freedom = 2
- *p*-value for 25.940 on χ^2 distribution with 2 d.f. < 0.001
- Reject H_0; keep age in the model

Likelihood ratio comparison tests: Example

We can test a more restricted model excluding education.

$$H_0 : \beta^{(\mathrm{Cons})}_{\mathrm{educ2}} = \beta^{(\mathrm{Lib})}_{\mathrm{educ2}} = \beta^{(\mathrm{Cons})}_{\mathrm{educ3}} = \beta^{(\mathrm{Lib})}_{\mathrm{educ3}} = 0$$

- -2 log likelihood of Model 1, without education = 992.019
- -2 log likelihood of Model 2, including education = 951.778
- Difference in -2 log likelihoods = 40.241
- Difference in degrees of freedom = 4
- *p*-value for 40.241 on χ^2 distribution with 4 d.f. < 0.001
- Reject H_0; keep education in the model

Ordinal response variables: An example

- Data from the U.S. General Social Survey in 1977 and 1989
- Response variable Y is the answer to the following item:
  "A working mother can establish just as warm and secure a relationship with her children as a mother who does not work."
  - 1=Strongly disagree (SD), 2=Disagree (D), 3=Agree (A) and 4=Strongly agree (SA)
- In this and many other examples, the categories of the response have a natural ordering
- A multinomial logistic model can be used here too, but it has the disadvantage of ignoring the ordering
- An **ordinal logistic model** (proportional odds model) does take the ordering into account
- This gives a model with fewer parameters to interpret

Cumulative probabilities

- Suppose response variable Y has C **ordered** categories $j = 1, 2, \ldots, C$, with probabilities

$$P(Y = j) = \pi^{(j)} \quad \text{for} \;\; j = 1, \ldots, C$$

- In multinomial logistic model, we considered the $C - 1$ ratios

$$P(Y = j)/P(Y = 1) = \pi^{(j)}/\pi^{(1)} \quad \text{for} \;\; j = 2, \ldots, C$$

  and wrote down a model for each of them
- Now we will consider the $C - 1$ **cumulative probabilities**

$$\gamma^{(j)} = P(Y \leq j) = \pi^{(1)} + \cdots + \pi^{(j)} \quad \text{for} \;\; j = 1, \ldots, C - 1$$

  and write down a model for each of them
  - Note that $\gamma^{(C)} = P(Y \leq C) = 1$ always, so it need not be modelled

The ordinal logistic model

- Data: $(Y_i, X_{1i}, \ldots, X_{ki})$ for observations $i = 1, \ldots, n$, where
  - Y is a response variable with C ordered categories $j = 1, \ldots, C$, and probabilities $\pi^{(j)} = P(Y = j)$
  - $X_1, \ldots, X_k$ are k explanatory variables
- Observations Y_i are statistically independent of each other
- The following holds for $\gamma_i^{(j)} = P(Y_i \leq j)$ for each unit i and each category $j = 1, \ldots, C-1$:

$$\log\left(\frac{\gamma_i^{(j)}}{1-\gamma_i^{(j)}}\right) = \log\left(\frac{P(Y_i \leq j)}{P(Y_i > j)}\right) = \alpha^{(j)} - (\beta_1 X_{1i} + \cdots + \beta_k X_{ki})$$

The ordinal logistic model

- In other words, the ordinal logistic model considers a set of dichotomies, one for each possible cut-off of the response categories into two sets, of “high” and “low” responses
- This is meaningful only if the categories of Y do have an ordering
- In our example, these cut-offs are
  - Strongly disagree vs. (Disagree, Agree, or Strongly agree),
    i.e. SD vs. (D, A, or SA)
  - (SD or D) vs. (A or SA)
  - (SD, D, or SA) vs. SA
- A binary logistic model is then defined for the log-odds of each of these cuts

Parameters of the model

The model for the cumulative probabilities is

$$\gamma^{(j)} = P(Y \leq j) = \frac{\exp[\alpha^{(j)} - (\beta_1 X_1 + \cdots + \beta_k X_k)]}{1 + \exp[\alpha^{(j)} - (\beta_1 X_1 + \cdots + \beta_k X_k)]}$$

The intercept terms must be $\alpha^{(1)} < \alpha^{(2)} < \cdots < \alpha^{(C-1)}$, to guarantee that $\gamma^{(1)} < \gamma^{(2)} < \cdots < \gamma^{(C-1)}$

$\beta_1, \beta_2, \ldots, \beta_k$ are the *same* for each value of j

- There is thus only one set of regression coefficients, not $C - 1$ as in a multinomial logistic model
- The curves for $\gamma^{(1)}, \ldots, \gamma^{(C-1)}$ are "parallel" as seen below
- This is the assumption of "proportional odds". The ordinal logistic model is also known as the **proportional odds model**

Probabilities from the model

The probabilities of individual categories are

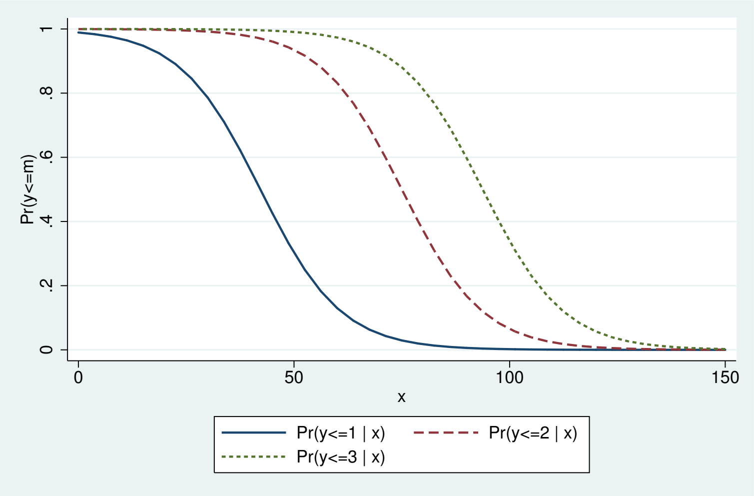
$$
\begin{aligned}
P(Y=1) &= \gamma^{(1)} = \frac{\exp[\alpha^{(1)} - (\beta_1 X_1 + \cdots + \beta_k X_k)]}{1 + \exp[\alpha^{(1)} - (\beta_1 X_1 + \cdots + \beta_k X_k)]} \\
P(Y=j) &= \gamma^{(j)} - \gamma^{(j-1)} = \frac{\exp[\alpha^{(j)} - (\beta_1 X_1 + \cdots + \beta_k X_k)]}{1 + \exp[\alpha^{(j)} - (\beta_1 X_1 + \cdots + \beta_k X_k)]} \\
&\qquad - \frac{\exp[\alpha^{(j-1)} - (\beta_1 X_1 + \cdots + \beta_k X_k)]}{1 + \exp[\alpha^{(j-1)} - (\beta_1 X_1 + \cdots + \beta_k X_k)]}
\end{aligned}
$$

for $j = 2, \ldots, C-1$, and

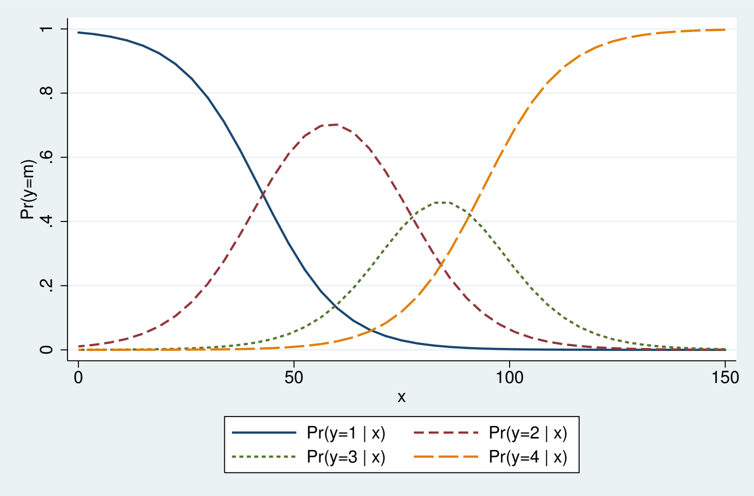
$$
P(Y=C) = 1 - \gamma^{(C-1)} = 1 - \frac{\exp[\alpha^{(C-1)} - (\beta_1 X_1 + \cdots + \beta_k X_k)]}{1 + \exp[\alpha^{(C-1)} - (\beta_1 X_1 + \cdots + \beta_k X_k)]}
$$

Illustrated below with plots for a case with $C = 4$ categories.

Cumulative probabilities: An example



Probabilities of individual categories: An example



Estimation and inference

Everything here is unchanged from binary logistic models:

- Parameters are estimated using maximum likelihood estimation
- Hypotheses of interest are typically of the form $\beta_j = 0$, for one or more coefficients β_j
- Wald tests, likelihood ratio tests and confidence intervals are defined and used as before

Back to the example

Response variable Y (variable `warm`): *"A working mother can establish just as warm and secure a relationship with her children as a mother who does not work."*, with levels SD, D, A, and SA

Explanatory variables:

- `yr89`: a dummy variable for survey year 1989 (1=1989, 0=1977)
- `white`: a dummy variable for ethnic group white (1=white, 0=non-white)
- `age` in years
- `ed`: years of education
- `male`: a dummy variable for men (1=Male, 2=Female)

Fitted model: An example

```
. ologit warm yr89 white age ed male

Ordered logistic regression                       Number of obs   =        22
                                                  LR chi2(5)      =      298.
                                                  Prob > chi2     =      0.00
Log likelihood = -2846.6132                       Pseudo R2       =      0.04

------------------------------------------------------------------------------
        warm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interva
-------------+----------------------------------------------------------------
        yr89 |   .5282808   .0798763     6.61   0.000     .3717262    .68483
       white |  -.3795009   .1182501    -3.21   0.001    -.6112669   -.14773
         age |  -.0207738   .0024195    -8.59   0.000    -.0255159   -.01603
          ed |   .0839738   .0131433     6.39   0.000     .0582135    .10973
        male |  -.7269441   .0783997    -9.27   0.000    -.8806048   -.57328
-------------+----------------------------------------------------------------
       /cut1 |  -2.443735   .2386412                     -2.911463   -1.9760
       /cut2 |  -.6096001   .2331233                     -1.066513   -.15268
       /cut3 |   1.279352   .2338585                      .8209981    1.7377
------------------------------------------------------------------------------
```

Interpretation of the coefficients

- Exponentiated coefficients are interpreted as partial odds ratios for being in the *higher* rather than the lower half of the dichotomy
  - here (SA, A, or D) vs. SD, (SA or A) vs. (D or SD), and SA vs. (A, D, or SD)
  - odds ratio is the same for each of these
- e.g. $\exp(\hat{\beta}_{male}) = 0.48$: Controlling for the other explanatory variables, men have 52% lower odds than women of giving a response that indicates higher levels of agreement with the statement
- e.g. $\exp(\hat{\beta}_{ed}) = 1.088$: Controlling for the other explanatory variables, 1 additional year of education is associated with a 8.8% increase in odds of giving a response that indicates higher levels of agreement with the statement

Example with an interaction

Consider adding an interaction between sex and education.

Just to show something new, include it in the form of two variables:

```
gen male_ed=male*ed
gen fem_ed=(1-male)*ed
```

instead of using `ed` and `male*ed` as previously

- Both versions give the same model
- In this version, the coefficients of `male_ed` and `fem_ed` are the coefficients of education for men and women respectively

Example with an interaction

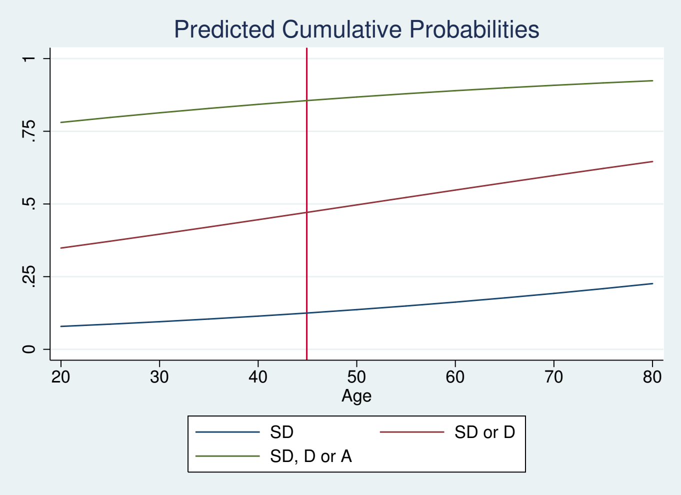
This version of interaction can be tested using a likelihood ratio test as before:

```
ologit warm yr89 white age male ed
estimates store mod1
ologit warm yr89 white age male male_ed fem_ed
lrtest mod1 .

Likelihood-ratio test                    LR chi2(1)  =      4.48
(Assumption: mod1 nested in .)           Prob > chi2 =    0.0344
```

or with a Wald test of the hypothesis that $\beta_{\text{male\_ed}} = \beta_{\text{fem\_ed}}$:

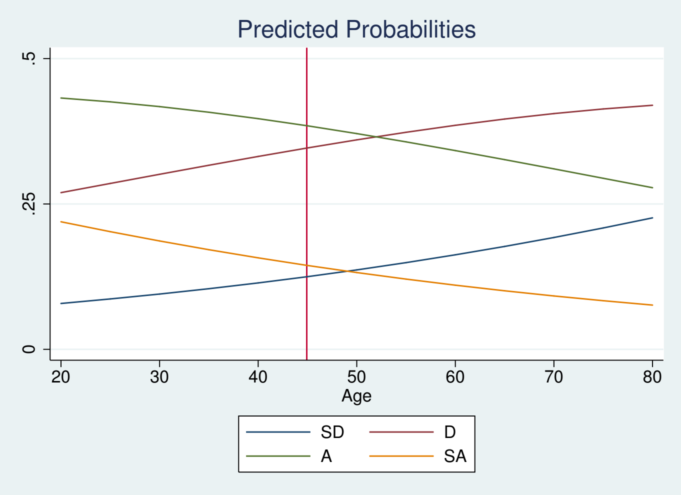
```
test male_ed=fem_ed

 ( 1)  [warm]male_ed - [warm]fem_ed = 0
           chi2(  1) =    4.47
         Prob > chi2 =    0.0345
```

Example: Fitted probabilities



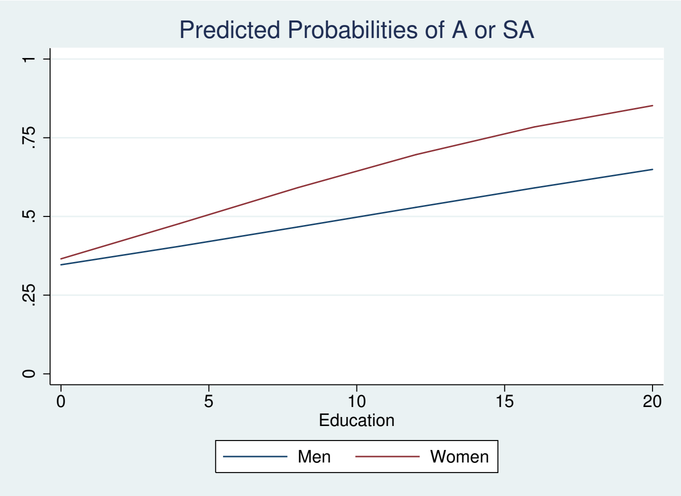
(Other variables fixed at: year 1989, white, male, 12 years of education)

Example: Fitted probabilities



(Other variables fixed at: year 1989, white, male, 12 years of education)

Example: Fitted probabilities



(year 1989, white, male, 45 years old)

Assessing the proportional odds assumption

- The ordinal logistic model

$$\log[\gamma^{(j)}/(1-\gamma^{(j)})] = \alpha^{(j)} - (\beta_1 X_1 + \dots \beta_k X_k)$$

  assumes the same coefficients $\beta_1, \dots, \beta_k$ for each cut-off j
- This is good for the parsimony of the model, because it means that the effect of an explanatory variable on the ordinal response is described by one parameter
- However, it is also a restriction on the flexibility of the model, which may or may not be adequate for the data
- There are a number of ways of checking the assumption
- Here we consider briefly only one, comparing (informally) estimates and fitted probabilities between ordinal and multinomial logistic model

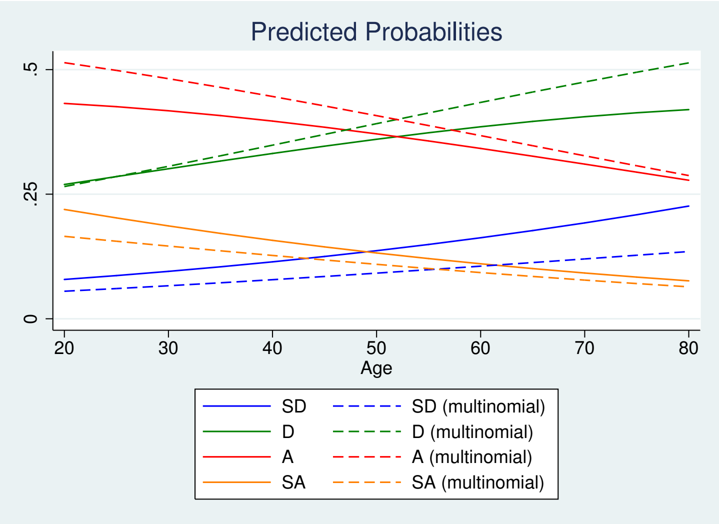
Multinomial logistic model in the example

```
. mlogit warm yr89 white age male male_ed fem_ed, base(1)
------------------------------------------------------------------------------
        warm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
2D           |
        yr89 |   .7358761   .1656333     4.44   0.000     .4112408    1.060511
       white |  -.4431508   .2464251    -1.80   0.072    -.9261352    .0398336
         age |  -.0038748   .0043447    -0.89   0.372    -.0123901    .0046406
        male |  -.2007845   .5167223    -0.39   0.698    -1.213542    .8119726
     male_ed |   .0780207   .0276012     2.83   0.005     .0239233    .1321181
      fem_ed |   .0538995   .0371462     1.45   0.147    -.0189057    .1267047
       _cons |   .6184256   .5438483     1.14   0.255    -.4474975    1.684349
-------------+----------------------------------------------------------------
3A           |
        yr89 |   1.097829   .1637353     6.70   0.000      .776914    1.418745
       white |  -.5317257   .2456104    -2.16   0.030    -1.013113   -.0503381
         age |  -.0245649   .0043948    -5.59   0.000    -.0331785   -.0159512
        male |  -.3240701   .5411645    -0.60   0.549    -1.384733    .7365928
     male_ed |   .1182817   .0289517     4.09   0.000     .0615373     .175026
      fem_ed |   .1214095   .0370969     3.27   0.001     .0487008    .1941181
       _cons |   1.060008   .5468476     1.94   0.053    -.0117936     2.13181
-------------+----------------------------------------------------------------
4SA          |
        yr89 |   1.159963   .1811322     6.40   0.000     .8049508    1.514976
       white |  -.8293461   .2633927    -3.15   0.002    -1.345586   -.3131058
         age |  -.0306687   .0051191    -5.99   0.000    -.0407019   -.0206354
        male |  -.3319753   .6680166    -0.50   0.619    -1.641264    .9773132
     male_ed |   .1173827    .038081     3.08   0.002     .0427454    .1920201
      fem_ed |   .1865388   .0403865     4.62   0.000     .1073827     .265695
       _cons |   .3026709   .6050227     0.50   0.617    -.8831518    1.488494
------------------------------------------------------------------------------
(warm==1SD is the base outcome)
```

Ordinal vs. multinomial models in the example

- In the multinomial model, (more or less) all the coefficients at least imply the same ordering of the categories
  - e.g. for age: 0 (SD) > -0.004 (D) > -0.025 (A) > -0.031 (SA)
  - this is not always the case: e.g. an example in the computer class
- For fitted probabilities in our example:
  - for most of the variables, the agreement in the strengths of association is reasonable if not perfect — see plot for age below
- The biggest difference is for period (1977 vs. 1989):
  - the ordinal model forces the change (in cumulative odds) to be the same for all cut-off points
  - according the multinomial model, some categories have shifted more than others between the two periods
  - in particular, the shift away from "Strongly disagree" has been a bit stronger than the ordinal model allows

Ordinal vs. multinomial: Fitted probabilities



Ordinal vs. multinomial: Fitted probabilities

Model	Year	SD	D	A	SA
Ordinal	1977	0.19	0.41	0.31	0.09
	1989	0.13	0.35	0.38	0.14
Multinomial	1977	0.19	0.40	0.32	0.08
	1989	0.08	0.37	0.43	0.12

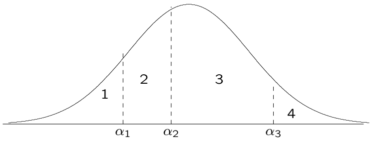
(white man, aged 45, 12 years of education)

Latent-variable motivation of the model

- The ordinal logit model can also be derived from a model for a hypothetical unobserved (latent) continuous response variable Y^*
- Suppose

$$Y^* = \beta_1 x_1 + \ldots + \beta_k x_k + \epsilon,$$

  where ϵ is a random error term which has *standard logistic distribution*: a “bell-shaped” distribution quite similar to the normal, with mean 0 and variance $\pi^2/3$
- Suppose that we do not observe Y^* but a grouped version Y:



- In other words, we record $Y = j$ if Y^* is in the interval $\alpha_{j-1} \leq Y^* < \alpha_j$ (where $\alpha_0 = -\infty$ and $\alpha_C = \infty$)

Latent-variable motivation of the model

- Then the model for Y (rather than Y^*) is an ordinal logit (proportional odds model)
- This derivation in terms of a hypothetical underlying Y^* is sometimes useful for motivating the model and deriving some of its properties
  - and sometimes Y^* even makes substantive sense
- Other models also have analogous latent-variable derivations
  - if $C = 2$ (only one cut-off point), we get the *binary* logit model
  - if ϵ is assumed to have a standard normal (rather than logistic) distribution, we get the (ordinal or binary) probit model
  - the latent-variable motivation of the multinomial logistic model is completely different