UL Summer School: Categorical Data Analysis

Brendan Halpin, Sociology
2022 Summer School

Outline

Association in tables

Association in tables

Association in tables

- Tables display association between categorical variables
- Made evident by patterns of percentages
- Tested by χ^2 test

Association

How do we characterise association?

- Is there association?
- What form does it take?
- How strong is it?

Q1: Is there association?

- This is what the χ^2 test determines – *evidence of association*
- Does not characterise nature or size!
- Depends on N
- Other tests exist, such as Fisher's exact test

Q2: What form does it take?

- Examine percentages
- Compare observed and expected: residuals
- *Standardised* residuals: behave like z, i.e., should lie in range $-2 : +2$ about 95% of time, if independence is true

$$z = \frac{O-E}{\sqrt{E(1-\text{row proportion})(1-\text{col proportion})}}$$

$$= \frac{O-E}{\sqrt{E(1-\frac{R}{T})(1-\frac{C}{T})}}$$

Q3: How strong is it?

Many possible measures of association

- Difference in proportions
- Ratio of proportions or "relative rate"
- Ratio of odds or "odds ratio"

(see `http://teaching.sociology.ul.ie:3838/apps/orrr/`)

Ordinal variables

- Ordinal variables may have more structured association
- Simpler pattern, analogous to correlation
- X high, Y high; X low, Y low

Characterising ordinal association

- Focus on concordant/discordant pairs
- Pairs of cases which differ on both variables
  - Concordant: case that is higher on one variable also higher on other
  - Discordant: higher on one, lower on the other
- Gamma, $\hat{\gamma} = \frac{C-D}{C+D}$
- Values range $-1 \leq \gamma \leq +1$
- Like correlation in interpretation
- Has asymptotic standard error $\Rightarrow$ t-test possible

Higher order tables

- We can consider association in higher-order tables, e.g., 3-way
- Is the association between A and B the same for different values of C?
- Does the association between A and B disappear1 if we control for C?

Simpson's paradox etc.

- Scouting example (ch 10): negative association between scouting and delinquency
- Control for family characteristics (church attendance) and it disappears
- See also death penalty example: note pattern of odds ratios
- Cochran-Mantel-Haenszel test: $2 \times 2 \times k$ table
- H_0 : within each of k 2×2 panels, OR = 1

Scouting 1/3

scout	delinq Yes	delinq No	Total
Yes	36	364	400
No	60	340	400
Total	96	704	800

Scouting 2/3

	church and delinq					
	--- Low --		-- Med ---		-- High --	
scout	Yes	No	Yes	No	Yes	No
Yes	10	40	18	132	8	192
No	40	160	18	132	2	48

Scouting 3/3

scout	church Low	church Med	church High	Total
Yes	50	150	200	400
No	200	150	50	400
Total	250	300	250	800

Loglinear modelling

- More complex questions and larger tables can be handled by loglinear modelling
- Treats all variables as "dependent variables"
- Can test null hypothesis of independence, as well as specified patterns of interaction

Logistic regression

Logistic regression

Logistic regression

- OLS regression requires interval dependent variable
- Binary or "yes/no" dependent variables are not suitable
- Nor are rates, e.g., n successes out of m trials
- Errors are distinctly not normal
- While predicted value can be read as a probability, can depart from 0:1 range
- Particular difficulties with multiple explanatory variables.

Linear Probability Model

- OLS gives the "linear probability model" in this case:

$$Pr(Y = 1) = a + bX$$

- data is 0/1, prediction is probability
- Assumptions violated, but if predicted probabilities in range 0.2–0.8, not too bad
- See credit card example: becomes unrealistic only at very low or high income

Logistic transformation

- Probability is bounded [0 : 1]
- OLS predicted value is unbounded
- How to transform probability to $-\infty : \infty$ range?
- Odds: $\frac{p}{1-p}$ – range is $0 : \infty$
- Log of odds: $\log \frac{p}{1-p}$ has range $-\infty : \infty$

Logistic regression

- Logistic regression uses this as the dependent variable:

$$\log\left(\frac{Pr(Y = 1)}{1 - Pr(Y = 1)}\right) = a + bX$$

- Alternatively:

$$\frac{Pr(Y = 1)}{1 - Pr(Y = 1)} = e^{a+bX}$$

- Or:

$$Pr(Y = 1) = \frac{e^{a+bX}}{1 + e^{a+bX}} = \frac{1}{1 + e^{-a-bX}}$$

Parameters

- The b parameter is the effect of a unit change in X on $\log\left(\frac{Pr(Y=1)}{1-Pr(Y=1)}\right)$
- This implies a multiplicative change of e^b in $\frac{Pr(Y=1)}{1-Pr(Y=1)}$, in the Odds
- Thus an odds ratio
- But the effect of b on P depends on the level of b
- See credit card example
- Death penalty example allows us to see the link between odds ratios and estimates

Logistic regression

Inference

Inference

- In practice, inference is similar to OLS though based on a different logic
- For each explanatory variable, $H_0 : \beta = 0$ is the interesting null
- $z = \frac{\hat{\beta}}{SE}$ is approximately normally distributed (large sample property)
- More usually, the Wald test is used: $\left(\frac{\hat{\beta}}{SE}\right)^2$ has a χ^2 distribution with one degree of freedom

Likelihood ratio tests

- The "likelihood ratio" test is thought more robust than the Wald test for smaller samples
- Where l_0 is the likelihood of the model without X_j, and l_1 that with it, the quantity

$$-2\left(\log \frac{l_0}{l_1}\right) = -2\left(\log l_0 - \log l_1\right)$$

  is χ^2 distributed with one degree of freedom

LR test in practice

```
. qui logit univ c.age##c.age i.sex

. est store base

. logit univ c.age##c.age i.sex i.gold

Iteration 0:   log likelihood = -258.63227
Iteration 1:   log likelihood = -235.46647
Iteration 2:   log likelihood = -224.18885
Iteration 3:   log likelihood = -223.79947
Iteration 4:   log likelihood = -223.79762
Iteration 5:   log likelihood = -223.79762

Logistic regression                                 Number of obs =    998
                                                    LR chi2(7)    =  69.67
                                                    Prob > chi2   = 0.0000
Log likelihood = -223.79762                         Pseudo R2     = 0.1347

------------------------------------------------------------------------------
           univ | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
----------------+-------------------------------------------------------------
            age |   .2135413   .0556893     3.83   0.000     .1043923    .3226903
                |
    c.age#c.age |  -.0025071   .0006445    -3.89   0.000    -.0037704   -.0012439
                |
            sex |
         female |  -.5470423   .2591863    -2.11   0.035    -1.055038   -.0390465
                |
           gold |
            RNM |  -1.241583   .5610744    -2.21   0.027    -2.341268   -.1418974
           Prop |  -1.388413   .3982332    -3.49   0.000    -2.168936   -.6078902
        Skilled |  -1.519483   .3206528    -4.74   0.000    -2.147951   -.8910149
Un/semi-skilled |  -2.334295   .4599521    -5.08   0.000    -3.235785   -1.432806
                |
          _cons |  -5.155577   1.135296    -4.54   0.000    -7.380716   -2.930438
------------------------------------------------------------------------------

. lrtest base

Likelihood-ratio test
Assumption: base nested within .

 LR chi2(4) =  43.01
Prob > chi2 = 0.0000
```


Nested models

- More generally, $-2\left(\log \frac{l_0}{l_1}\right)$ tests nested models: where model 1 contains all the variables in model 0, plus m extra ones, it tests the null that all the extra βs are zero (χ^2 with m df)
- If we compare a model against the null model (no explanatory variables, it tests

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$$

- Strong analogy with F test in OLS

Logistic regression

Maximum likelihood

Maximum likelihood estimation

- What is this "likelihood"?
- Unlike OLS, logistic regression (and many, many other models) are extimated by *maximum likelihood estimation*
- In general this works by choosing values for the parameter estimates which maximise the probability (likelihood) of observing the actual data
- OLS can be ML estimated, and yields exactly the same results

Iterative search

- Sometimes the values can be chosen analytically
  - A likelihood function is written, defining the probability of observing the actual data given parameter estimates
  - Differential calculus derives the values of the parameters that maximise the likelihood, for a given data set
- Often, such "closed form solutions" are not possible, and the values for the parameters are chosen by a systematic computerised search (multiple iterations)
- Extremely flexible, allows estimation of a vast range of complex models within a single framework

Likelihood as a quantity

- Either way, a given model yields a specific maximum likelihood for a give data set
- This is a probability, henced bounded [0 : 1]
- Reported as log-likelihood, hence bounded $[-\infty : 0]$
- Thus is usually a large negative number
- Where an iterative solution is used, likelihood at each stage is usually reported – *normally* getting nearer 0 at each step

Logistic regression

Tabular data

Tabular data

- If all the explanatory variables are categorical (or have few fixed values) your data set can be represented as a table
- If we think of it as a table where each cell contains n yeses and $m-n$ noes (n successes out of m trials) we can fit grouped logistic regression
- n successes out of m trials implies a binomial distribution of degree m

$$\log \frac{n}{m-n} = \alpha + \beta X$$

- The parameter estimates will be exactly the same as if the data were treated individually

Tabular data and goodness of fit

- But unlike with individual data, we can calculate goodness of fit, by relating observed successes to predicted in each cell
- If these are close we cannot reject the null hypothesis that the model is incorrect (i.e., you want a high p-value)
- Where l_i is the likelihood of the current model, and l_s is the likelihood of the "saturated model" the test statistic is

$$-2\left(\log \frac{l_i}{l_s}\right)$$

- The saturated model predicts perfectly and has as many parameters as there are "settings" (cells in the table)
- The test has *df* of number of settings less number of parameters estimated, and is χ^2 distributed

Logistic regression

Goodness of fit and accuracy of classification

Fit with individual data

- Where the number of "settings" (combinations of values of explanatory variables) is large, this approach to fit is not feasible
- Cannot be used with continuous covariates
- Hosmer-Lemeshow statistic attempts to create an analogy
  - Divide sample into deciles of predicted probability
  - Calculate a fit measure based on observed and predicted numbers in the ten groups
  - Simulation shows this is χ^2 distributed with 2 df
  - Not a perfect solution, sensitive to how the cuts are made
- Pseudo-R^2 measures exist, but none approaches the clean interpretation as in OLS
- See `http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm`

Predicting outcomes

- Another way of assessing the adequacy of a logit model is its accuracy of classification:

	True yes	True no
Predicted yes	a	c
Predicted no	b	d

- Proportion correctly classified: $\frac{a+d}{a+b+c+d}$
- Sensitivity: $\frac{a}{a+b}$; Specificity: $\frac{d}{c+d}$
- False positive: $\frac{c}{a+c}$; False negative: $\frac{b}{b+d}$
- Stata: `estat class`

Some problems

- Zero cells in tables can cause problems: no yeses or no noes for particular settings
- Not automatically a problem but can give rise to attempts to estimate a parameter as $-\infty$ or $+\infty$
- If this happens, you will see a large parameter estimate and a huge standard error
- In individual data, sometimes certain combinations of variables have only successes or only failures
- In Stata, these cases are dropped from estimation – you need to be aware of this as it changes the interpretation (you may wish to drop one of the offending variables instead)

Multinomial logistic regression

Baseline-category extension of binary logistic

What if we have multiple possible outcomes, not just two?

- Logistic regression is binary: yes/no
- Many interesting dependent variables have multiple categories
  - voting intention by party
  - first destination after second-level education
  - housing tenure type
- We can use binary logistic by
  - recoding into two categories
  - dropping all but two categories
- But that would lose information

Multinomial logistic regression

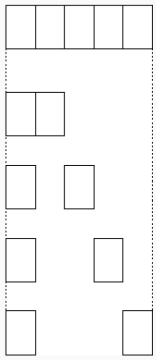
- Another idea:
- Pick one of the J categories as baseline
- For each of $J-1$ other categories, fit binary models contrasting that category with baseline
- Multinomial logistic effectively does that, fitting $J-1$ models simultaneously

$$\log \frac{P(Y=j)}{P(Y=J)} = \alpha_j + \beta_j X, \;\; j = 1, \ldots, c-1$$

- Which category is baseline is not critically important, but better for interpretation if it is reasonably large and coherent (i.e. "Other" is a poor choice)

$J-1$ contrasts

Compare each of
$J-$ categories
against a baseline



Predicting p from formula

$$\log \frac{\pi_j}{\pi_J} = \alpha_j + \beta_j X$$

$$\frac{\pi_j}{\pi_J} = e^{\alpha_j + \beta_j X}$$

$$\pi_j = \pi_J e^{\alpha_j + \beta_j X}$$

$$\pi_J = 1 - \sum_{k=1}^{J-1} \pi_k = 1 - \pi_J \sum_{k=1}^{J-1} e^{\alpha_k + \beta_k X}$$

$$\pi_J = \frac{1}{1 + \sum_{k=1}^{J-1} e^{\alpha_k + \beta_k X}} = \frac{1}{\sum_{k=1}^{J} e^{\alpha_k + \beta_k X}}$$

$$\Rightarrow \pi_j = \frac{e^{\alpha_j + \beta_j X}}{\sum_{k=1}^{J} e^{\alpha_k + \beta_k X}}$$

Multinomial logistic regression

Interpreting example, inference

Example

- Let's attempt to predict housing tenure
  - Owner occupier
  - Local authority renter
  - Private renter
- using age and employment status
  - Employed
  - Unemployed
  - Not in labour force
- `mlogit ten3 age i.eun`

Stata output

```
Multinomial logistic regression                 Number of obs     =      15490
                                                LR chi2(6)        =    1256.51
                                                Prob > chi2       =     0.0000
Log likelihood = -10204.575                     Pseudo R2         =     0.0580

------------------------------------------------------------------------------
        ten3 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1            |  (base outcome)
-------------+----------------------------------------------------------------
2        age |  -.0103121   .0012577    -8.20   0.000     -.012777   -.0078471
             |
         eun |
          2  |   1.990774   .1026404    19.40   0.000     1.789603    2.191946
          3  |    1.25075   .0522691    23.93   0.000     1.148304    1.353195
             |
       _cons |  -1.813314   .0621613   -29.17   0.000    -1.935148    -1.69148
-------------+----------------------------------------------------------------
3        age |  -.0389969   .0018355   -21.25   0.000    -.0425945   -.0353994
             |
         eun |
          2  |   .4677734   .1594678     2.93   0.003     .1552223    .7803245
          3  |   .4632419    .063764     7.26   0.000     .3382668    .5882171
             |
       _cons |    -.76724   .0758172   -10.12   0.000     -.915839   -.6186411
------------------------------------------------------------------------------
```


Interpretation

- Stata chooses category 1 (owner) as baseline
- Each panel is similar in interpretation to a binary regression on that category versus baseline
- Effects are on the log of the odds of being in category j versus the baseline

Inference

- At one level inference is the same:
  - Wald test for $H_o : \beta_k = 0$
  - LR test between nested models
- However, each variable has $J-1$ parameters
- Better to consider the LR test for dropping the variable across all contrasts: $H_0 : \forall j \, : \, \beta_j k = 0$
- Thus retain a variable even for contrasts where it is insignificant as long as it has an effect overall
- Which category is baseline affects the parameter estimates but not the fit (log-likelihood, predicted values, LR test on variables)

Predicting ordinal outcomes

- While `mlogit` is attractive for multi-category outcomes, it is imparsimonious
- For nominal variables this is necessary, but for ordinal variables there should be a better way
- We consider three useful models
  - Stereotype logit
  - Proportional odds logit
  - Continuation ratio or sequential logit
- Each approaches the problem is a different way

Ordinal logit

Stereotype logit

Stereotype logit

- If outcome is ordinal we should see a pattern in the parameter estimates:

```
. mlogit educ c.age i.sex if age>30
[...]
Multinomial logistic regression                   Number of obs   =      10905
                                                  LR chi2(4)      =    1171.90
                                                  Prob > chi2     =     0.0000
Log likelihood = -9778.8701                       Pseudo R2       =     0.0565

------------------------------------------------------------------------------
        educ |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Hi           |
         age |  -.0453534   .0015199   -29.84   0.000    -.0483323   -.0423744
       2.sex |  -.4350524   .0429147   -10.14   0.000    -.5191636   -.3509411
       _cons |   2.503877    .086875    28.82   0.000     2.333605    2.674149
-------------+----------------------------------------------------------------
Med          |
         age |  -.0380206   .0023874   -15.93   0.000    -.0426999   -.0333413
       2.sex |  -.1285718   .0674878    -1.91   0.057    -.2608455    .0037019
       _cons |   .5817336   .1335183     4.36   0.000     .3200425    .8434246
-------------+----------------------------------------------------------------
Lo           |  (base outcome)
------------------------------------------------------------------------------
```


Ordered parameter estimates

- Low education is the baseline
- The effect of age:
  - -0.045 for high vs low
  - -0.038 for medium vs low
  - 0.000, implicitly for low vs low
- Sex: -0.435, -0.129 and 0.000
- Stereotype logit fits a scale factor ϕ to the parameter estimates to capture this pattern

Scale factor

- Compare `mlogit`:

$$\log \frac{P(Y=j)}{P(Y=J)} = \alpha_j + \beta_{1j}X_1 + \beta_{2j}X,\ j = 1, \ldots, J-1$$

- with `slogit`

$$\log \frac{P(Y=j)}{P(Y=J)} = \alpha_j + \phi_j\beta_1X_1 + \phi_j\beta_2X_2,\ \ j = 1, \ldots, J-1$$

- ϕ is zero for the baseline category, and 1 for the maximum
- It won't necessarily rank your categories in the right order: sometimes the effects of other variables do not coincide with how you see the ordinality

Slogit example

- Age and sex predicting education for those 30yrs-plus

```
. slogit educ age i.sex if age>30
[...]
Stereotype logistic regression                    Number of obs   =      10905
                                                  Wald chi2(2)    =     970.21
Log likelihood =  -9784.863                       Prob > chi2     =     0.0000

 ( 1)  [phi1_1]_cons = 1
------------------------------------------------------------------------------
        educ |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0457061   .0015099    30.27   0.000     .0427468    .0486654
       2.sex |   .4090173   .0427624     9.56   0.000     .3252045    .4928301
-------------+----------------------------------------------------------------
     /phi1_1 |          1  (constrained)
     /phi1_2 |   .7857325   .0491519    15.99   0.000     .6893965    .8820684
     /phi1_3 |          0  (base outcome)
-------------+----------------------------------------------------------------
     /theta1 |   2.508265   .0869764    28.84   0.000     2.337795    2.678736
     /theta2 |   .5809221    .133082     4.37   0.000     .3200862     .841758
     /theta3 |          0  (base outcome)
------------------------------------------------------------------------------
(educ=Lo is the base outcome)
```


Interpreting ϕ

- With low education as the baseline, we find ϕ estimates thus:

High	1
Medium	0.786
Low	0

- That is, averaging across the variables, the effect of medium vs low is 0.786 times that of high vs low
- The `/theta` terms are the α_js

Surprises from `slogit`

- `slogit` is not guaranteed to respect the order
- if we include younger people as well as those over 30, lifecourse and cohort effects mean age has a non-linear effect
- $\Rightarrow$changes the order of ϕ

```
. slogit educ age i.sex
[...]
Stereotype logistic regression                    Number of obs   =      14321
                                                  Wald chi2(2)    =     489.72
Log likelihood =  -13792.05                       Prob > chi2     =     0.0000

 ( 1)  [phi1_1]_cons = 1
------------------------------------------------------------------------------
        educ |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0219661   .0009933    22.11   0.000     .0200192    .0239129
       2.sex |   .1450657   .0287461     5.05   0.000     .0887244    .2014071
-------------+----------------------------------------------------------------
     /phi1_1 |          1  (constrained)
     /phi1_2 |   1.813979   .0916542    19.79   0.000     1.634341    1.993618
     /phi1_3 |          0  (base outcome)
-------------+----------------------------------------------------------------
     /theta1 |   .9920811   .0559998    17.72   0.000     .8823235    1.101839
     /theta2 |   .7037589   .0735806     9.56   0.000     .5595436    .8479743
     /theta3 |          0  (base outcome)
------------------------------------------------------------------------------
(educ=Lo is the base outcome)
```

Recover by including non-linear age

```
Stereotype logistic regression                  Number of obs     =      14321
                                                Wald chi2(3)      =     984.66
Log likelihood = -13581.046                     Prob > chi2       =     0.0000

 ( 1)  [phi1_1]_cons = 1
------------------------------------------------------------------------------
        educ |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.1275568   .0071248   -17.90   0.000    -.1415212   -.1135924
             |
 c.age#c.age |   .0015888   .0000731    21.74   0.000     .0014456    .0017321
             |
       2.sex |   .3161976   .0380102     8.32   0.000     .2416989    .3906963
-------------+----------------------------------------------------------------
     /phi1_1 |          1  (constrained)
     /phi1_2 |   .5539747   .0479035    11.56   0.000     .4600854    .6478639
     /phi1_3 |          0  (base outcome)
-------------+----------------------------------------------------------------
     /theta1 |  -1.948551   .1581395   -12.32   0.000    -2.258499   -1.638604
     /theta2 |  -2.154373    .078911   -27.30   0.000    -2.309036   -1.999711
     /theta3 |          0  (base outcome)
------------------------------------------------------------------------------
(educ=Lo is the base outcome)
```

sociology

Stereotype logit

- Stereotype logit treats ordinality as ordinality in terms of the explanatory variables
- There can be therefore disagreements between variables about the pattern of ordinality
- It can be extended to more dimensions, which makes sense for categorical variables whose categories can be thought of as arrayed across more than one dimension
- See Long and Freese, Ch 6.8

sociology

Ordinal logit

Proportional odds

The proportional odds model

- The most commonly used ordinal logistic model has another logic
- It assumes the ordinal variable is based on an unobserved latent variable
- Unobserved cutpoints divide the latent variable into the groups indexed by the observed ordinal variable
- The model estimates the effects on the log of the odds of being higher rather than lower across the cutpoints

sociology

The model

- For $j = 1$ to $J-1$,

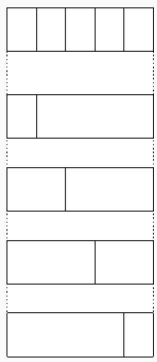
$$\log \frac{P(Y > j)}{P(Y <= j)} = \alpha_j + \beta x$$

- Only one β per variable, whose interpretation is the effect on the odds of being higher rather than lower
- One α per contrast, taking account of the fact that there are different proportions in each one

sociology

$J-1$ contrasts again, but different

But rather than compare categories against a baseline it splits into high and low, with all the data involved each time



sociology

An example

- Using data from the BHPS, we predict the probability of each of 5 ordered responses to the assertion "homosexual relationships are wrong"
- Answers from 1: strongly agree, to 5: strongly disagree
- Sex and age as predictors – descriptively women and younger people are more likely to disagree (i.e., have high values)

sociology

Ordered logistic: Stata output

```
Ordered logistic regression                     Number of obs     =      12725
                                                LR chi2(2)        =    2244.14
                                                Prob > chi2       =     0.0000
Log likelihood = -17802.088                     Pseudo R2         =     0.0593

------------------------------------------------------------------------------
     ropfamr |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      2.rsex |   .8339045    .033062    25.22   0.000     .7691041    .8987048
        rage |  -.0371618   .0009172   -40.51   0.000    -.0389595    -.035364
-------------+----------------------------------------------------------------
       /cut1 |  -3.833869   .0597563                     -3.950989   -3.716749
       /cut2 |  -2.913506   .0547271                      -3.02077   -2.806243
       /cut3 |  -1.132863   .0488522                     -1.228612   -1.037115
       /cut4 |   .3371151   .0482232                      .2425994    .4316307
------------------------------------------------------------------------------
```

sociology

Interpretation

- The betas are straightforward:
  - The effect for women is .8339. The OR is $e^{.8339}$ or 2.302
  - Women's odds of being on the "approve" rather than the "disapprove" side of each contrast are 2.302 times as big as men's
  - Each year of age reduced the log-odds by .03716 (OR 0.964).
- The cutpoints are odd: Stata sets up the model in terms of cutpoints in the latent variable, so they are actually $-\alpha_j$

Linear predictor

- Thus the $\alpha + \beta X$ or linear predictor for the contrast between strongly agree (1) and the rest is (2-5 versus 1)

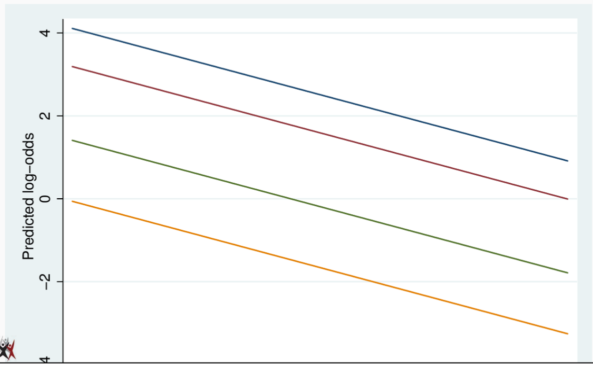
$$3.834 + 0.8339 \times \text{female} - 0.03716 \times \text{age}$$

- Between strongly disagree (5) and the rest (1-4 versus 5)

$$-0.3371 + 0.8339 \times \text{female} - 0.03716 \times \text{age}$$

and so on.

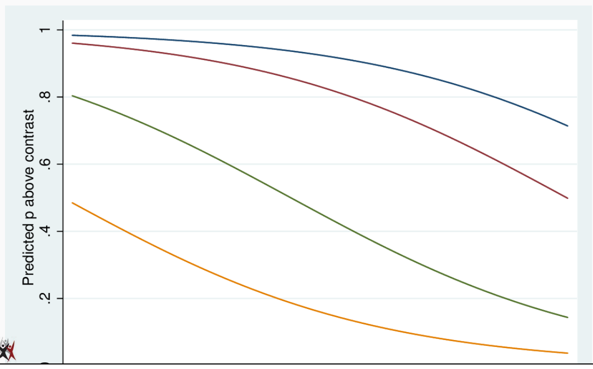
Predicted log odds



Predicted log odds per contrast

- The predicted log-odds lines are straight and parallel
- The highest relates to the 1-4 vs 5 contrast
- Parallel lines means the effect of a variable is the same across all contrasts
- Exponentiating, this means that the multiplicative effect of a variable is the same on all contrasts: hence "proportional odds"
- This is a key assumption

Predicted probabilities relative to contrasts



Predicted probabilities relative to contrasts

- We predict the probabilities of being above a particular contrast in the standard way
- Since age has a negative effect, downward sloping sigmoid curves
- Sigmoid curves are also parallel (same shape, shifted left-right)
- We get probabilities for each of the five states by subtraction

Inference

- The key elements of inference are standard: Wald tests and LR tests
- Since there is only one parameter per variable it is more straightforward than MNL
- However, the key assumption of proportional odds (that there *is* only one parameter per variable) is often wrong.
- The effect of a variable on one contrast may differ from another
- Long and Freese's `SPost` Stata add-on contains a test for this

Testing proportional odds

- It is possible to fit each contrast as a binary logit
- The `brant` command does this, and tests that the parameter estimates are the same across the contrast
- It needs to use Stata's old-fashioned `xi:` prefix to handle categorical variables:

```
xi: ologit ropfamr i.rsex rage
brant, detail
```

Brant test output

```
. brant, detail

Estimated coefficients from j-1 binary regressions

                  y>1         y>2         y>3         y>4
_Irsex_2    1.0198492   .91316651   .76176797    .8150246
    rage   -.02716537  -.03064454  -.03652048  -.04571137
   _cons    3.2067856   2.5225826   1.1214759  -.00985108

Brant Test of Parallel Regression Assumption

    Variable |      chi2   p>chi2    df
-------------+--------------------------
         All |    101.13    0.000     6
-------------+--------------------------
    _Irsex_2 |     15.88    0.001     3
        rage |     81.07    0.000     3
----------------------------------------

A significant test statistic provides evidence that the parallel
regression assumption has been violated.
```


What to do?

- In this case the assumption is violated for both variables, but looking at the individual estimates, the differences are not big
- It's a big data set (14k cases) so it's easy to find departures from assumptions
- However, the departures can be meaningful. In this case it is worth fitting the "Generalised Ordinal Logit" model

Generalised Ordinal Logit

- This extends the proportional odds model in this fashion

$$\log \frac{P(Y > j)}{P(Y <= j)} = \alpha_j + \beta_j x$$

- That is, each variable has a per-contrast parameter
- At the most imparsimonious this is like a reparameterisation of the MNL in ordinal terms
- However, can constrain βs to be constant for some variables
- Get something intermediate, with violations of PO accommodated, but the parsimony of a single parameter where that is acceptable
- Download Richard William's `gologit2` to fit this model:

```
ssc install gologit2
```


Ordinal logit

Sequential logit

Sequential logit

- Different ways of looking at ordinality suit different ordinal regression formations
  - categories arrayed in one (or more) dimension(s): `slogit`
  - categories derived by dividing an unobserved continuum: `ologit` etc
  - categories that represent successive stages: the continuation-ratio model
- Where you get to higher stages by passing through lower ones, in which you could also stay
  - Educational qualification: you can only progress to the next stage if you have completed all the previous ones
  - Promotion: you can only get to a higher grade by passing through the lower grades

"Continuation ratio" model

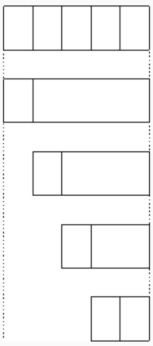
- Here the question is, given you reached level j, what is your chance of going further:

$$\log \frac{P(Y > j)}{P(Y = j)} = \alpha + \beta X_j$$

- For each level, the sample is anyone in level j or higher, and the outcome is being in level $j + 1$ or higher
- That is, for each contrast except the lowest, you drop the cases that didn't make it that far

$J - 1$ contrasts again, again different

But rather than splitting high and low, with all the data involved each time, it drops cases below the baseline



Fitting CR

- This model implies one equation for each contrast
- Can be fitted by hand by defining outcome variable and subsample for each contrast (`ed` has 4 values):

```
gen con1 = ed>1
gen con2 = ed>2
replace con2 = . if ed<=1
gen con3 = ed>3
replace con3 = . if ed<=2
logit con1 odoby i.osex
logit con2 odoby i.osex
logit con3 odoby i.osex
```

seqlogit

- Maarten Buis's `seqlogit` does it more or less automatically:

```
seqlogit ed odoby i.osex, tree(1 : 2 3 4 , 2 : 3 4 , 3 : 4 )
```

- you need to specify the contrasts
- You can impose constraints to make parameters equal across contrasts