ICPSR - 2017

Categorical Data Analysis

Lectures

Scott Long

CDAicpsr2017 lec 2017-05-04 V2.docx

Table of contents

Part 1: Introduction to CDA

Read

Long & Freese: Chapters 1 and 2

What is this class about?

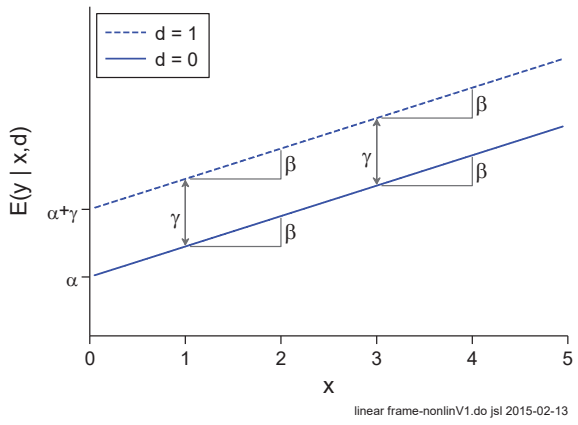
1. The most fundamental *regression models* for categorical outcomes
   - Models for cross-sectional data generalize to panel, hierarchical structures, and more
2. *Telling a story* with data in the presence of *nonlinearity*
   - Interpretations that go beyond signs and stars
3. How you interpret models depends on the *software* used
   - If post-estimation analysis is hard, you are unlikely to do it

Models for categorical outcomes are nonlinear

1. *Nonlinear* models are *implicitly interactive*
   - Nonlinearity is in the form of the model, not product terms (e.g., x1*x2)
2. *The effect of a regressors* depends on:
   - The value of the regressor
   - The values of all other regressors
3. With the LRM, the work is largely done when you estimate the model
   - Unless you add nonlinearities on the RHS
4. With CDA, the work begins when you estimate the model
   - Why does nonlinearity make things so hard (and realistic)?

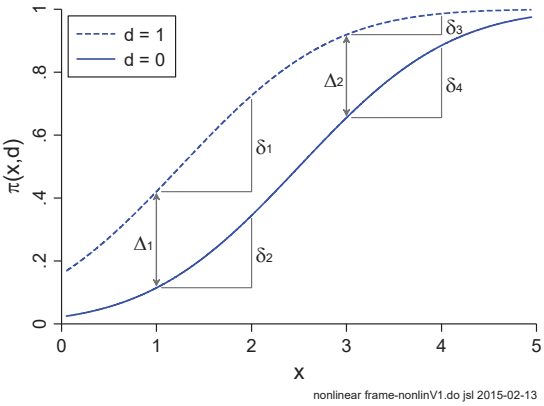
Linear model

$$y = \alpha + \beta x + \gamma d$$



Inherently nonlinear model

$$y = \frac{\exp(\alpha + \beta x + \gamma d)}{1 + \exp(\alpha + \beta x + \gamma d)}$$



RHS (right-hand-side) variables are *linear combinations*

1. Notation
   a. $\mathbf{x}_i\boldsymbol{\beta} = \alpha + \beta x_i$
   b. $\mathbf{x}_i\boldsymbol{\beta} = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Income}_i$
   c. $\mathbf{x}_i\boldsymbol{\beta} = \beta_0 + \beta_1 \text{x}_{i1} + \beta_2 \text{x}_{i2} + \cdots + \beta_K \text{x}_{iK}$
2. Linear combinations can include
   a. *Indicator variables* (binary indicators of characteristics)
   b. *Interaction* variables (products of variables)
   c. *Transformed regressors* such as $x_1 = \sqrt{w_1}$ or $x_2 = w_1^2$
3. With CDA, these tricks lead to unexpected subtleties

Topics considered

Models

1. Continuous outcomes: linear regression
2. Binary outcomes: binary logit and probit
3. Nominal outcomes: multinomial logit
4. Ordinal outcomes: ordinal logit and probit
5. Count outcomes: Poisson regression, negative binomial, ZIP/ZINB

Methods

1. ML estimation and estimation with complex samples
2. Wald and LR tests
3. Measures of internal and external fit

Interpretation

1. Parameters
2. Predictions

What are *your* applications?

1. Think of examples you are interested in
   - Translate my example into your example
2. Read published papers using each methods
   - Collect *exemplars*

Overview of the models we consider

Linear regression model (LRM)

Examples of outcomes

1. *Income*: income; log of income.
2. *Prestige of graduate program*: Scale from 100 to 500.
3. *Number of friends*: How many close friends do you have?
4. *Health index*: sum of ordinal indicators of health.

Binary regression models (BRM)

1. Logit or logistic regression
2. Probit
3. * Linear probability model

Examples

1. *Job status*: Did a person quit her job?
2. *Voting*: Did someone vote? Democrat or Republican?
3. *Schooling*: Does a high school student decide to go on to college?

Nominal regression models (NRM)

1. Multinomial logit
2. * Conditional logit

Examples

1. *Occupation*: manual; craft; white collar; blue collar; pink collar; professional
2. *Marital status*: single; married; divorced; widowed
3. *Preferred job location*: West; Midwest; South; Northwest; Northeast

Ordinal regression models (ORM)

1. Ordinal logit and probit
2. * Generalized ordered logit
3. * Stereotype model; continuation ratio model; adjacent category model

Examples

1. *Political party*: 1=Strongly Democrat to 5=Strongly Republican
2. Likert scale on attitudes toward working mothers: 1=SA; 2=A; 3=D; 4=SD.
3. *Rank attainment*: 1=Assistant Prof.; 2=Associate Prof.; 3=Full Prof.
4. *Social class*: 1=lower; 2=middle; 3=upper.

Count regression model (CRM)

1. Poisson regression
2. Negative binomial
3. Zero-inflated models
4. * Truncated and hurdle regression

Examples

1. *Strikes*: How many strikes occurred?
2. *Articles*: How many articles did a scientist publish?
3. *Demonstrations*: How many political demonstrations occurred?
4. *Number of extramarital affairs or number of partners*

Measurement of the outcome

1. Models are characterized by the level of measurement of the outcome
2. If you assume the wrong level of measurement

   Bias: on average the wrong answer

   Inefficiency: not using the data as well as you could

   Inappropriate answers
3. What is the true level of measurement?

   "Assumptions that a variable is somehow 'intrinsically' interval (ordinal, nominal) are analytically misleading." – Lew Carter, *Social Forces* 1971
4. What makes a variable ordinal? Nominal?
   - What level of measurement does the concept *education* have?
   - Is scientific productivity continuous?

Examples in lectures

1. The do-files for the results in lectures are available, but are often more complex than you need
2. Shorter versions are provided that you *must* run these before you start your assignments
   - **`cdalec*-`**<topic>**`-`**<dataset>**`.do`**
   - Ideally, use these as *templates* for your analysis
3. The output in the lectures is sometimes edited

Math and CDA

Read and run

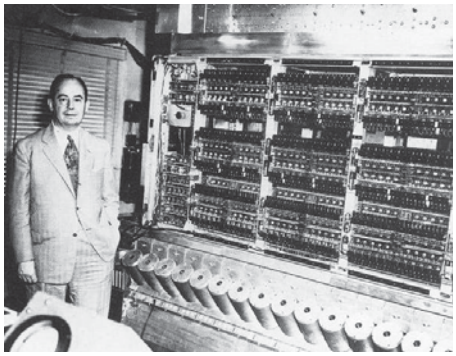
cdaiu math review 2007.pdf

Objectives

1. Getting comfortable with notation is critical.
2. Overcoming math anxiety: How do you "learn" math?
   a. Lewis Carol
   b. Sir Isaac Newton
   c. John Von Neumann had some advice...
3. How would you graph your understanding over time?

John von Neumann (1903-1957)

The world's smartest man. --Time Magazine



In mathematics you don't understand things.
You just get used to them. --John Von Neumann

From simple to complex

1. The *same* rules apply
   - A *simple* equation

     $$x = y$$

   - A *complex* equation

     $$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + u$$

2. Don't confuse *messy* with *hard*!

Basic Rules

Distributive law

Simple

$$a \times (b + c) = (a \times b) + (a \times c)$$

$$4 \times (2 + 3) = (4 \times 2) + (4 \times 3)$$

Complex

$$\begin{aligned}(\varphi_1 - \varphi_2)(\beta_0 + \beta_1 x_1 + \beta_2 x_2) &= (\varphi_1 - \varphi_2)\Delta \\ &= \varphi_1 \Delta - \varphi_2 \Delta \\ &= \varphi_1 (\beta_0 + \beta_1 x_1 + \beta_2 x_2) - \varphi_2 (\beta_0 + \beta_1 x_1 + \beta_2 x_2) \\ &= [\varphi_1 \beta_0 + \varphi_1 \beta_1 x_1 + \varphi_1 \beta_2 x_2] - [\varphi_2 \beta_0 + \varphi_2 \beta_1 x_1 + \varphi_2 \beta_2 x_2]\end{aligned}$$

Multiplying by 1

$$\frac{a}{b} = 1 \times \frac{a}{b} = \frac{k}{k} \times \frac{a}{b} = \frac{ka}{kb}$$

$$\frac{2}{3} = 1 \times \frac{2}{3} = \frac{4}{4} \times \frac{2}{3} = \frac{4 \times 2}{4 \times 3} = \frac{8}{12} = \frac{2}{3}$$

Adding 0

$$y = 0 + y$$

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \\ &= 0 + \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \\ &= (\delta - \delta) + \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \\ &= (\beta_0 + \delta) + \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + (\varepsilon_i - \delta) \\ &= \beta_0^* + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i^* \end{aligned}$$

Part 2: Linear regression

Read and run

Long & Freese	Chapters 3 and 4
cdalec*.do	lrm-anscombe.do, lrm-regjob.do, lrm-science.do, lrm-slid-ontario-.do

Objectives

1. Establish *notation* and *terminology*
2. Reinforce the ideas of *linearity* and *nonlinearity*
3. Explain the concept of *identification*
4. Introduce *maximum likelihood estimation*
5. Introduce `m*` commands for *post-estimation*

Notation

Outcome = linear combination + error

1. $y_i = \alpha + \beta x_i + \varepsilon_i$
2. $Occupation = \beta_0 + \beta_1 Education + \beta_2 ParentEd + \beta_3 ParentOcc + \varepsilon$
3. $y_i = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i$

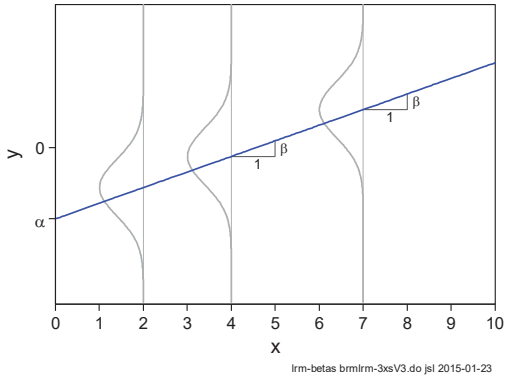
$$= \begin{bmatrix} 1 & x_{i1} & \dots & x_{iK} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} + \varepsilon_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_K x_{iK} + \varepsilon_i$$

ε is unexplained variation

1. Randomness
2. Unobserved heterogeneity.

Assumptions

1. Linearity
2. Not perfect collinearity
3. E(ε|x)=0
4. Homoscedasticity
5. Uncorrelated errors
6. Normality



Conditional mean error and identification

An important concept for understanding the BLM

1. We *assume* the average error is 0:

$$E(\varepsilon_i \mid \mathbf{x}_i) = 0$$

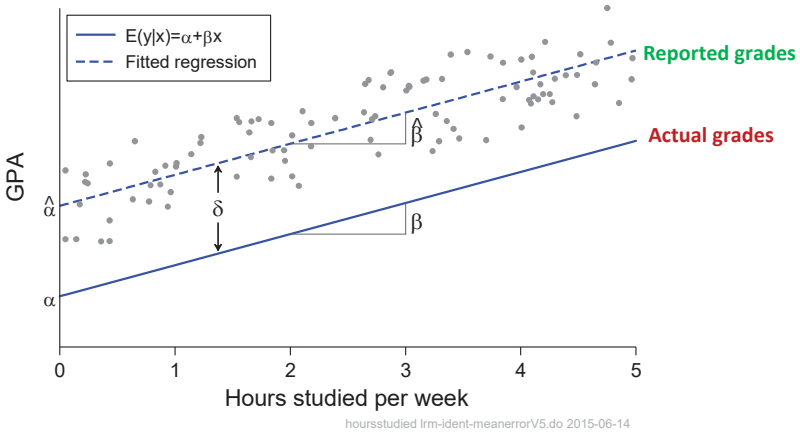
2. This implies

$$E(y_i \mid \mathbf{x}_i) = E(\mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \mid \mathbf{x}_i)$$
$$= \mathbf{x}_i\boldsymbol{\beta} + E(\varepsilon_i \mid \mathbf{x}_i) = \mathbf{x}_i\boldsymbol{\beta} = \hat{y}_i$$

3. How do you *know* the error is on average 0?

Identification: a pre-requisite to estimation

You can estimate β but not α.



General principles of identification

1. An unidentified parameter *cannot be estimated* with more data.
   - Doubling the data does not help.
2. Parameters are identified by:
   a. Adding assumptions.
   b. New kinds of data.
3. *Identification is not all or nothing*: Some parameters can be identified while others are not.
4. *Combinations of unidentified parameters can be identified*, while the individual parameters are not.
   α+δ is identified, but α or δ are not individually identified.

Interpretation with marginal effects

1. *Marginal effects* measure
   - How much the outcome changes
   - for a change in *one* regressor
   - holding other regressors *constant*.
2. Two types of marginal effects
   a. *Discrete change* in $\hat{y}$ as a regressor changes by a fixed amount.
   b. *Marginal change* in $\hat{y}$ for an infinitely small change in a regressors.

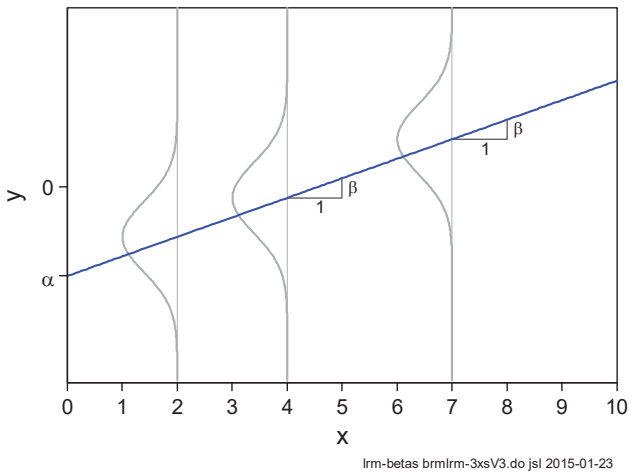
Discrete change in E(y|x)

1. Starting at $E(y \mid \mathbf{x}, x_3)$: expected value before change in x_3
2. Ending at $E(y \mid \mathbf{x}, x_3 + 1)$: expected value after change in x_3.
3. The *discrete change* for a change of 1 in x_3:

$$\begin{aligned}\frac{\Delta E(y \mid \mathbf{x}, x_3)}{\Delta x_3} &= \text{Ending} - \text{Starting} = \beta_3 \\ &= E(y \mid \mathbf{x}, x_3 + 1) - E(y \mid \mathbf{x}, x_3) \\ &= \left[\beta_0 + \beta_1 x + \beta_2 x_2 + \underline{\beta_3 (x_3 + 1)}\right] - \left[\beta_0 + \beta_1 x + \beta_2 x_2 + \underline{\beta_3 x_3}\right] \\ &= \beta_3\end{aligned}$$

4. The change does *not* depend on
   a. The specific value of x_3
   b. The specific values of the x_k's that are held constant
5. Graphically,…

Discrete change



Marginal change in E(y|x)

1. The *instantaneous rate of change* in the expected value of *y* as x_k changes, holding other x's constant

$$\frac{\partial E(y \mid \mathbf{x})}{\partial x_k} = \frac{\partial \mathbf{x}\boldsymbol{\beta}}{\partial x_k} = \beta_k$$

2. The marginal change is the slope at a *specific* location
3. In the LRM, the marginal does *not* depend on
   a. The value of x_k
   b. The values at which other *x*'s are held constant

Marginal and discrete change in LRM

In linear models (without interactions)

$$\frac{\partial E(y \mid \mathbf{x})}{\partial x_k} = \frac{\Delta E(y \mid \mathbf{x})}{\Delta x_k} = \beta_k$$

.

Simple interpretation due to linearity

Continuous variables

1. For a unit increase in x_k the expected change in y is β_k, holding other variables constant.
2. For each additional year of education, income is expected to increase by $1,400, holding other variables constant.

Dummy variables coded as 0 and 1:

1. Having characteristic x_k (as opposed to not having the characteristic) results in an expected change of β_k in y, holding other variables constant.
2. Being a female decreases the expected salary by $1,400, holding other variables constant.

Can you hold other variables constant?

1. These interpretations assume one variable changes without changing other variables
2. With linked variables this is mathematically impossible
   - o x and x^2 must change together
3. More generally
   - o Does it make *substantive* sense to change one regressor holding others constant?
   - o Can you increase education holding everything else constant?

What does it mean when we say a variable is changing?

1. What does the *counterfactual* mean: *increasing a person's education while holding income and occupation constant?*
2. Does it make sense to imagine changing gender?

Example: academic job prestige (-lrm-regjob.do)

#1 Descriptive statistics

```
. use regjob3, clear
(Long's data on academic jobs of biochemists \ 2009-03-13)

. codebook job100 fem phd100 ment fel art cit, compact

Variable Obs Unique      Mean  Min       Max  Label
------------------------------------------------------------------------------
job100   408     80  223.3431  100       480  Prestige of 1st job on 100 to...
fem      408      2  .3897059    0         1  Gender: 1=female 0=male
phd100   408     89  320.0564  100       480  PhD prestige on 100 to 500 scale
ment     408    123  45.47058    0  531.9999  Citations received by mentor
fel      408      2  .6176471    0         1  Fellow: 1=yes 0=no
art      408     14  2.276961    0        18  # of articles published
cit      408     87  21.71569    0       203  # of citations received
------------------------------------------------------------------------------
```

In serious research, the variables would be carefully examined before modeling begins.

#7 Estimating the LRM

```
  <cmd>   <lhs>  <rhs> [, <options>]

. regress job100 i.fem c.phd100 c.ment i.fel c.art c.cit

      Source |       SS           df       MS      Number of obs   =       408
-------------+----------------------------------   F(6, 401)       =     17.78
       Model |  810584.791         6  135097.465   Prob > F        =    0.0000
    Residual |  3047379.21       401  7599.44941   R-squared       =    0.2101
-------------+----------------------------------   Adj R-squared   =    0.1983
       Total |     3857964       407  9479.02704   Root MSE        =    87.175

------------------------------------------------------------------------------
      job100 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         fem |
    1Female  |  -13.91939   9.023442    -1.54   0.124    -31.65856    3.819769
      phd100 |   .2726826   .0493183     5.53   0.000     .1757278    .3696375
        ment |   .1186708   .0701164     1.69   0.091    -.0191709    .2565125
         fel |
    1Fellow  |   23.41384   9.482065     2.47   0.014     4.773075    42.05461
         art |   2.280112   2.888427     0.79   0.430    -3.398239    7.958464
         cit |   .4478843   .1968665     2.28   0.023      .060865    .8349036
       _cons |   106.7184   16.61357     6.42   0.000     74.05785     139.379
------------------------------------------------------------------------------
```

Factor-variables notation

1. **`i.`***varname* creates indicator variables of all but the base category of *varname*
   - o **`i.fem`** has some advantages over **`fem`** when using **`margins`**
2. For **`i.fem`**:
   a. **`1.fem`** equals 1 if **`fem`** is 1.
   b. **`0.fem`** equals 1 if **`fem`** is 0.
3. For **`i.agecat`**, **`2.agecat`** and **`3.agecat`** indicate if **`agecat`** is 2 or 3

Continuous variables

1. By default a variable is *not* an indicator variable
2. To make this explicit, use **`c.`***varname*

   `c.art` could have been specified **`art`**

Interpretations follow...

Interpreting unstandardized coefficients

1. Being a female scientist decreases the *expected* prestige of the first job by 14 points on a 400 point scale, holding other variables constant.

```
     job100 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------+----------------------------------------------------------------
        fem |  -13.91939   9.023442    -1.54   0.124    -31.65856    3.819769
```

2. For each additional citation, the prestige of the first job is *expected* to increase by .45 units, holding other variables constant.

```
     job100 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------+----------------------------------------------------------------
        cit |   .4478843   .1968665     2.28   0.023      .060865    .8349036
```


Standardized coefficients

1. Standardized coefficients are often used to interpret the LRM.
2. In binary & ordinal models, standardization is required due to identification.

Tool: Standardizing to 1

1. Standard deviation of x_k : $sd(x_k) = \sigma$
2. Standard deviation of αx_k : $sd(\alpha x_k) = \alpha \sigma$
3. Then: $sd(1/\sigma\, x_k) = (1/\sigma)\, sd(x_k) = \sigma/\sigma = 1$

#8 Standardizing variables

```
. egen job100std = std(job100) // job100 standardized
. egen artstd = std(art) // articles standardized
. sum job100 job100std art artstd

    Variable |     Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      job100 |     408    223.3431    97.36029        100        480
   job100std |     408   -8.73e-09           1  -1.266873   2.636156
         art |     408    2.276961    2.256143          0         18
      artstd |     408   -1.68e-08           1  -1.009227   6.968991

* unstandardized variables
regress job100      fem phd100 ment fel art    cit

* none standardized
  regress job100    fem phd100 ment fel art    cit

* y & x standardized
  regress job100std fem phd100 ment fel artstd cit

* x standardized
  regress job100    fem phd100 ment fel artstd cit

* y standardized
  regress job100std fem phd100 ment fel art    cit
```

This is what listcoef does

#7 Standardized coefficients with listcoef

```
. listcoef, cons help

regress (N=408): Unstandardized and Standardized Estimates

 Observed SD: 97.360295
 SD of Error: 87.174821

------------------------------------------------------------------------------
      job100 |      b         t     P>|t|    bStdX    bStdY   bStdXY     SDofX
-------------+----------------------------------------------------------------
         fem | -13.91939   -1.543   0.124   -6.7966  -0.1430  -0.0698    0.4883
      phd100 |   0.27268    5.529   0.000   26.0071   0.0028   0.2671   95.3751
        ment |   0.11867    1.692   0.091    7.7765   0.0012   0.0799   65.5299
         fel |  23.41384    2.469   0.014   11.3922   0.2405   0.1170    0.4866
         art |   2.28011    0.789   0.430    5.1443   0.0234   0.0528    2.2561
         cit |   0.44788    2.275   0.023   14.8070   0.0046   0.1521   33.0599
       _cons | 106.71842    6.424   0.000
------------------------------------------------------------------------------
       b = raw coefficient
       t = t-score for test of b=0
   P>|t| = p-value for t-test
   bStdX = x-standardized coefficient
   bStdY = y-standardized coefficient
  bStdXY = fully standardized coefficient
```

SDofX = standard deviation of X

y-standardized coefficients

```
regress job100std fem phd100 ment fel art    cit
```

Standardize *y* to a unit variance:

$$\frac{y}{\sigma_y} = \frac{\beta_0}{\sigma_y} + \frac{\beta_1}{\sigma_y}x_1 + \frac{\beta_2}{\sigma_y}x_2 + \frac{\beta_3}{\sigma_y}x_3 + \frac{\varepsilon}{\sigma_y}$$

$$= \beta_0^{S_y} + \beta_1^{S_y}x_1 + \beta_2^{S_y}x_2 + \beta_3^{S_y}x_3 + \varepsilon^{S_y}$$

For a continuous variable

For a unit increase in x_k, *y* is expected to change by $\beta_k^{S_y}$ *standard deviations*, holding other variables constant.

For a dummy variable

Having characteristic x_k (as opposed to not having it) results in an expected change in *y* of $\beta_k^{S_y}$ *standard deviations*, holding other variables constant.

Examples

1. Being a woman decreases the expected prestige of the first job by .14 standard deviations, holding other variables constant.

job100	b	t	P>|t|	bStdX	bStdY	bStdXY	SDofX
fem	-13.91939	-1.543	0.124	-6.7966	-0.1430	-0.0698	0.4883

2. For *every* additional citation, the prestige of the first job is expected to increase by .005 standard deviations, holding other variables constant.

job100	b	t	P>|t|	bStdX	bStdY	bStdXY	SDofX
cit	0.44788	2.275	0.023	14.8070	0.0046	0.1521	33.0599

Fully standardized ("beta") coefficients

```
regress job100std fem phd100 ment fel artstd cit
```

Combine *y* and *x*'s standardization:

$$\frac{y}{\sigma_y} = \frac{\beta_0}{\sigma_y} + \left(\frac{\sigma_1\beta_1}{\sigma_y}\right)\frac{x_1}{\sigma_1} + \left(\frac{\sigma_2\beta_2}{\sigma_y}\right)\frac{x_2}{\sigma_2} + \left(\frac{\sigma_3\beta_3}{\sigma_y}\right)\frac{x_3}{\sigma_3} + \frac{\varepsilon}{\sigma_y}$$

$$= \beta_0^S + \beta_1^S x_1^S + \beta_2^S x_2^S + \beta_3^S x_3^S + \varepsilon^{S_y}$$

For a continuous variable

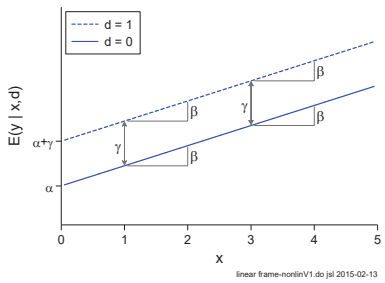
For a *standard deviation increase in x_k*, *y* is expected to change by β_k^S *standard deviations*, holding other variables constant.

- For *every* standard deviation increase in citations, the prestige of the first job is expected to increase by .15 standard deviations, holding other variables constant.

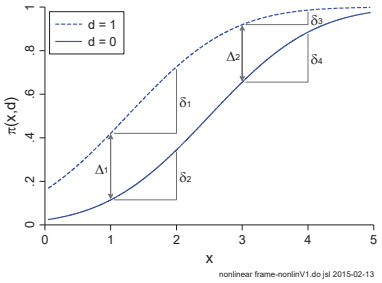
job100	b	t	P>|t|	bStdX	bStdY	bStdXY	SDofX
cit	0.44788	2.275	0.023	14.8070	0.0046	0.1521	33.0599

Linear and nonlinear models

A: Linear model



B: Nonlinear model



Nonlinear compared to linear models

Marginal effect of x_k in linear models

1. The size of the effect *does not* depend on the value of x_k
2. The size of the effect *does not* depend on the values of other *x*'s
3. Marginal change and discrete change are *equal*

$$\frac{\partial \mathrm{E}(\cdot)}{\partial x_k} = \frac{\Delta \mathrm{E}(\cdot)}{\Delta x_k}$$

Marginal effect of x_k in nonlinear models

1. The size of the effect *does* depend on the value of x_k
2. The size of the effect *does* depend on the values of the other *x*'s
3. Marginal and discrete change are usually *unequal*

$$\frac{\partial \mathrm{E}(\cdot)}{\partial x_k} \neq \frac{\Delta \mathrm{E}(\cdot)}{\Delta x_k}$$

Nonlinear linear regression models

1. By transforming regressors, effects of variables can be "nonlinear"
2. The *x*'s enter in the linear form $\mathbf{x\beta} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots$
3. But the *x*'s can be *transformations* of variables, such as.

Quadratic:
$$\begin{aligned} y &= \beta_0 + \beta_1 w_1 + \beta_2 w_1^2 + \varepsilon \\ &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \end{aligned}$$

Square root:
$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2 \sqrt{w_2} + \varepsilon \\ &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \end{aligned}$$

Loglinear:
$$y = \ln z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

Loglinear models: essential for later models

1. Review: $\exp(a + b) = \exp(a)\exp(b)$

   $\log[\exp(a + b)] = a + b$
2. An exponential model is multiplicative on the y metric

$$\begin{aligned} y &= \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon) \\ &= \exp(\beta_0)\exp(\beta_1 x_1)\exp(\beta_2 x_2)\exp(\varepsilon) \end{aligned}$$

3. Taking the log makes the model *loglinear* on the log(y) metric

$$\begin{aligned} \ln(y) &= \ln\left[\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon)\right] \\ &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \end{aligned}$$

4. β_1 is interpreted as:
   - For a unit increase in x_1, *the log of y* is expected to increase by β_1 units, holding other variables constant.
5. A change in the *log of y* is often substantively unclear

Interpreting loglinear models with factor change

What is a "factor change"?

y = 10.

Factor change of 2: y is doubled or twice as large

2*10 = 20

Factor change of .5: y is made half as large

.5*10 = 5

Factor and percentage change

If y is 2 times larger, y increases 100%

If y is 1.5 times larger, y increases 50%

If y is .5 times smaller, y decreases 50%

Factor change in y

1. **Start value**: Let y(**x**,x_1) be the value of *y* focusing on x_1:

$$y(\mathbf{x}, x_1) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon)$$
$$= \exp(\beta_0)\exp(\beta_1 x_1)\exp(\beta_2 x_2)\exp(\varepsilon)$$

2. **End value**: y(**x**,x_1+1) is the value of y after increasing x_1 by 1:

$$y(\mathbf{x}, x_1 + 1) = \exp(\beta_0)\exp\left[\beta_1 (x_1 + 1)\right]\exp(\beta_2 x_2)\exp(\varepsilon)$$
$$= \exp(\beta_0)\exp(\beta_1 x_1)\exp(\beta_1)\exp(\beta_2 x_2)\exp(\varepsilon)$$

3. The ratio y(**x**,x_1+1)/y(**x**,x_1) is the *factor change* in *y* for a unit increase in x_1:

$$\frac{y(\mathbf{x}, x_1 + 1)}{y(\mathbf{x}, x_1)} = \frac{y_{end}}{y_{start}}$$
$$= \frac{\exp(\beta_0)\exp(\beta_1 x_1)\exp(\beta_1)\exp(\beta_2 x_2)\exp(\varepsilon)}{\exp(\beta_0)\exp(\beta_1 x_1)\exp(\beta_2 x_2)\exp(\varepsilon)} = \exp(\beta_1)$$

For a unit increase in x_1, y is expected be $\exp(\beta_1)$ times larger (or smaller), holding other variables *constant*.

Percentage and factor change

1. We can translate the factor change $\exp(\beta_1)$ to percentage change:

$$100\left[\frac{y(\mathbf{x}, x_1 + 1) - y(\mathbf{x}, x_1)}{y(\mathbf{x}, x_1)}\right] = 100\left[\exp(\beta_1) - 1\right]$$

   - If you make \$10/hour and get a raise to \$11/hour, a 10% raise:

$$100\frac{\$11 - \$10}{\$10} = 100\frac{\$1}{\$10} = 100\left[\frac{\$11}{\$10} - \frac{\$10}{\$10}\right] = 10$$

2. Either factor or percentage change can be used
   - For a unit increase in x_1, *y* is expected to change by $100[\exp(\beta_1)-1]$ percent, holding other variables constant.
   - For a unit increase in x_1, y is expected be $\exp(\beta_1)$ times larger (or smaller), holding other variables constant.

Example: Wages in Canada (-lrm-slid-ontario.do)

Fox (2008) *Applied Regression Analysis and Generalized Linear Models* 2nd, p267.
Survey of Labour & Income Dynamics, Ontario, Canada, 1994.

Descriptive statistics:

```
. use slid-ontario01, clear
(Canada's 1994 Survey of Labor and Income Dynamics \ 2011-04-04)

. codebook, compact

Variable     Mean   StdDev  Minimum  Maximum  Label
------------------------------------------------------------------------------
logwages     2.62     0.50     0.83     3.91  Log(wages) in base e
male         0.50     0.50     0.00     1.00  Is male?
age         36.96    12.00    16.00    65.00  Age in years.
edyears     13.21     3.04     0.00    20.00  Years of education completed.
------------------------------------------------------------------------------
N=3,997
```

M1: baseline loglinear regression

$$\ln(wages) = \beta_0 + \beta_1 male + \beta_2 edyears + \beta_3 age + \varepsilon$$

Estimates follow with a focus on age...

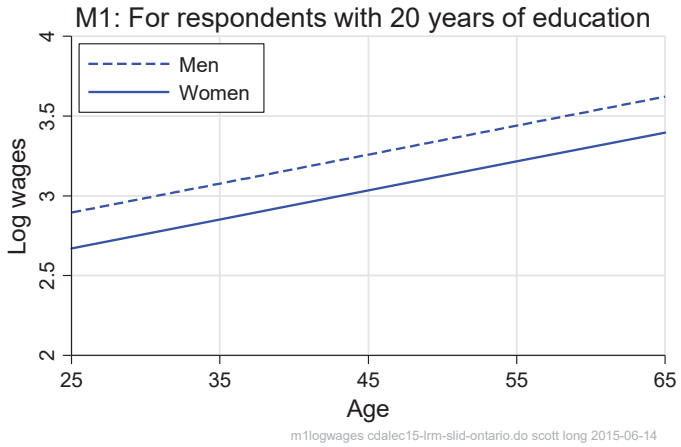
#11 estimating M1

```
. regress logwages male age edyears
      Source |       SS       df       MS              Number of obs =    3997
-------------+------------------------------           F(  3,  3993) =  629.73
       Model |  324.777885     3  108.259295            Prob > F      =  0.0000
    Residual |  686.449784  3993  .171913294           R-squared     =  0.3212
-------------+------------------------------           Adj R-squared =  0.3207
       Total |  1011.22767  3996  .253059977           Root MSE      =  .41462

------------------------------------------------------------------------------
    logwages |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   .2244959   .0131208    17.11   0.000     .1987718    .2502201
         age |   .0181548   .0005491    33.06   0.000     .0170782    .0192315
     edyears |   .0558764   .0021713    25.73   0.000     .0516195    .0601334
       _cons |   1.099018   .0379649    28.95   0.000     1.024585     1.17345
```

Linear in log wages

1. For each additional year of age, the *log of wages* is expected to increase by .018, holding other variables constant.
2. Graphically, *on the next page...*

3. At all ages, getting one year older leads to the same increase in *log wages*.

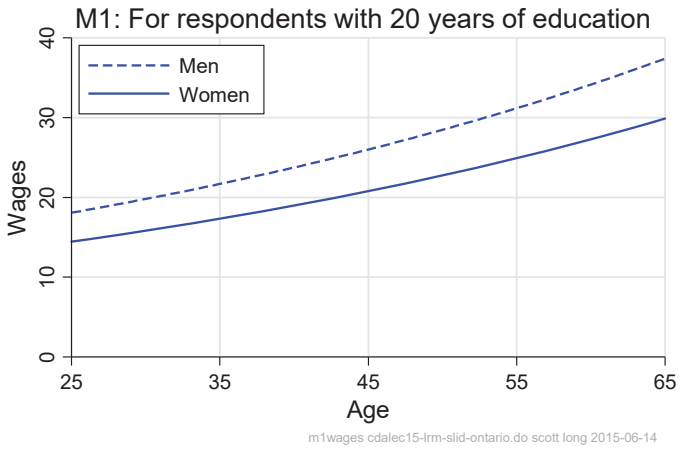


#13 Transforming log(wages) to wages

1. Wages are easier to understand than log of wages
2. To plot *wages* against *age*, take the exponential:

   wages = exp[log(wages)]
3. The plot shows a specific form of nonlinearity.
   - Details on graphing discussed in later chapters

Graph on next page...



Interpretation of coefficients follows...

#11 The rate of increase in wages

```
regress, eform(Factor) // "Factor" is a name I use to label coefficients
```

logwages	Factor	Std. Err.	t	P>|t|	[95% Conf.	Interval]
male	1.251692	.0164232	17.11	0.000	1.219904	1.284308
age	1.018321	.0005592	33.06	0.000	1.017225	1.019418
edyears	1.057467	.0022961	25.73	0.000	1.052975	1.061978
_cons	3.001216	.1139408	28.95	0.000	2.78594	3.233128

1. For each year of age, wages increase by a factor of 1.018 or 1.8%.

   $\exp(b_{age}) = \exp(.018) = 1.018$

   $100[\exp(b_{age})-1] = 100[\exp(.018)-1] = 1.8\%$
2. For each 5-years of age, wages increase by a factor of 1.093 or 9.3%.

   $[\exp(.018)]^5 = (1.018)(1.018)(1.018)(1.018)(1.018) = 1.018^5$

   $= 1.093$
   - Same rate of change from 30 to 35 or from 40 to 45.

OLS estimates of β's

1. Guess the values of the coefficients: $\boldsymbol{\beta}^*$
2. For this guess, the residuals for case *i* are:

$$r_i^* = y_i - \mathbf{x}_i\boldsymbol{\beta}^* = y_i - y_i^*$$

3. Compute the sum of squared residuals *SSR*= $\sum_{i=1}^{N}\left(r_i^*\right)^2$
4. Try other values of $\boldsymbol{\beta}^*$ to see if you can make *SSR* smaller.
5. *The OLS estimate* $\hat{\boldsymbol{\beta}}$ *minimizes the SSR*:

$$SSR = \sum_{i=1}^{N}\left(y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right)^2 = \sum_{i=1}^{N}\left(\hat{\varepsilon}_i\right)^2$$

6. OLS has a simple "closed-form" formula:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

Properties of the OLS estimator

1. When the assumptions hold, OLS estimates are BLUE.
   a. Best: smallest possible sampling variance.
   b. Linear: a linear combination of the data.
   c. Unbiased: on average the correct answer across samples.
2. These properties are asymptotic, but OLS works very well for small N.

Residuals

1. First we estimate the intercept and the slope coefficients.
2. With these estimates and the observed data we compute residuals.

$$\hat{\varepsilon}_i = y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}$$

3. The residuals are used to estimate the variance of the error:

$$Var\left(\hat{\varepsilon}\right) = \frac{1}{N-K-1}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)_i^2 = \frac{1}{N-K-1}\sum_{i=1}^{N}\hat{\varepsilon}_i^2 = s^2$$

R^2: explained variation

R^2 is the percent of the variation in *y* explained by the regressors.

$$R^2 = \frac{Var\left(\hat{y}\right)}{Var\left(\hat{y}\right)+Var\left(\hat{\varepsilon}\right)} = \frac{Explained\ Variation}{Total\ Variation}$$

Knowing the β's does not tell you the R^2

1. Consider models for men and women:

$$y = \beta_0^W + \beta_1^W edyears + \beta_2^W age + \varepsilon$$

$$y = \beta_0^M + \beta_1^M edyears + \beta_2^M age + \varepsilon$$

2. Equal coefficients for men and women *does not* imply $R_W^2 = R_M^2$ if $Var_W\left(\hat{\varepsilon}\right) \neq Var_M\left(\hat{\varepsilon}\right)$.

If the slopes are equal, but the R^2's differ, are the "effects" the same for both groups? Are the social processes the same?

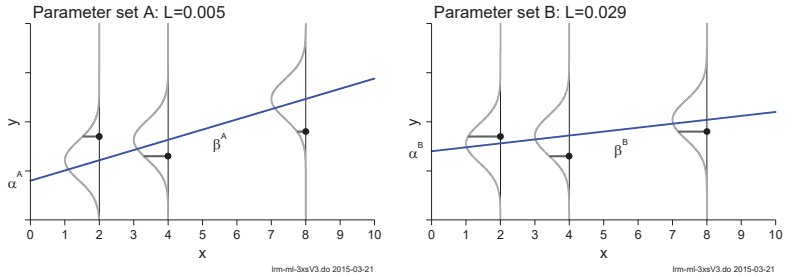
Maximum likelihood estimation (MLE)

1. *ML estimates maximize the likelihood of what you observe*.
2. No other values of parameters increase your chance of observing your data.
3. In LRM OLS estimates are the same as ML estimates.
4. In later models, we cannot use OLS and rely on MLE.
5. Graphically, here is how MLE works in the LRM.

Graph on next page...

Graphical view of ML estimation for LRM

1. Which set of parameters makes the data more likely?



2. The likelihood curve considers all possible values of the parameters.

Properties of ML estimators

1. Under general conditions, the MLE is:
   a. *Consistent*: the mean of the sampling distribution approaches the true value.
   b. *Asymptotically efficient*: Data are used as well as possible.
   c. *Asymptotically normal*: The sampling distribution becomes normal.
2. How big must N be to approximate infinity? If N is small, are the estimates necessarily bad? No.
3. OLS tends to work well with very small samples.
4. Part 3 considers how large N needs to be for other models.

Variance in estimated β's used for testing

1. The covariance matrix the X and Z coefficients:

$$\sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1} = Var\left(\hat{\boldsymbol{\beta}} \text{ for X and Z}\right) = \begin{pmatrix} Var\left(\hat{\beta}_X\right) & Cov\left(\hat{\beta}_X, \hat{\beta}_Z\right) \\ Cov\left(\hat{\beta}_Z, \hat{\beta}_X\right) & Var\left(\hat{\beta}_Z\right) \end{pmatrix}$$

2. Let *X* be the right wall, *Z* the left, and *Y* the height of the room.
3. Off-diagonal elements indicate how the regression plane "rocks".
4. Off-diagonal elements are critical for tests of multiple coefficients.
5. What affects $Var\left(\hat{\beta}_X\right)$?
6. Imagine holding a large sheet of plywood. Why does it wobble?

What affects the variance of the estimated slope?

1. Let: $y = \beta_0 + \beta_x x + \beta_z z + \varepsilon$
2. ρ_{XZ} is the correlation between *X* and *Z*.
3. Then:

$$Var\left(\hat{\beta}_X\right) = \frac{\sigma_\varepsilon^2}{N\sigma_X^2\left(1-\rho_{XZ}^2\right)}$$

Each component affects the variance

1. Increasing **sample size (*N*):** **Decreases** $Var\left(\hat{\beta}_X\right)$
2. Increasing **variance in X (σ_X^2):** **Decreases** $Var\left(\hat{\beta}_X\right)$
3. Increasing **collinearity (ρ_{XZ}^2):** **Increases** $Var\left(\hat{\beta}_X\right)$
4. Increasing **error variance (σ_ε^2):** **Increases** $Var\left(\hat{\beta}_X\right)$

Overview of hypothesis testing

1. Consider the hypothesis

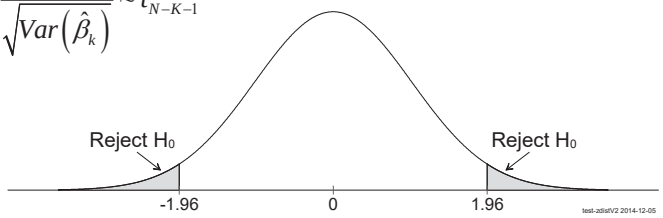
   H_0: $\beta_k=0$

2. Two types of errors are possible when testing H_0: β=0

	Decision	
H_0: β=0	**Accept H_0**	**Reject H_0**
In fact β=0	No error	**Type I: Pr(reject true)=α** Area in the shaded tail. Size of the test.
In fact β≠0	**Type II: accept false** Power of test.	No error

3. If the errors are normal and β_k=0, then

4. $$t_k = \frac{\hat{\beta}_k - 0}{\sqrt{Var\left(\hat{\beta}_k\right)}} \sim t_{N-K-1}$$



5. For a *two-tailed test*, H_0 is rejected at the .05 level when t/z falls in the shaded region of either tail.

#17 Example of t-tests in regression (-lrm-regjob.do)

```
. regress job100 fem phd100 ment fel art cit
```

job100	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
fem	-13.91939	9.023442	-1.54	0.124	-31.65856	3.819769
phd100	.2726826	.0493183	5.53	0.000	.1757278	.3696375
ment	.1186708	.0701164	1.69	0.091	-.0191709	.2565125
fel	23.41384	9.482065	2.47	0.014	4.773075	42.05461
art	2.280112	2.888427	0.79	0.430	-3.398239	7.958464
cit	.4478843	.1968665	2.28	0.023	.060865	.8349036
_cons	106.7184	16.61357	6.42	0.000	74.05785	139.379

1. Doctoral origin has a significant effect on the prestige of the current job (t=5.53, p<0.01 for a *two-tailed* test).
2. Being female does not significantly affect job prestige (p>.05 for a 2-tailed test).

Overview of continuous LHS

1. LRM is the foundation for CDA models
   - But be careful about generalizing from LRM to other models!
2. Variables enter the model as **xβ**, called the *index function*
3. **xβ** allows flexible specifications through interactions and transformations
4. Nonlinearity makes interpretation more complicated even in linear models
5. All of the models in later sections are nonlinear

Part 3: Binary outcomes

Read and run

Long & Freese Chapters 5 and 6

cdalec*.do cdalec-brm-lfp.do; cdalec-brm-science.do

Objectives

1. Why are binary variables coded as 0 and 1?
2. What are the limitations of the LRM for binary outcomes?
3. Derive the binary regression model (BRM) as
   - Latent variable model
   - Probability model
   - Random utility model
   - Generalized linear model
4. Illustrate fundamental methods of interpretation using predictions.
   - Parameters versus predictions

Binary variables, means and expectations

1. Consider probabilities of observing the binary values 0 and 1

y – value	Probability
0	1/4
1	3/4

2. The mean mixes the 0's and 1's weighted by their probabilities

$$\begin{aligned} Mean &= \left[0 \times \Pr(y=0)\right] + \left[1 \times \Pr(y=1)\right] \\ &= \left[0 \times 1/4\right] + \left[1 \times 3/4\right] \\ &= 3/4 \end{aligned}$$

3. More formally

$$E(y) = \left[0 \times \Pr(y=0)\right] + \left[1 \times \Pr(y=1)\right] = \Pr(y=1)$$

4. Conditional on values of other variables

$$E(y \mid \mathbf{x}) = \Pr(y=1 \mid \mathbf{x})$$

The linear probability model (LPM)

1. The structural model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$$

2. Taking expectations

$$\begin{aligned} E(y_i \mid \mathbf{x}) &= \Pr(y_i = 1 \mid \mathbf{x}) \\ &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} \end{aligned}$$

3. Example

   a. $$LFP = \begin{cases} 1 \text{ if in labor force} \\ 0 \text{ if not} \end{cases}$$

   b. $$LFP_i = \alpha + \beta Educ_i + \varepsilon_i$$

   c. $$E(LFP_i \mid Educ_i) = \Pr(LFP_i = 1 \mid Educ_i) = \alpha + \beta Educ_i$$

4. The model is *linear in the probability of y*

Example: LPM for labor force participation (-brm-lfp.do)

#1 Summary statistics

```
. use binlfp4, clear
(binlfp4.dta | Mroz data on labor force participation of women | 2014-10-20)

. regress lfp c.k5 c.k618 i.agecat i.wc i.hc c.lwg c.inc

<snip>

         lfp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          k5 |  -.2827435   .0354894    -7.97   0.000    -.3524148   -.2130721
        k618 |  -.0125888   .0141043    -0.89   0.372    -.0402778    .0151003
      agecat |
      40_49  |  -.1232177   .0417838    -2.95   0.003    -.2052458   -.0411896
        50+  |  -.2663019   .0529478    -5.03   0.000    -.3702467    -.162357
          wc |
    College  |   .1616348   .0458621     3.52   0.000     .0716002    .2516693
          hc |
    College  |   .0235894   .0424713     0.56   0.579    -.0597884    .1069672
          lwg |   .1235474   .0302773     4.08   0.000     .0641082    .1829865
          inc |  -.0068515   .0015758    -4.35   0.000    -.0099451   -.0037578
        _cons |   .7060677   .0576341    12.25   0.000     .5929229    .8192124
------------------------------------------------------------------------------
```

Interpretations follow…

Unstandardized Coefficients for Continuous Variables:

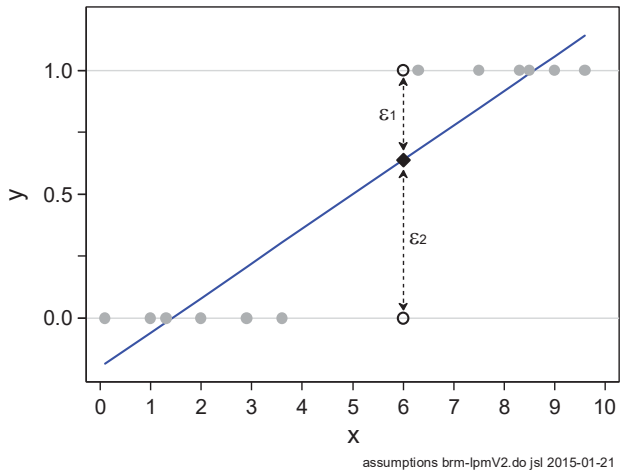
```
         lfp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          k5 |  -.2827435   .0354894    -7.97   0.000    -.3524148   -.2130721
```

For *each additional child* under six, the predicted probability of a woman being employed decreases by .28, holding other *variables constant*.

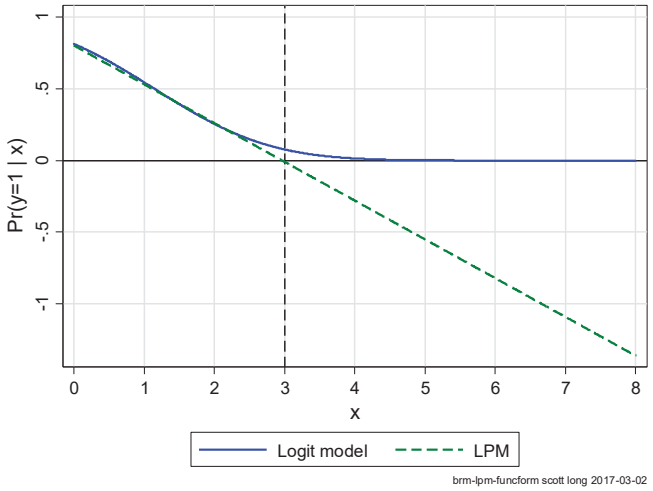
Problems with the LPM

Next page

LPM

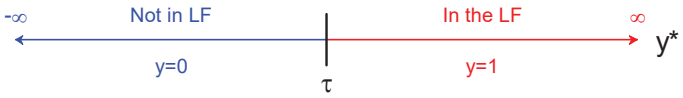


LPM versus BRM



BRM as a latent variable model

1. The *unobserved propensity to work* y^* generates the observed y.



2. y^* is linked to the observed y:

$$y_i = \begin{cases} 1 & \text{if } y_i^* > \tau \\ 0 & \text{if } y_i^* \le \tau \end{cases}$$

3. Not all women with y=1 are in the labor force with the same certainty or "propensity y* to work".
   a. One working woman is about to leave the labor force.
   b. Another working woman is firm in her decision to be in the labor force.
4. Near τ a tiny change in y^* changes the value of y.
5. Suppose, y* was propensity to vote? Propensity to engage in risky behavior?

What if you don't believe in a latent variables?

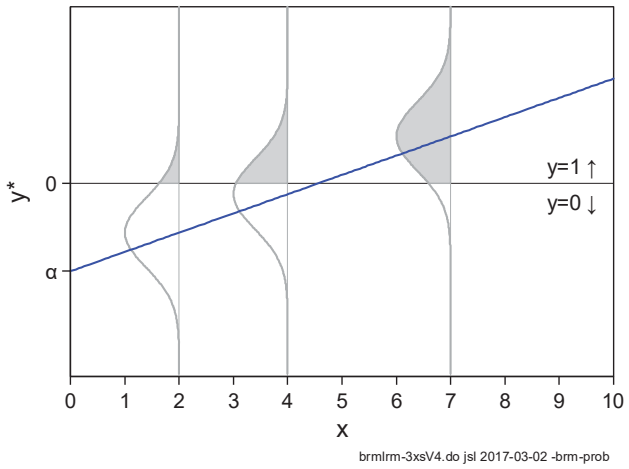
Identical statistical model can be derived four ways.

1. A *latent variable model*
2. A *probability model* without a latent variable
3. A *random utility model* (RUM) from economic theory
4. A *generalized linear model* (GLM)

Why use a latent variable model?

1. It builds on what you know about the LRM.
2. It generalizes to models not considered in this class.
   - Measurement models such as IRT.
   - Sample selection models.
   - Models for censored outcomes such as tobit.

The structural model y* = α + βx + ε



Where y* and ϵ are *both* latent.

Identification: the scale of y* cannot be estimated

What we can know about y affects interpretation in fundamental ways.*

1. Our structural model regresses latent y* on x:

   $y^* = \alpha + \beta x + \varepsilon$

2. Since y* and ε are unobserved, we *do not know* their means or variances.

What if someone doubled the *unobserved* y*?

1. Since the relationship with x is unchanged

   $2y^* = 2\alpha + 2\beta x + 2\varepsilon$

2. Changing notation, where underlining indicates "two times"

   $\underline{y}^* = \underline{\alpha} + \underline{\beta} x + \underline{\varepsilon}$

3. You cannot tell if β or $\underline{\beta}$ is the "true" parameter since you can't observe y*
4. We can't directly interpret the β's since we don't know the metric

Tools: the PDF and CDF

1. *y values*: -2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0
1. *PDF for y*: Probability density function

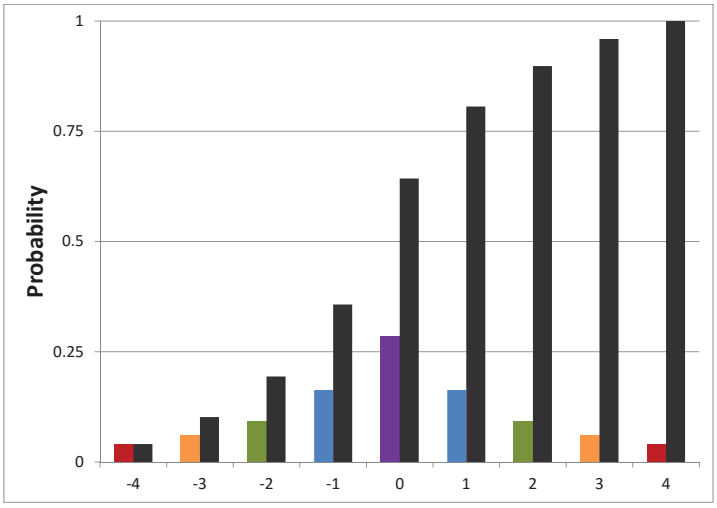
   For example: Pr(*y*=-2)
   Pr(*y* =-1.5)
   Pr(*y*=-1)

2. *CDF for y*: Cumulative density function is sum the PDFs up to a given value.

   For example: $\Pr(y \le -1.0) = \Pr(y = -2.0) + \Pr(y = -1.5) + \Pr(y = -1.0)$
   $\Pr(y \le -1.5) = \Pr(y = -2.0) + \Pr(y = -1.5)$

Graphically,...

The **PDF** is in colors; the **CDF** is in black (pdfcdf-example.xls)



Errors in the latent variable model

Normal errors for probit

1. **Normal PDF**: standard deviation σ

$$\varphi(\varepsilon_P;\mu=0,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-\varepsilon_P^2}{2\sigma^2}\right)$$

2. **Standardized normal PDF**: standard deviation σ=1

$$\varphi^S(\varepsilon_P)=\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-\varepsilon_P^2}{2}\right)$$

3. **Standardized normal CDF**

$$\Phi^S(\varepsilon_P)=\int_{-\infty}^{\varepsilon}\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-t^2}{2}\right)dt$$

Logistic errors for the logit model

1. **Standardized logistic PDF**: σ=1

$$\lambda^S(\varepsilon_L)=\frac{\frac{\pi}{\sqrt{3}}\exp\left(\frac{\pi}{\sqrt{3}}\varepsilon_L\right)}{\left[1+\exp\left(\frac{\pi}{\sqrt{3}}\varepsilon_L\right)\right]^2}$$

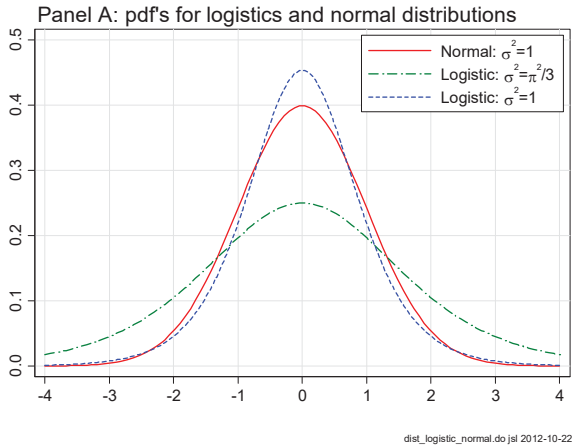
2. **Standard logistic PDF**: σ=π/√3=1.81...

$$\lambda(\varepsilon_L)=\frac{\exp(\varepsilon_L)}{\left[1+\exp(\varepsilon_L)\right]^2}$$

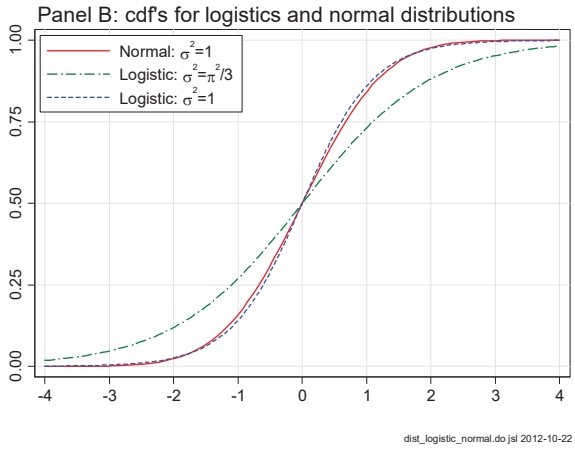
3. **Standard logistic CDF**: σ=π/√3=1.81...

$$\Lambda(\varepsilon_L)=\frac{\exp(\varepsilon_L)}{1+\exp(\varepsilon_L)}$$

PDF: Comparing logit and probit distributions



CDF: Comparing logit and probit distributions



Computing Pr(y=1|x)

We will now show where these formulas come from.

1. For *probit* with standardized normal errors

$$\Pr(y=1\,|\,\mathbf{x}) = \Phi(\mathbf{x\beta}) = \int_{-\infty}^{\mathbf{x\beta}} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right) dt$$

2. For *logit* with standard logistic errors

$$\Pr(y=1\,|\,\mathbf{x}) = \Lambda(\mathbf{x\beta}) = \frac{\exp(\mathbf{x\beta})}{1+\exp(\mathbf{x\beta})}$$

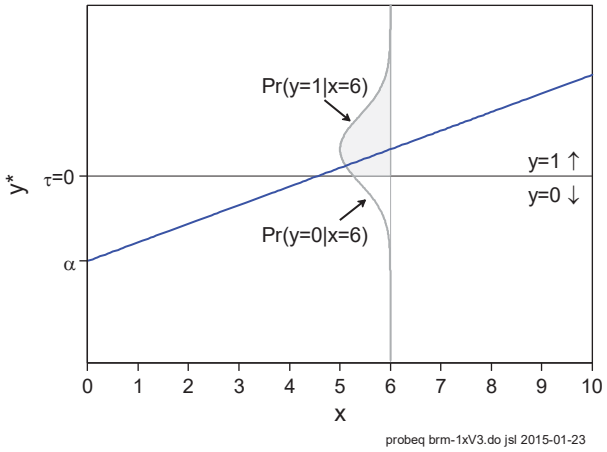
3. Generally, using π() as shorthand for Pr(y=1|)

$$\pi(\mathbf{x\beta}) = \Pr(y=1\,|\,\mathbf{x}) = F(\mathbf{x\beta})$$

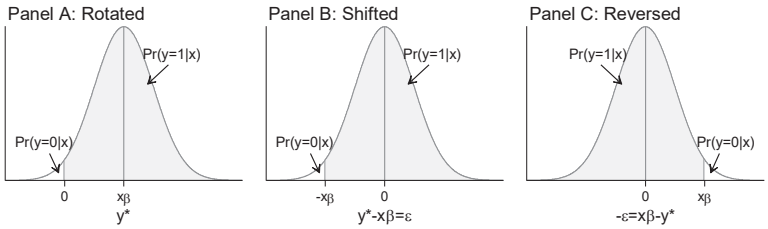
4. Where do these formula come from?

Graphically...

We need a formula for the shaded region



1. The tedious algebra involves these steps...



2. Resulting in

Probit: $$\Phi(\mathbf{x\beta}) = \int_{-\infty}^{\mathbf{x\beta}} \frac{1}{\sqrt{2\pi}} \left(\frac{-t^2}{2} \right) dt$$

Logit: $$\Lambda(\mathbf{x\beta}) = \frac{\exp(\mathbf{x\beta})}{(1+\exp(\mathbf{x\beta}))}$$

y* and Pr(y=1|x) for a single regressor

1. The structural equation is:

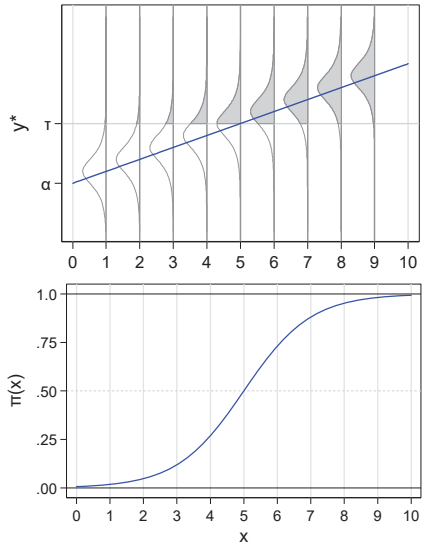
$$y^* = \alpha + \beta x + \varepsilon \quad \text{where } \varepsilon \sim N(0,1)$$

2. The probability equation is:

$$\Pr(y = 1 \mid x) = F(\alpha + \beta x)$$

3. The link between *y** and Pr(*y*=1) leads to the classic S-shaped curve relating *x* to Pr(*y*=1|*x*).

Next page...

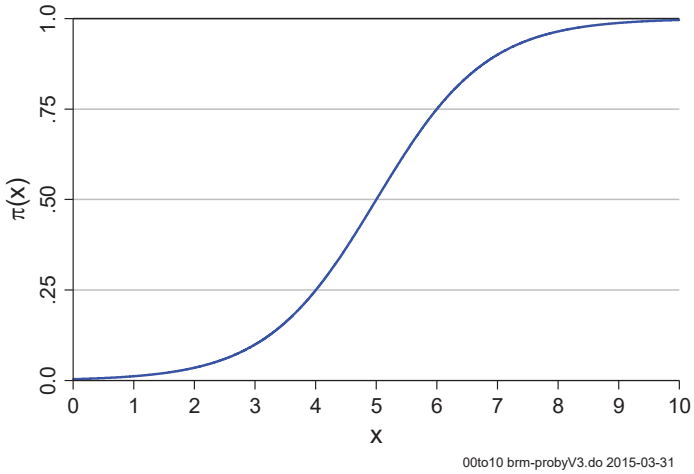


Does the empirical relationship need to be S-shaped?

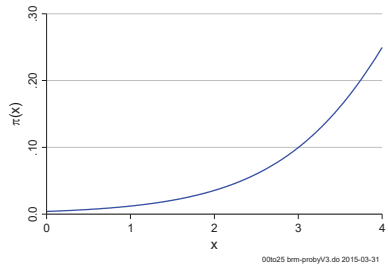
1. Does anyone in the sample need to have a probability near 0 or 1?
2. Can the relationship be linear?
3. Can changes in x's change the probability from 0 to 1 in the sample?

Consider the probability curve...

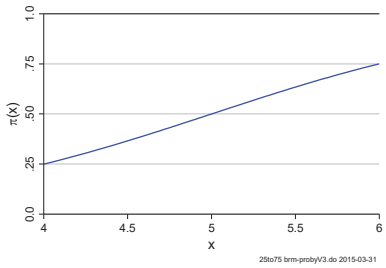
Where is your observed data located?



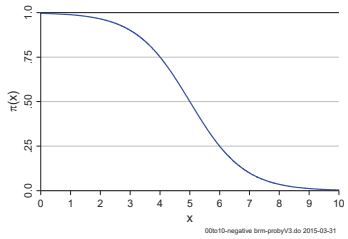
Only low probabilities

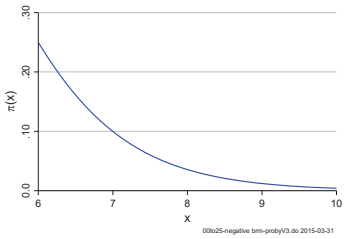


Change is linear



Negative relations are possible

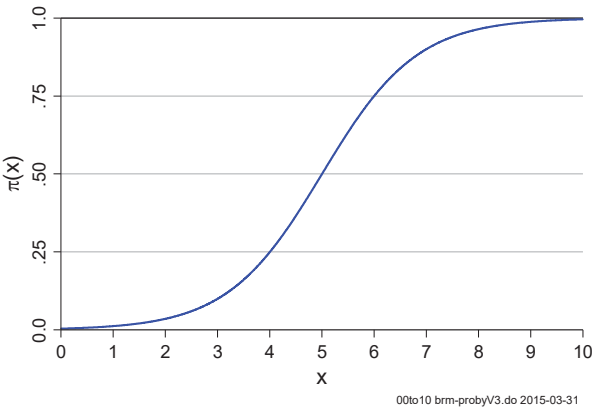




And so on...

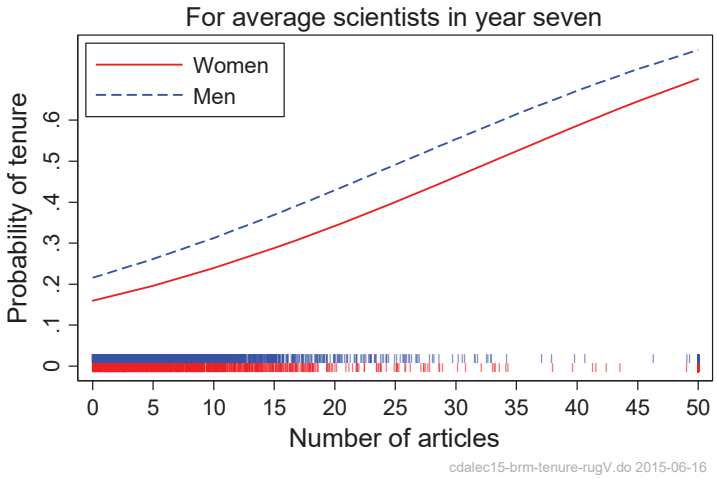
On the support of the data

Where is your data? Where do you want to explore



For example,...

How confident are you about the predictions?



Scale change and regression coefficients

The following tools and ideas are essential for understanding identification.

The variance and rescaling

$$Var(x) = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}$$

Scale change

$$Var(b\,x) = b^2 Var(x)$$

$$SD(b\,x) = b\,SD(x)$$

Normalizing a variable

$$Var\left(\frac{1}{\sigma}x\right) = \frac{1}{\sigma^2}Var(x) = 1$$

#3 Rescaling by a factor of 100 (-brm-science.do)

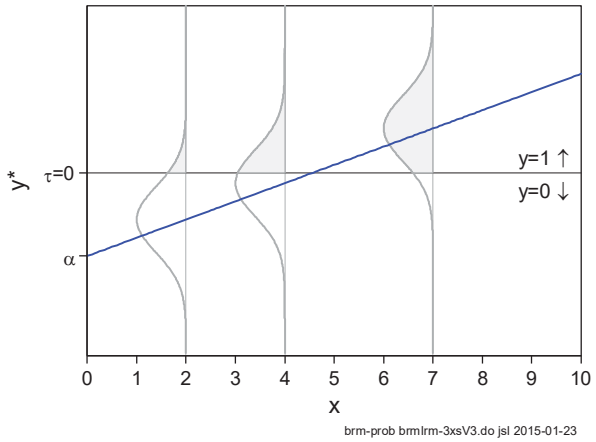
```
. gen job100=100*job
. label var job100 "job*100"
. gen phd100=100*phd
. label var phd100 "phd*100"

. sum job job100 phd phd100 pub1 nopub9 female

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         job |       163    2.967117     .880396        1.01       4.69
       job100 |       163    296.7117     88.0396        101        469
-------------+--------------------------------------------------------
         phd |       308    3.177987    1.012738          1       4.77
      phd100 |       308    317.7987    101.2738        100        477
-------------+--------------------------------------------------------
        pub1 |       308    2.545455    3.092685          0         24
      nopub9 |       308    .1980519    .3991801          0          1
      female |       308    .3474026    .4769198          0          1
```

#4 LRM with rescaling

```
. qui regress job phd pub1 female
. estimates store m1job
. qui regress job100 phd pub1 female
. estimates store m2job100
. estimates table m1job m2job100, stats(r2 ll) b(%8.3f) t(%8.2f)

-----------------------------------
    Variable |    m1job   m2job100
-------------+---------------------
         phd |     0.357     35.709
             |      5.54       5.54
        pub1 |     0.032      3.212
             |      1.61       1.61
      female |    -0.246    -24.570
             |     -1.65      -1.65
       _cons |     1.736    173.555
             |      7.62       7.62
-------------+---------------------
          r2 |     0.190      0.190
          ll |  -192.819   -943.462
-----------------------------------
                        legend: b/t
```


Identification in the BRM

Scaling and **identification** are critical for understanding the BRM



Identifying assumptions

Three *arbitrary* but *necessary* identifying assumptions:

Assumption 1: Value of threshold

$\tau = 0$

Assumption 2: Mean of the errors

$E(\varepsilon \mid \mathbf{x}) = 0$

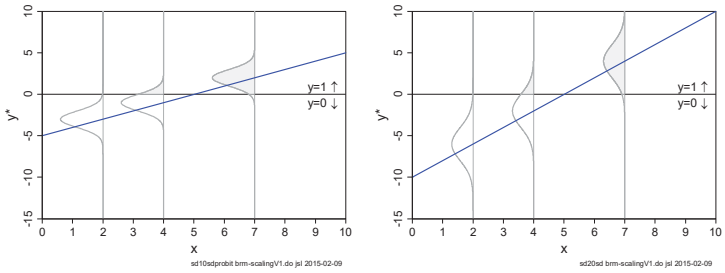
Assumption 3: Variance of the errors

$Var(\varepsilon \mid \mathbf{x}) = 1$ for probit

$Var(\varepsilon \mid \mathbf{x}) = \pi^2 / 3$ for logit

Graphical illustration of identification assumption 3

1. The β's cannot be interpreted directly since their magnitude reflects:
   a. The relationship between the *x*'s and *y**.
   b. *Arbitrary* identifying assumptions.
2. Pr(*y*=1|**x**) is unaffected by the identifying assumption about Var(ε |**x**).



3. Demonstration: see *CDAlec BRM ident prob demo 2017-03-02.docx*

#3 Comparing logit and probit with Mroz data (-brm-lfp.do)

Estimate the two models

```
. // logit
. logit lfp k5 k618 i.agecat i.wc i.hc lwg inc, nolog robust

<snip>

. estimates store blm

. // probit
. probit lfp k5 k618 i.agecat i.wc i.hc lwg inc, nolog

<snip>

. estimates store bpm

. // create table
. estimates table blm bpm, stats(aic bic r2_p ll) b(%8.3f) t(%8.3f) p(%8.3f)
```


Comparing estimates of coefficients

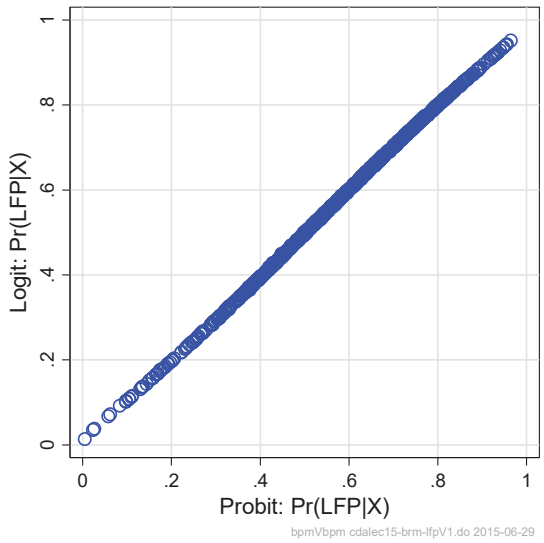
1. β's differ by a factor of about 1.7
2. z's are roughly equal

	blm		bpm		ratio	
	b	z	b	z	b	z
lfp						
k5	-1.392	-7.182	-0.840	-7.480	1.657	0.960
k618	-0.066	-0.916	-0.041	-0.975	1.593	0.939
1.agecat	.b	.	.b	.	.	.
2.agecat	-0.627	-3.042	-0.382	-3.107	1.643	0.979
3.agecat	-1.279	-4.956	-0.780	-5.031	1.640	0.985
0.wc	.b	.	.b	.	.	.
1.wc	0.798	3.367	0.482	3.481	1.655	0.967
0.hc	.b	.	.b	.	.	.
1.hc	0.136	0.659	0.074	0.596	1.841	1.106
lwg	0.610	3.677	0.371	3.894	1.644	0.944
inc	-0.035	-3.989	-0.021	-4.136	1.665	0.965
_cons	1.014	3.329	0.622	3.493	1.630	0.953

Predicted probabilities

```
. estimates restore blm
. predict prblm
(option pr assumed; Pr(lfp))
. label var prblm "Logit: Pr(LFP|X)"

. estimates restore bpm
(results bpm are active now)
. predict prbpm
(option pr assumed; Pr(lfp))
. label var prbpm "Probit: Pr(LFP|X)"

. pwcorr prblm prbpm

             |    prblm    prbpm
-------------+------------------
       prblm |   1.0000
       prbpm |   0.9998   1.0000

. twoway scatter prblm prbpm, ysize(7) xsize(7) mcol(blue) ///
>     msym(Oh) xlabel(0(.2)1,grid) ylabel(0(.2)1,grid) ///
>     caption("#24 `tag'", size(vsmall))
```

Graph follows...



Review of identification in BRM

1. The magnitude of the slopes depends on the scale of the outcome
2. Since *y** is latent, we do not know its scale (i.e., variance)
3. Therefore, *the slopes are not identified*
4. The estimated β's *cannot be directly interpreted* since they reflect
   a. The relationship between the *x*'s and *y**
   b. The arbitrary identifying assumption for Var(ε|**x**)
5. *The identifying assumption does not affect Pr(y=1|x)*
   - Probabilities can be interpreted without concern about identification
6. This identification issue has profound implications for interpreting the models
   - Group comparisons
   - Nested models
   - Mediation effects

* Alternative derivations of the BRM

Probability model

$$\ln\left[\frac{\Pr(y=1|\mathbf{x})}{\Pr(y=0|\mathbf{x})}\right]=\ln\left[\frac{\Pr(y=1|\mathbf{x})}{1-\Pr(y=1|\mathbf{x})}\right]=\mathbf{x}\boldsymbol{\beta}$$

BRM as a Random Utility Model (RUM)

1. Two choices where

   Choice 0 provides utility u_{0i}

   Choice 1 provides utility u_{1i}

2. The utility received from a choice is modeled as

   $u_{0i} = \mathbf{x}_i\boldsymbol{\beta}_0 + \varepsilon_{0i}$

   $u_{1i} = \mathbf{x}_i\boldsymbol{\beta}_1 + \varepsilon_{1i}$

3. A person chooses 0 if $u_{0i} > u_{1i}$ with $\Pr(u_{0i} > u_{1i}|\mathbf{x})=\Pr(0|\mathbf{x})$

BRM as a generalized linear model (GLM)

1. The observed y has a binomial distribution with mean μ
2. The linear predictor is η = **xβ**
3. Link function for logit is $\ln[\mu/(1-\mu)] = \eta = \mathbf{x}\boldsymbol{\beta}$ and for probit is $\Phi^{-1}(\mu) = \eta = \mathbf{x}\boldsymbol{\beta}$

All lead to the same estimates and predictions

ML estimation

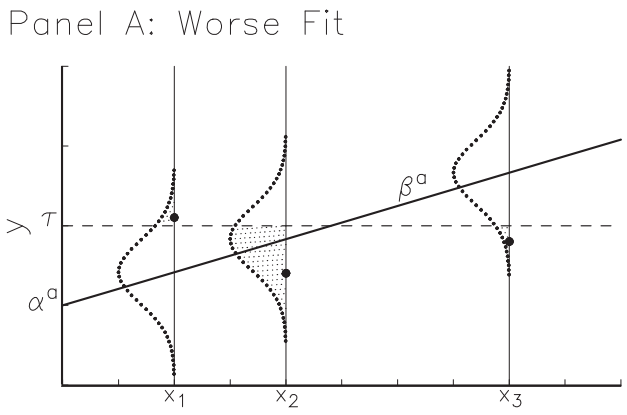
1. Probability of what was observed for each observation

$$p_i=\begin{cases}\Pr(y_i=1|\mathbf{x}_i) & \text{if } y_i=1 \text{ is observed}\\ 1-\Pr(y_i=1|\mathbf{x}_i) & \text{if } y_i=0 \text{ is observed}\end{cases}$$

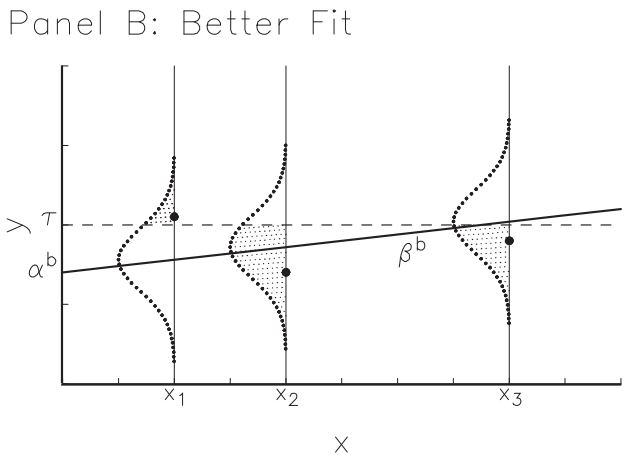
2. If observations are independent, Pr(HH) = Pr(H)*Pr(H). Thus,

$$L(\boldsymbol{\beta}|\mathbf{y},\mathbf{X})=\prod_{i=1}^{N}p_i$$

MLE shown graphically: a worse fit



MLE shown graphically: a better fit



MLE and sample size

1. *Consistency*, *normality*, and *efficiency* are asymptotic properties
2. ML estimators are *not necessarily bad in small samples*
3. But small sample behavior is largely unknown

When is the N large enough to justify MLE?

1. It is risky to use MLE for N<100. N>500 is generally safe
2. N's should be larger in some cases
    a. If there are a *lot of parameters*, more observations are needed
        - At least ten observations per parameter seems reasonable
    b. If data are *ill-conditioned* or if there is little variation in the dependent variable, a larger N is required
3. Some models seem to require more observations (e.g., ordinal regression)
4. Small depends on the size of the smallest outcome category. "Rare events" methods deal with this.

Adjusting p-values for small samples?

1. In small samples do *not* use *larger* p-values as evidence against the null hypothesis
2. Since the degree to which MLEs are normal is unknown, it is reasonable to require *smaller* p-values in small samples

Exact and Firth estimation for small samples

1. ML is biased in small samples
2. **`exlogistic`** computes exact estimates in small samples but is computationally intensive

   R.A. Fisher devised exact tests when Muriel Bristol claimed she could detect whether tea or milk was added first to her cup. He asked her to taste 8 cups of tea, 4 with tea added first and 4 not, and decide which was added first. He had to figure out how to perform a test with so few observations. Essentially, he counted every possible outcome.
3. Penalized maximum likelihood (Firth estimation) is a computationally simpler way to address small sample bias

Maximizing the likelihood and numerical methods

1. Algebraic maximization of $ln\ L(\boldsymbol{\beta}|\mathbf{X},\mathbf{y})$ is rarely possible
2. Numerical methods search for the maximum using the slope and change in slope of the likelihood equation (i.e., first and second derivatives)
3. The process corresponds to what you would do to find the top of a hill if you were blindfolded
   - What would it take to make sure you were at the top?
   - What would you want to know before playing this game?
   - Will you end up at the same place as another person? Why? Why not?
   - How big of a step will you take? Always the same?
   - Why would you play this silly game?
4. Estimates of coefficients are usually very close in different software, with perhaps small differences in standard errors

Possible problems with ML

1. A flat likelihood function makes convergence difficult. Errors might be: (a) *Convergence not obtained* after 250 iterations; (b) Hessian *not of full rank.*
2. *Little variation in the outcome* or *ill conditioned data* cause problems
3. Some models (not the BRM), have a *local maxima*
4. *Perfect prediction* is a pseudo problem (-brm-science.do).

```
. tabulate hipub mmale, miss
```

Pubs greater than 10?	Mentor male? 0FemMent	1MalMent	.	Total
0_LoPub	4	293	5	302
1_10plus	0	6	0	6
Total	4	299	5	308

   - The odds of LoPub if female mentor are 4/0 which is undefined.
   - The odds of 10plus if female mentor are 0/4=0.

Here is the logit results:

The four cases with female mentors are dropped by **logit**.

```
. logit hipub i.mmale phd, nolog

note: 0.mmale != 0 predicts failure perfectly
      0.mmale dropped and 4 obs not used
```

This means: female mentors are low publishers with probability 1.

```
note: 1.mmale omitted because of collinearity

Logistic regression                                     Number of obs   =        299
                                                        LR chi2(1)      =       0.23
                                                        Prob > chi2     =     0.6320
Log likelihood = -29.276794                             Pseudo R2       =     0.0039

------------------------------------------------------------------------------
       hipub |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       mmale |
   0FemMent  |          0  (empty)
   1MalMent  |          0  (omitted)
         phd |  -.1927085   .4023944    -0.48   0.632    -.9813871    .5959701
       _cons |  -3.293021   1.272882    -2.59   0.010    -5.787824   -.7982179
------------------------------------------------------------------------------
```


When problems occur with ML, what to do?

1. Check that the model is correctly specified
2. Verify that variables are correct
3. A large ratio between the largest and smallest standard deviations of regressors causes problems with ML
   - For example, rescale income in $1's to income in $1000's
4. If a very large proportion of cases are in one of the categories of the outcome, convergence may be difficult

Overall

1. Numerical methods for ML estimation work very well "when your model is appropriate for your data"
2. Cramer (1986:10) gives excellent advice

   Check the data, check their transfer into the computer, check the actual computations (preferably by repeating at least a sample by a rival program), and always remain suspicious of the results, regardless of the appeal.

Parameters and the probability curve

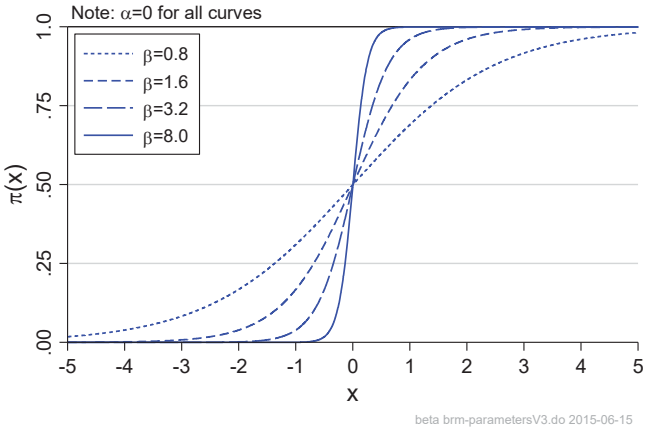
1. In the LRM
   a. The intercept moves the curve "up and down"
   b. The slope changes the rate of change
2. Consider the BRM with a single *x*:

Logit: $\Pr(y=1\,|\,x)=\Lambda(\alpha+\beta x)$

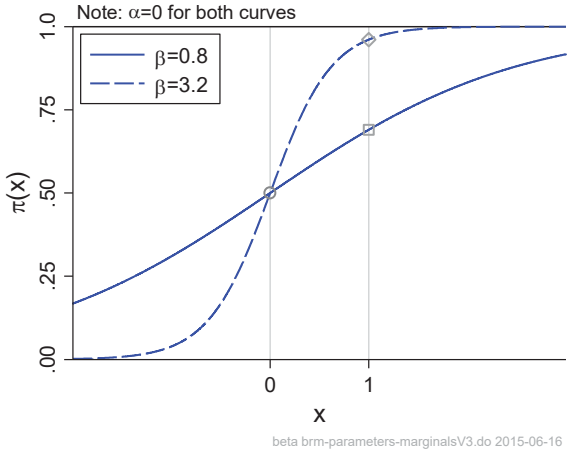
Probit: $\Pr(y=1\,|\,x)=\Phi(\alpha+\beta x)$

Changing the slope

Smaller the slope, greater change in X required for a given change in Pr(y).

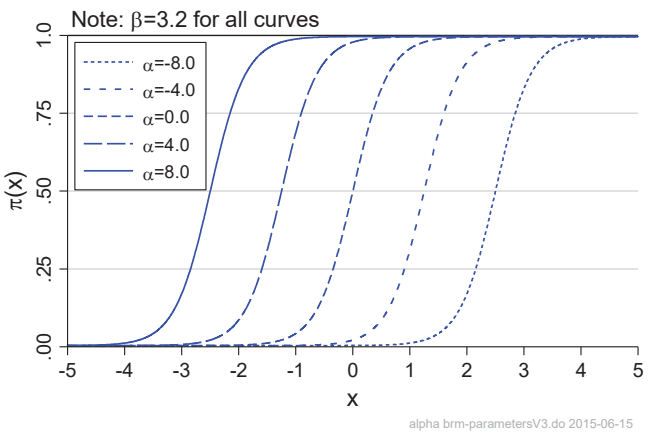


Changing β changes the effect of x

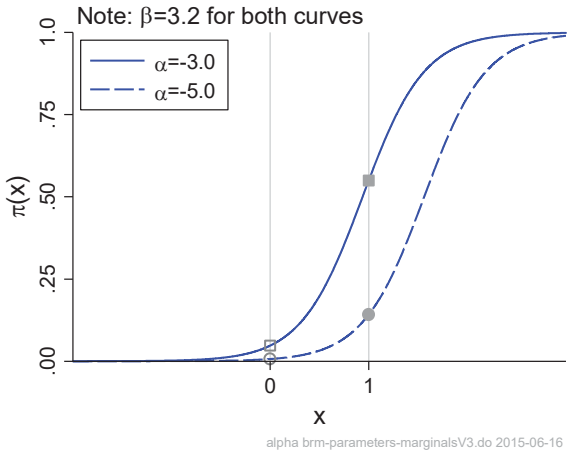


Changing the intercept

Smaller intercept, curve moves **right**.



Changing α changes the effect of x



How Z changes the effect of x

The intercept absorbs the level of x_2

1. The model is:

$$\Pr(y=1 \mid x_1, x_2) = \Phi(-4 + .6x_1 + .5x_2)$$

2. If x_2=0 (curve with squares on the next page):

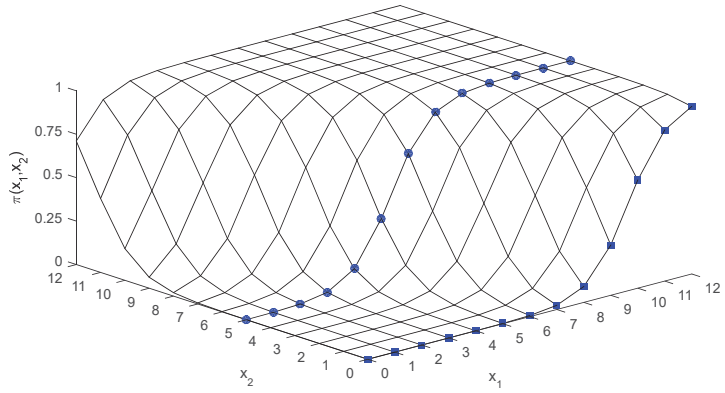
$$\Pr(y=1 \mid x_1, x_2 = 0) = \Phi(-4 + .6x_1 + [.5 \times 0]) = \Phi(-4 + .6x_1)$$

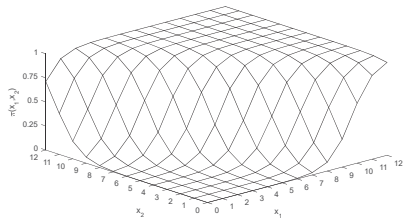
3. If x_2=5 (curve with circles on next page):

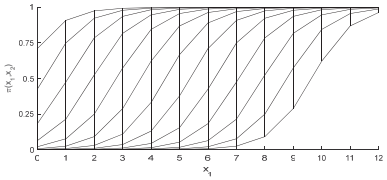
$$\Pr(y=1 \mid x_1, x_2 = 5) = \Phi(-4 + .6x_1 + [.5 \times 5]) = \Phi([-4 + 2.5] + .6x_1) = \Phi(-1.5 + .6x_1)$$

4. When looking at the effect of x_1, values of other variable change the intercept.
5. *Graphically...*

How x_2 "affects the effect" of x_1







Interpretation using predictions

1. Probabilities are the fundamental statistic for interpreting the BRM:

$$\text{Logit: } \widehat{\Pr}(y=1\,|\,\mathbf{x}) = \Lambda\left(\mathbf{x}\hat{\boldsymbol{\beta}}\right) = \frac{\exp\left(\mathbf{x}\hat{\boldsymbol{\beta}}\right)}{1+\exp\left(\mathbf{x}\hat{\boldsymbol{\beta}}\right)} = F\left(\mathbf{x}\hat{\boldsymbol{\beta}}\right)$$

$$\text{Probit: } \widehat{\Pr}(y=1\,|\,\mathbf{x}) = \Phi\left(\mathbf{x}\hat{\boldsymbol{\beta}}\right) = \int_{-\infty}^{\mathbf{x}\hat{\boldsymbol{\beta}}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt = F\left(\mathbf{x}\hat{\boldsymbol{\beta}}\right)$$

2. Since the model is nonlinear, no single method of interpretation can fully describe the relationship between a variable and the outcome.
3. Search for an elegant method that reflects the substantive complexities.
4. The critical decision is at which values of **x** you want to examine the predictions.

Let's explore this...

Value of the regressors for Pr(y=1|x)

1. *In-sample predictions* use *x*-values from the sample: Pr(y=1 | $\mathbf{x}_i$)
2. *Out-sample* or *counterfactual predictions* use any values of ***x***

On the support When making counterfactual predictions, the values should be where real data might be found now or in the future.

Counterfactual experiments involves imagining the value of a variable changes while other variables are held constant.

Average: Is an *average person* a reasonable counterfactual? Even if that person is .53 female?

Ways of using predictions for interpretation

1. *Predictions at observed values*
2. *Marginal effects*: changes in predictions
3. *Ideal types* or *profiles*: predictions at values of substantive interest
4. *Tables*: predictions at multiple levels of regressors
5. *Graphs*: predictions at many levels of regressors
6. *Odds ratios*: ratios of predicted probabilities

Commands for predictions

Official Stata 11+

`predict`: Predictions for observations in the dataset

`margins`: Predictions at specific values or averaged over observed values

`marginsplot`: Plot predictions from **`margins`**

SPost13

1. SPost13 has m* commands that are "wrappers" to make margins easier and more powerful. The most important commands are:

   `mchange`: Changes in predictions

   `mgen`: Predictions as one variable changes over a range

   `mtable`: Tables of predictions

   `mlincom`: Tests of predictions

2. They work with most models and with complex surveys.

*Tool: specifying values of regressors in margins and m**

atmeans: all regressors at their means.

```
mtable, atmeans
```

at() for single values of regressors

```
mtable, at(age=25 male=1 edyears=20) atmeans
```

Variables not specified are held at their mean.

at() with linked variables

```
mtable, at(age=25) atmeans
```

If **`c.age#c.age`** is a regressor, predictions are made at 25*25 for age-squared.

at() for multiple values using a *numlist*

```
mtable, at(age=(25(5)75) male=1 edyears=20) atmeans
```

Predictions are computed for age = 25, 30, 35, etc.

at() at multiple specified values

```
mtable, at(age=25 male=1 edyears=20) ///
        at(age=60 male=0 edyears=12) atmeans
```

In-sample predicted probabilities

1. *In-sample predictions* use $\mathbf{x}_i$ values from the sample

$$\text{Logit: } \widehat{\Pr}(y_i = 1 \mid \mathbf{x}_i) = \Lambda\left(\mathbf{x}_i\hat{\boldsymbol{\beta}}\right) = \frac{\exp\left(\mathbf{x}_i\hat{\boldsymbol{\beta}}\right)}{1+\exp\left(\mathbf{x}_i\hat{\boldsymbol{\beta}}\right)}$$

$$\text{Probit: } \widehat{\Pr}(y_i = 1 \mid \mathbf{x}_i) = \Phi\left(\mathbf{x}_i\hat{\boldsymbol{\beta}}\right) = \int_{-\infty}^{\mathbf{x}_i\hat{\boldsymbol{\beta}}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

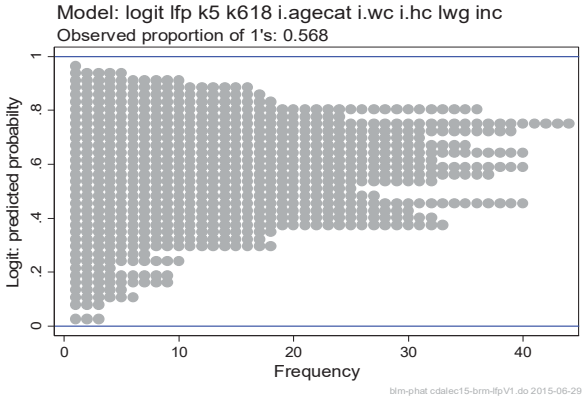
Start with the distribution of predictions

1. The range suggests how large the "effects" of regressors can be
2. Clumping suggest "types" of respondents or strong effects of categorical regressors
3. Outliers can indicate incorrectly coded variables.
4. If things stand out or are unexpected, figure out what is going on before interpreting the model.

#4 In-sample predictions (-brm-lfp.do)

```
estimates restore blm

predict prblm
label var prblm "Logit: predicted probability"

* mean prediction
qui sum prblm
local mn = string(r(mean),"%5.3f") // store formatted string

* distribution of predictions
dotplot prblm, ylab(0(.2)1, nogrid) ylin(0 1, lcol(blue)) mcol(gs10) ///
  title(Model: logit lfp k5 k618 i.agecat i.wc i.hc lwg inc, pos(11)) ///
  subtitle("Observed proportion of 1's: `mn'", position(11)) ///
  caption(#34a `tag', size(vsmall))
```

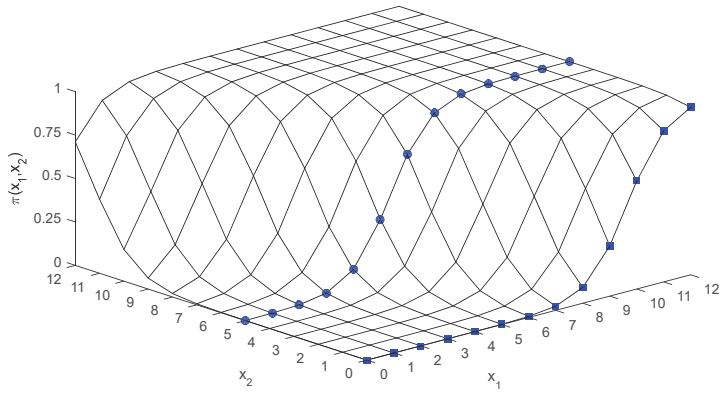

#34 dotplot of predicted probabilities



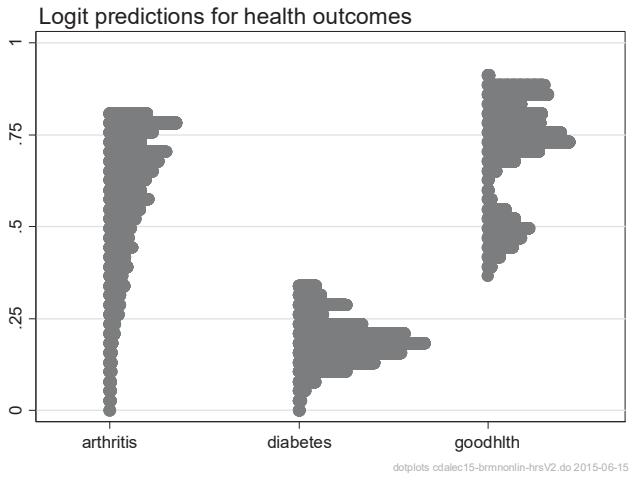
- Are there observations you want to explore?

What does this imply about "effects"? Consider the graph...

Effects are largest at π()=.5



Example of predictions for health outcomes (details later)



Marginal effects: changes in probabilities

1. A *marginal effect* is

   The change in Pr(y|**x**) for a change of δ in x_k holding other regressors at specific values.

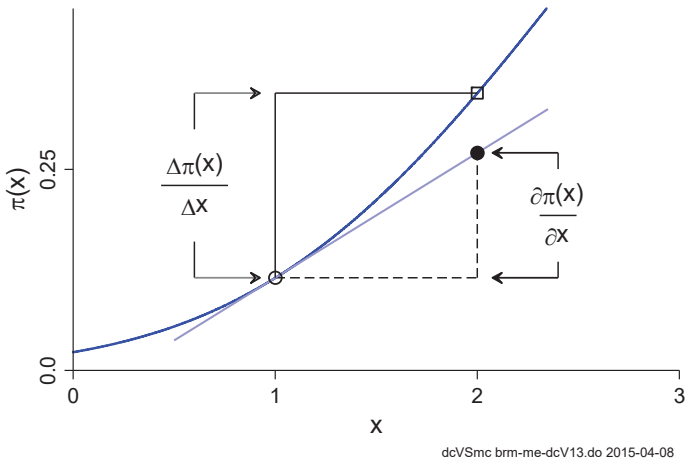
2. It is often the best summary of the effect of a variable.

Decisions when using MEs

1. How much change?
   - o An infinitely small change leads to the *marginal change*.
   - o A finite change leads to a *discrete change* or *first difference*.
2. Where is the change computed?
   - o The size of the ME depends on where it is computed.
3. How should the marginal effect be summarized?

These important issues are now examined, starting with a graph...

Marginal change and discrete change



- Speedometer reading compared to average speed for day.

Marginal change versus discrete change

1. **Marginal change** tells you how much the probability would change for a unit change in x_k *if the probability curve was linear*.
2. **Discrete change** is the change that occurs over a fixed distance.
3. The more nonlinear the curve near x_k, the greater the difference between the MC and the DC.
4. Unless your field uses MC, DC is more intuitive.

* Marginal change (MC)

1. Assuming there are no product terms (e.g., x1*x2) the MC is the partial derivative of Pr(y) with respect to x_k:

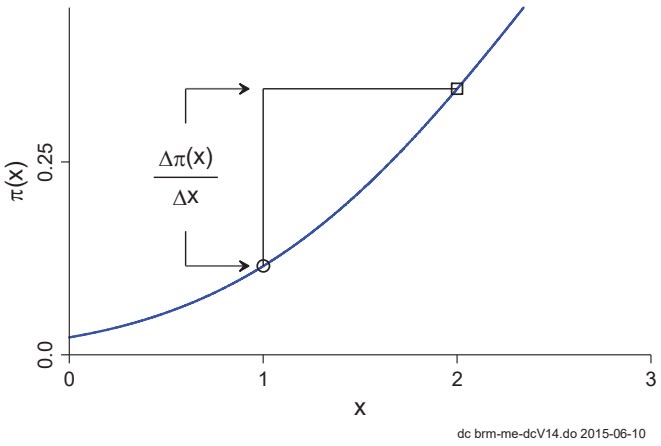
Logit: $$\frac{\partial \Pr(y=1|\mathbf{x})}{\partial x_k} = \lambda(\mathbf{x\beta})\beta_k = f(\mathbf{x\beta})\beta_k$$

Probit: $$\frac{\partial \Pr(y=1|\mathbf{x})}{\partial x_k} = \varphi(\mathbf{x\beta})\beta_k = f(\mathbf{x\beta})\beta_k$$

- *instantaneous rate of change in the probability* with respect to x_k holding other variables at specific values
- slope of the Pr(y=1|**x**) curve at x_k holding other variables a specific values

2. *Sign* of MC determined by β_k since f(**xβ**) is always positive.
3. *Magnitude* depends on β_k and f(**xβ**), thus by all variables and coefficients.

Discrete change (DC) or first difference



Here's how we compute the DC...

1. Compute probabilities at start and end values of x_k

   $\Pr(y=1 \mid \mathbf{x}^*, \text{Start } x_k)$: **Start** probability given **x*** & start value x_k.

   $\Pr(y=1 \mid \mathbf{x}^*, \text{End } x_k)$: **End** probability after changing only x_k.

2. Discrete change

$$\frac{\Delta \Pr(y=1 \mid \mathbf{x})}{\Delta x_k} = \Pr(y=1 \mid \mathbf{x}^*, \text{End } x_k) - \Pr(y=1 \mid \mathbf{x}^*, \text{Start } x_k)$$

3. Interpretation

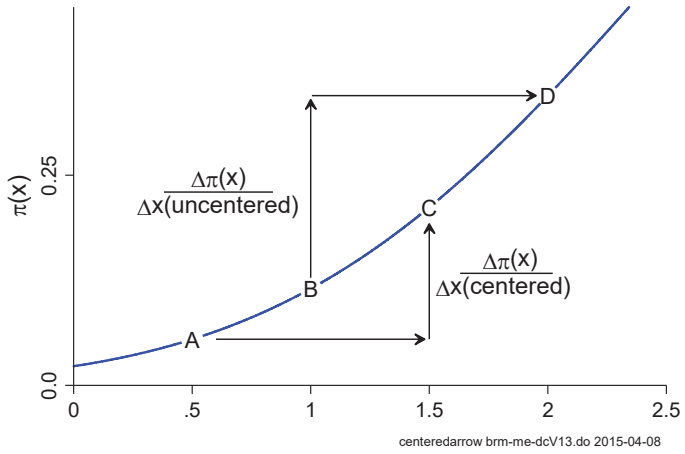
   - For a change in x_k from **start x_k** to **end x_k**, the probability changes by $\frac{\Delta \Pr(y=1 \mid \mathbf{x})}{\Delta x_k}$, holding other variables at the specific values.

4. Example

   Attending college increases the probability of women being in the labor force by .19, holding other variables at their means.

$$\frac{\Delta \Pr(y=1 \mid \mathbf{x})}{\Delta x_k} = \Pr(y=1 \mid wc=1, \bar{\mathbf{x}}) - \Pr(y=1 \mid wc=0, \bar{\mathbf{x}})$$

Centered and uncentered changes of 1



Mathematically, ...

Uncentered change of 1

$$\frac{\Delta \Pr\left(y=1 \mid \mathbf{x}^*\right)}{\Delta x_k} = \Pr\left(y=1 \mid \mathbf{x}^*, x_k^* + 1\right) - \Pr\left(y=1 \mid \mathbf{x}^*, x_k^*\right)$$

Centered change of 1

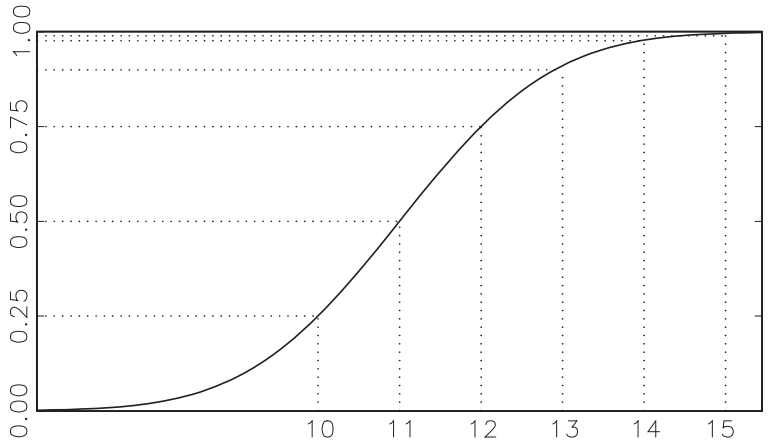
$$\frac{\Delta \Pr\left(y=1 \mid \mathbf{x}^*\right)}{\Delta x_k} = \Pr\left(y=1 \mid \mathbf{x}^*, x_k^* + \tfrac{1}{2}\right) - \Pr\left(y=1 \mid \mathbf{x}^*, x_k^* - \tfrac{1}{2}\right)$$

Overview of what affects the size of the marginal effect?

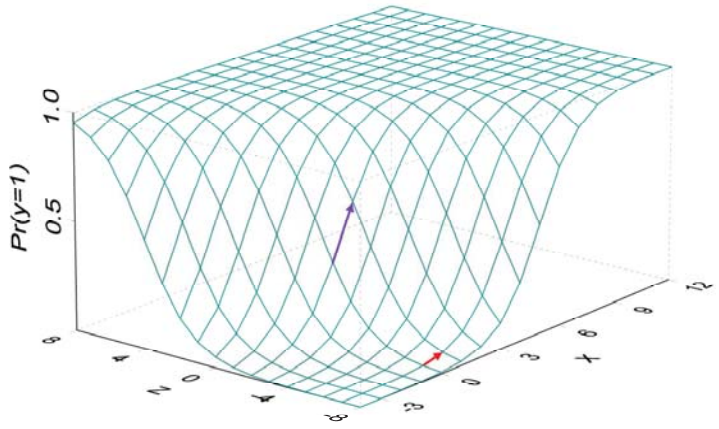
1. The regression coefficient β_k
   : The larger the magnitude the larger the effect
2. Start value of x_k
   : The curve changes more rapidly at some places
3. The amount of change in x_k
   : Bigger changes have bigger effects (assuming no polynomials)
4. Value at which other variables are evaluated along with their coefficents
   : These change the intercept which changes the effect

Graphically...

Effect of start value on DC of 1



Effect of other variables on DC of 1



Options for amount of change in x_k

1. Infinitely small
2. 0 to 1 for binary variables
3. Unit change
4. Standard deviation change
5. Minimum to maximum
6. Four years of education, $10,000, or whatever makes substantive sense
7. Changes in several variables, such as white males to black females
8. Changing linked variables

Summarizing the marginal effects

1. The ME depends on the levels of *all* variables in the model
2. Where to hold variables constant and how to summarize this variation is an important *substantive decision*.

Common summary measures

1. **Marginal effects at the mean** (MEM) is the effects with all variables at their means
   - Is anyone average?
   - *Skewed variables*
   - Mean of a *dummy variable*?
2. **Marginal effects at substantively representative values** (MER)
   - At values that are representative of substantive interests
3. **Average marginal effect** (AME) is the mean across all observations
   - Compute ME for each observation and average

Marginal effect at the mean (MEM)

Often used, perhaps due to ease of computation and tradition.

1. Hold all variables held at theirs means:

$$MCM: \quad \frac{\partial \Pr\left(y=1 \mid \bar{\mathbf{x}}\right)}{\partial x_k} = f(\bar{\mathbf{x}}\boldsymbol{\beta})\beta_k \qquad DCM: \quad \frac{\Delta \Pr\left(y=1 \mid \bar{\mathbf{x}}\right)}{\Delta x_k}$$

2. Is the mean representative of what you want to know?

Marginal effect at representative values (MER)

1. Think of a specific set of values **x*** and compute the ME there.

$$MCR: \quad \frac{\partial \Pr\left(y=1 \mid \mathbf{x}^*\right)}{\partial x_k} \qquad DCR: \quad \frac{\Delta \Pr\left(y=1 \mid \mathbf{x}^*\right)}{\Delta x_k}$$

2. Maddala in 1980s recommended using MER's at multiple locations
   - We apply this by creating "ideal types" and computing MER's for these

Average marginal effect (AME)

1. Compute the ME at each $\mathbf{x}_i$. For example:

$$MC_i: \frac{\partial \Pr\left(y=1 \mid \mathbf{x}_i\right)}{\partial x_{ik}} \qquad DC_i: \frac{\Delta \Pr\left(y=1 \mid \mathbf{x}_i\right)}{\Delta x_{ik}}$$

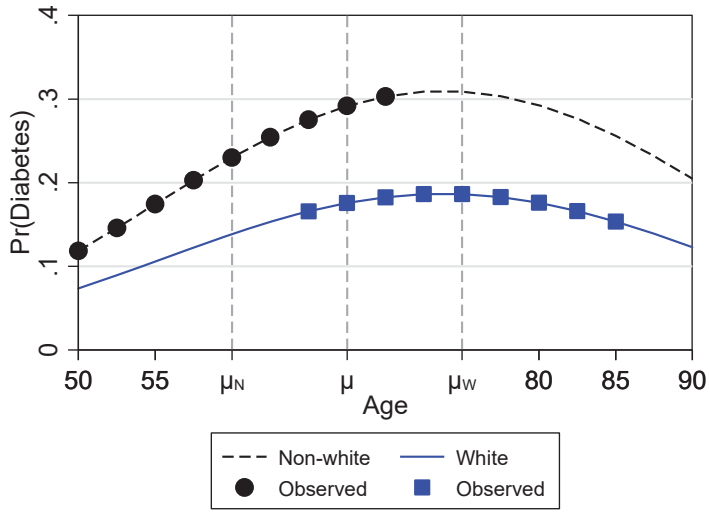
2. AME averages over all cases:

$$AMC = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial \Pr\left(y=1 \mid \mathbf{x}_i\right)}{\partial x_{ik}} \quad ADC = \frac{1}{N}\sum_{i=1}^{N}\frac{\Delta \Pr\left(y=1 \mid \mathbf{x}_i\right)}{\Delta x_{ik}}$$

3. Generally, AME's are the most useful summary measure.

Which measure of change? AME, MEM, MER

1. AME and MEM can be similar, but are **not** asymptotically equivalent
2. Traditionally, MEM prevailed
   a. AME requires N times more computation
   b. MEM was in common software like SPost9
3. Newer software computes both measures
4. A critique MEM is that the mean might not correspond to anyone
   a. Nobody is .47 female.
   b. But the ME at the mean of a binary variable roughly averages the ME for the two groups.
   c. MER can use modal values of the binary variables, but this ignores everyone who is in a less well represented group.
5. Consider this example…



diabetes-youngW groups-didactic-AMEvMEMV8.do 2016-06-22

Limitations of the AME

1. *The AME replaces one mean with another.*
   a. Computation **at the mean** is replaced by **the mean of**.
   b. Means are only one characteristic of a distribution.
2. The AME might not be close to the effect for anyone in the sample.
   a. Suppose effects are small for men and large for women. The AME does not indicate this difference.
   b. If you are planning an intervention, are you interested in the average effect or the average for those you want to target (e.g., high risk youth)?

What do you want to know?

1. The best measure is the one that addresses the goals of your research
2. Think about what you want to know

mchange computes marginal effects

Overview of mchange

```
. mchange, atmeans dec(2)
logit: Changes in Pr(y) | Number of obs = 753
Expression: Pr(lfp), predict(pr)
                |    Change    p-value
----------------+----------------------
k5              |
             +1 |     -0.32       0.00
            +SD |     -0.18       0.00
:::
Predictions at base value
             | not in LF      in LF
-------------+----------------------
  Pr(y|base) |      0.42       0.58
Base values of regressors
            |                          2.         3.         1.         1.
            |       k5       k618     agecat     agecat     wc         hc
------------+-------------------------------------------------------------
         at |      .24        1.4        .39        .22        .28        .39
            |
            |      lwg        inc
------------+----------------------
         at |      1.1         20
```

Hypothesis testing for marginal effects

1. Standard errors are computed by *delta method*
2. You can test if change is 0 or to compute a confidence interval
   - Is the effect of having another child significant?
3. And more test complex hypotheses, such as the equality of effects by group
   - Is effect of age the same for men and women?

Examples of marginal effects (-brm-lfp.do)

#5.1 MEM: marginal effects at the mean

```
. mchange, atmeans dec(2) amount(one sd)

logit: Changes in Pr(y) | Number of obs = 753

Expression: Pr(lfp), predict(pr)

                |    Change    p-value
----------------+----------------------
k5              |
             +1 |     -0.32       0.00
            +SD |     -0.18       0.00
       Marginal |     -0.34       0.00
k618            |
             +1 |     -0.02       0.34
            +SD |     -0.02       0.34
       Marginal |     -0.02       0.34
agecat          |
 40-49 vs 30-39 |     -0.15       0.00
   50+ vs 30-39 |     -0.31       0.00
   50+ vs 40-49 |     -0.16       0.00
wc              |
  college vs no |      0.19       0.00
hc              |
  college vs no |      0.03       0.51
```



```
lwg             |
             +1 |      0.14       0.00
            +SD |      0.08       0.00
       Marginal |      0.15       0.00
inc             |
             +1 |     -0.01       0.00
            +SD |     -0.10       0.00
       Marginal |     -0.01       0.00

Predictions at base value

             | not in LF      in LF
-------------+----------------------
  Pr(y|base) |      0.42       0.58

Base values of regressors

          |                                 2.         3.         1.         1.
          |       k5       k618     agecat     agecat         wc         hc
----------+-------------------------------------------------------------------
       at |       .24        1.4        .39        .22        .28        .39

          |      lwg        inc
----------+----------------------
       at |      1.1         20

1: Estimates with margins option atmeans.
```

A unit change: +1

$$\frac{\Delta \Pr\left(y=1 \mid \mathbf{x}^*\right)}{\Delta x_k} = \Pr\left(y=1 \mid \mathbf{x}^*, x_k^*+1\right) - \Pr\left(y=1 \mid \mathbf{x}^*, x_k^*\right)$$

1. A unit increase in x_k from x_k^* to x_k^*+1 results in a change of $\frac{\Delta \Pr}{\Delta x_k}$ in the predicted probability, holding other variables at the specified values.
2. For example:

```
                |    Change    p-value
----------------+----------------------
k5              |
          +1    |     -0.32       0.00
```

For a woman who is average on all characteristics, an additional young child decreases the probability of being in the labor force by .32 (p<.01).

The additional child is added to the average number of children (.24), which isn't ideal. More on this later.

A standard deviation change: +SD

1. The effect of a standard deviation change:

$$\frac{\Delta \Pr\left(y=1 \mid \mathbf{x}^*\right)}{\Delta x_k} = \Pr\left(y=1 \mid \mathbf{x}^*, x_k^*+s_k\right) - \Pr\left(y=1 \mid \mathbf{x}^*, x_k^*\right)$$

2. For example:

```
                |    Change    p-value
----------------+----------------------
inc             |
            +1  |     -0.01       0.00
           +SD  |     -0.10       0.00
```

A standard deviation increases in family income, about $20,000, decreases the probability of being in the labor force by .10 (p<.01, two-tailed test), with variables held at their means.

A change from 0 to 1: 0 to 1

1. Binary variables were entered as (for example) **`i.wc`** so **`mchange`** knows to change them from 0 to 1.
2. For example,

```
                |    Change    p-value
----------------+----------------------
wc              |
  college vs no |      0.19       0.00
hc              |
  college vs no |      0.03       0.51
```

If an average woman attends college, her probability of being in the labor force is .19 greater than that of a woman who does not attend college (p<.01).

#5.2 Centered changes in x_k

$$\frac{\Delta \Pr\left(y=1 \mid \mathbf{x}^*\right)}{\Delta x_k} = \Pr\left(y=1 \mid \mathbf{x}^*, x_k^* + \tfrac{1}{2}\right) - \Pr\left(y=1 \mid \mathbf{x}^*, x_k^* - \tfrac{1}{2}\right)$$

$$\frac{\Delta \Pr\left(y=1 \mid \mathbf{x}^*\right)}{\Delta x_k} = \Pr\left(y=1 \mid \mathbf{x}^*, x_k^* + \tfrac{s_k}{2}\right) - \Pr\left(y=1 \mid \mathbf{x}^*, x_k^* - \tfrac{s_k}{2}\right)$$

```
. mchange, atmeans dec(2) centered
<snip>
               |    Change    p-value
---------------+----------------------
k5             |
    +1 centered |     -0.33       0.00
   +SD centered |     -0.18       0.00
      Marginal |     -0.34       0.00
<snip>
inc            |
    +1 centered |     -0.01       0.00
   +SD centered |     -0.10       0.00
      Marginal |     -0.01       0.00
<snip>
```

For example…

1. For k5

   For a woman who is average on all characteristics, an additional young child centered around the mean decreases the probability of being in the labor force by .33 (p<.01).

   The added child is centered on the average number of children (.24), which leads to a negative start value!

2. For inc

   A standard deviation change in family income (about $20,000) centered around the mean income increases the probability of being in the labor force by .10 (p<.01, two-tailed test), with other variables held at their means.

Change from the minimum to the maximum

Even when the change is substantively unrealistic, this is a useful measure of the total possible effect of a variable:

$$\frac{\Delta \Pr\left(y=1 \mid \mathbf{x}^*\right)}{\Delta x_k} = \Pr\left(y=1 \mid \mathbf{x}^*, \max x_k\right) - \Pr\left(y=1 \mid \mathbf{x}^*, \min x_k\right)$$

```
. mchange k5 k618 wc hc lwg inc agecat, ///
>     atmeans amount(range) stat(from to change pvalue) dec(2) brief

logit: Changes in Pr(y) | Number of obs = 753

Expression: Pr(lfp), predict(pr)

               |      From         To     Change    p-value
---------------+-----------------------------------------
k5             |
         Range |      0.66       0.03      -0.63       0.00
k618           |
         Range |      0.60       0.47      -0.13       0.34
wc             |
  college vs no |      0.52       0.71       0.19       0.00
hc             |
  college vs no |      0.56       0.60       0.03       0.51
```

<continued>

```
               |      From         To     Change    p-value
---------------+-----------------------------------------
lwg            |
         Range |      0.17       0.83       0.67       0.00
inc            |
         Range |      0.74       0.09      -0.65       0.00
agecat         |
 40-49 vs 30-39 |      0.70       0.55      -0.15       0.00
   50+ vs 30-39 |      0.70       0.39      -0.31       0.00
   50+ vs 40-49 |      0.55       0.39      -0.16       0.00
```

1. There is little to be learned if the total change is small, such as **hc**. Variables **k5**, **wc**, and **inc** have *potentially* large effects.
2. Option **trim()** can be used to remove extreme values (**help mchange**).

```
* trim upper and lower 5th percentiles
mchange inc, atmeans amount(range) trim(5)
```


#5.4 AME: average marginal effects

```
. mchange // <= no atmeans

logit: Changes in Pr(y) | Number of obs = 753

Expression: Pr(lfp), predict(pr)

                |    Change    p-value
----------------+----------------------
k5              |
             +1 |    -0.281      0.000
            +SD |    -0.153      0.000
       Marginal |    -0.289      0.000
k618            |
             +1 |    -0.014      0.337
            +SD |    -0.018      0.337
       Marginal |    -0.014      0.335
agecat          |
 40-49 vs 30-39 |    -0.124      0.002
   50+ vs 30-39 |    -0.262      0.000
   50+ vs 40-49 |    -0.138      0.002
```



```
wc              |
  college vs no |     0.162      0.000
hc              |
  college vs no |     0.028      0.508
lwg             |
             +1 |     0.120      0.000
            +SD |     0.072      0.000
       Marginal |     0.127      0.000
inc             |
             +1 |    -0.007      0.000
            +SD |    -0.086      0.000
       Marginal |    -0.007      0.000

Average predictions

             | not in LF      in LF
-------------+----------------------
  Pr(y|base) |     0.432      0.568
```

Interpretations (excluding p-values)

1. AME for k5

   On average having one more young child decreases the probability of being in the labor force by .28.

2. MEM for k5

   For someone who is average on all characteristics, having an additional young child is expected to decrease the probability of LFP by .32.

3. AME for wc

   On average women who attend college have a probability of being in the labor force that is .16 greater than those who do not attend college.

4. MEM for wc

   If an average woman attends college, her probability of being in the labor force is .19 greater than that of an average woman who does not attend college.

MEM vs AME

```
mchange, amount(sd) // AME

mchange, amount(sd) atmeans // MEM
```

	AME	MEM	
	Change	Change	AME-MEM
k5 +SD	-0.153	-0.180	0.027
k618 +SD	-0.018	-0.021	0.003
agecat 40vs39	-0.124	-0.146	0.021
50+ vs 30-39	-0.262	-0.307	0.044
50+ vs 40-49	-0.138	-0.161	0.023
wc college vs	0.162	0.186	-0.024
inc +SD	-0.086	-0.101	0.016

Distribution of marginal effects

1. *On average* the marginal effect for the log of wages on the probability of labor force is .127.
2. *On average* if a woman attends college her probability of labor force participation increase by .162.
3. Averages do not indicate variation in the sample.
   - The effect of college might depend on the levels of other variables.
4. This suggests the importance of looking at the distribution of marginal effects.
   - This is not commonly done, but should be.
5. How do you do this?
   - In Stata 13 you can do this with some simple programming as shown in the do-file at #54 for details on how to create these plots
   - Stata 14 added an undocumented feature to do this!

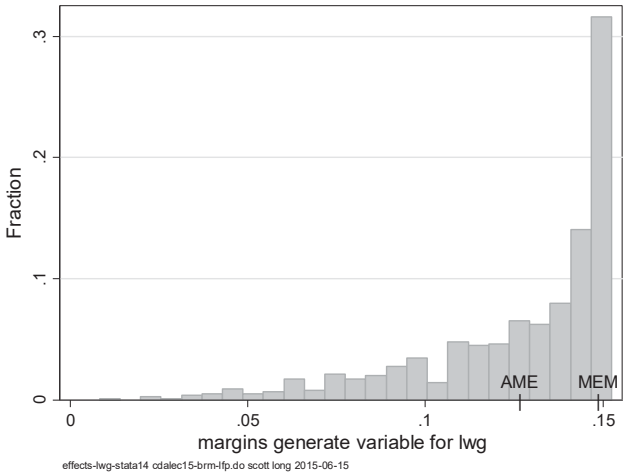
Distribution of effects for lwg

1. For the first case, compute the effect

```
. list k5 k618 agecat wc hc lwg inc in 1, clean

       k5   k618   agecat      wc      hc        lwg       inc
  1.    0      3    30_39   NoCol   NoCol   .8532125   28.363

. mchange lwg, at(k5=0 k618=3 agecat=1 wc=0 hc=0 lwg=.8532125 inc=28.363)
    atmeans

logit: Changes in Pr(lfp) | Number of obs = 753

             |    Change       P>|z|
-------------+----------------------
lwg          |
         +1 |     0.137      0.000
        +SD |     0.084      0.000
    Marginal |     0.148      0.000

Base values of predictors

    |     k5    k618   agecat      wc      hc     lwg      inc
----+-------------------------------------------------------------
 at |      0       3        1       0       0    .853     28.4
```

2. Do this for all cases and plot...

Distribution of marginal change for lwg



#5.5 computing the distribution of effects for wc

Stata 14 distribution of marginal effects

```
margins, dydx(wc) generate(dydxwc)
```

1. The variable **dydxwc** has the DC for each observation
2. Plot using **graph**
3. **help margins undocumented** for details on the **generate()** option

Using Stata 12 and Stata 13

1. Estimate the model.
2. Change **wc** to 0 for *all cases* and compute predictions:

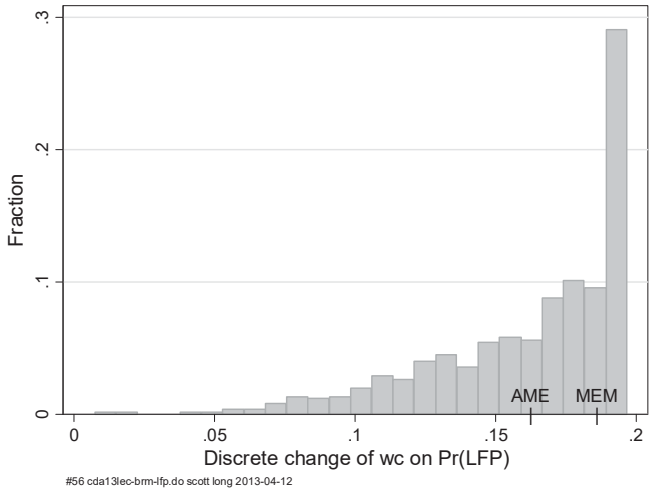
```
gen wc_orig = wc
label var wc_orig "source wc variable"
replace wc = 0
predict pratwc0
label var pratwc0 "PR if wc=0 for all cases"
```

3. Change **wc** to 1 for *all cases* and compute predictions in **pratwc1**.

4. Create **prdif** = **pratwc0** - **pratwc1** and take the mean of **prdif**.

```
. // AME for wc
. gen wc_orig = wc // save original values for wc
. replace wc = 0 // make all cases wc=0
(212 real changes made)
. predict pratwc0 // make predictions
(option pr assumed; Pr(lfp))
label var pratwc0 "PR if wc=0 for all cases"

. replace wc = 1 // make all cases wc=1
(753 real changes made)
. predict effects_phat1 // make predictions
(option pr assumed; Pr(lfp))
label var pratwc1 "PR if wc=1 for all cases"

. replace wc = wc_orig // restore original values
(541 real changes made)
. drop wc_orig

. gen double me_wc = pratwc1 - pratwc0 // DCwc
. label var me_wc "Discrete change of wc on Pr(LFP)"

. * sum me_wc matches result from margins, dydx(wc)
. sum me_wc

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       me_wc |       753    .1624037    .0344572   .0074083    .196826
```


Distribution of DC for wc



Summary of marginal effects

1. A summary measure of the "effect" of a variable is often useful.
2. In the BRM, regression coefficients do not directly indicate the magnitude of the effect.
3. OR's are often used, but are limited as discussed below.
4. In most cases measures of the change in the probability for a change in a regressor are the best way to summarize the effect of a regressor.
5. AME and MEM are often close, but AME is preferred as a single measure in most cases.
6. Multiple MER's might be the best approach.
   - Look at effects at "interesting" locations in the data space.
7. Summary measures are only summaries.
8. Remember, *the model is nonlinear....*

#6 Predictions for ideal types (-brm-lfp.do)

1. Next, think about the substantive issue. What types of people are you interested in? What interesting clusters of characteristics occur together?
2. These sets of characteristics are called *ideal types* or *profiles*
3. Defining profiles forces you to think about where you want to look in the data
4. Comparing predictions across profiles helps you understand your data and the effects of variables
5. We will compute these types and later compare them statistically

```
                           |    Pr(y)        ll        ul
---------------------------+-----------------------------
            Average person |    0.578     0.539     0.616
      Young lower ed kids  |    0.159     0.068     0.251
     Young higher ed kids  |    0.394     0.234     0.554
Middle age higher ed kids  |    0.748     0.672     0.823
          Older higher ed  |    0.631     0.528     0.734
```


An "average person"

```
. estimates restore blm

. mtable, rowname(Average person) atmeans ci clear

Expression: Pr(lfp), predict()

                |    Pr(y)        ll        ul
----------------+-----------------------------
 Average person |    0.578     0.539     0.616

Specified values of covariates

          |                                2.        3.        1.        1.
          |       k5      k618      agecat    agecat        wc        hc
----------+-------------------------------------------------------------------
  Current |     .238      1.35        .385      .219      .282      .392

          |
          |      lwg       inc
----------+--------------------
  Current |      1.1      20.1

. local Average "`r(atspec)'" // see next page
```


A very useful feature for comparing profiles

```
. di "`Average'"
k5=.2377158 k618=1.3532537 1b.agecat=.39575033 2.agecat=.38512616
> 3.agecat=.21912351 0b.wc=.7184595 1.wc=.2815405 0b.hc=.60823373
> 1.hc=.39176627 lwg=1.0971148 inc=20.128965.
```

This is the at-spec we saved earlier that we now use with **`at()`**

```
. mtable, rowname(Average person) at(`Average') ci clear

Expression: Pr(lfp), predict()

                |    Pr(y)        ll        ul
----------------+-----------------------------
 Average person |    0.578     0.539     0.616

Specified values of covariates

          |                                2.        3.        1.        1.
          |       k5      k618      agecat    agecat        wc        hc
----------+-------------------------------------------------------------------
  Current |     .238      1.35        .385      .219      .282      .392

          |
          |      lwg       inc
----------+--------------------
  Current |      1.1      20.1
```

Young, lower class, less educated mom

```
* note: in 1975 $2.10 is min wage; .75 for lwg
mtable, rowname(Young lower ed kids) atmeans ci ///
    at(agecat=1 k5=2 k618=0 inc=10 lwg=.75 hc=0 wc=0) below
local YoungLowEdMom "`r(atspec)'"
```

Young, middle class, more educated mom

1. I define this profile as

   `agecat==1 & k5==2 & k618==0 & wc==1 & hc==1`

2. Where should I hold **`lwg`** and **`inc`**?
   - *Global means* based on the entire sample might be inconsistent with the idea of the profile?
   - *Local means* are based only on individuals who meet this condition

     `if agecat==1 & k5==2 & k618==0 & wc==1 & hc==1`

3. I create a *selection variable* equal to 1 if you have these characteristics:

```
. gen isYoungHiEdMom = agecat==1 & k5==2 & k618==0 & wc==1 & hc==1
. label var isYoungHiEdMom "Select young, higher ed mothers"
```



```
. tab isYoungHiEdMom, miss

     Select |
     young, |
  higher ed |
    mothers |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        747       99.20       99.20
          1 |          6        0.80      100.00
------------+-----------------------------------
      Total |        753      100.00
```

4. Comparing global and local means

```
. sum lwg inc // global

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         lwg |       753    1.097115    .5875564  -2.054124   3.218876
         inc |       753    20.12897     11.6348  -.0290001         96

. sum lwg inc if isYoungHiEdMom // local

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         lwg |         6    1.621039    .3411624   1.230121   2.230264
         inc |         6    16.64083    3.490015     14.245       23.6
```

5. I can use these values with the **`at()`** option – next page.


```
mtable, at(agecat=1 wc=1 hc=1 k5=2 k618=0 lwg=1.62 inc=16.64)
```

6. A simpler way is to use **`if`** and **`atmeans`**

```
mtable if isYoungHiEdMom, rowname(Young higher ed kids) ///
    atmeans ci below
```

Middle aged, educated dad with kids

```
gen isMiddleEdDad = agecat==2 & k5==0 & k618>=1 & wc==1 & hc==1
mtable if isMiddleEdDad, rowname(Middle age higher ed kids) ///
    atmeans ci below
```

Highly educated older couples

```
gen isOlderHiEd = agecat==3 & wc==1 & hc==1 & k618==0
mtable if isOlderHiEd, rowname(Older higher ed) ///
    atmeans ci below twidth(25)
```

Summary of ideal types

```
Expression: Pr(lfp), predict()

                          |    Pr(y)        ll        ul
--------------------------+------------------------------
           Average person |    0.578     0.539     0.616
      Young lower ed kids |    0.159     0.068     0.251
     Young higher ed kids |    0.394     0.234     0.554
Middle age higher ed kids |    0.748     0.672     0.823
          Older higher ed |    0.631     0.528     0.734

Specified values of covariates
<snip>
```

1. Which variables seem important explaining LFP?
2. What is your next step to verify this?
3. Before pursuing this, let's make sure that the differences in the predictions are statistically significant.

Standard errors of predictions

1. How precise are the predictions? Are predictions equal?
2. The *delta method* computes the standard error for the sampling distribution of the estimated predictions
3. Standard errors can be used for confidence intervals and significance tests

Confidence intervals

1. The *confidence interval* (CI) is: [Lower level, Upper level]
2. With a 95% CI, we conclude:

   We are 95% certain that our CI includes the true value of the parameter.

   Or:

   With repeated samples we would expect our prediction to be within the CI 95% of the time.

Examples

1. *The predicted probability of labor force participation is .59 with a 95% confidence interval from .48 to .62.*
2. *The estimated probability of labor force participation is .59 (95%CI: .48, .62).*
3. *Our results suggest that the predicted probability of labor force participation could be as small as .48 or as large as .62 with 95 percent confidence.*

Summary of ideal types

In our commands for ideal types, we could add the option **`statistics(ci)`** or **`ci`** to add confidence intervals to the table.

```
                          |    Pr(y)        ll        ul
--------------------------+------------------------------
           Average person |    0.578     0.539     0.616
      Young lower ed kids |    0.159     0.068     0.251
     Young higher ed kids |    0.394     0.234     0.554
Middle age higher ed kids |    0.748     0.672     0.823
          Older higher ed |    0.631     0.528     0.734
```

4. Are they significantly different?

Comparing profiles/predictions

1. Consider two predictions:

   p1 with CI [p1LB, p1UB] and **p1** with CI [p1LB, p1UB]

2. If the CI's *do not overlap*, the predictions are *significantly different*
3. People *incorrectly* assume: If the CI's overlap, the predictions are not significantly different
4. *To test if predictions differ, I suggest testing if the predictions are equal.*
5. The easiest way to do this is to estimate all of the predictions simultaneously
6. To do this we take advantage of the return **`r(atspec)`**

#6.4 Estimate all predictions simultaneously and post them

```
mtable, at(`Average') at(`YoungLowEdMom') at(`YoungHiEdMom') ///
    at(`MiddleEdDad') at(`OlderHiEd') post atright

Expression: Pr(lfp), predict()

          |                                        2.        3.        1.
          |    Pr(y)        k5      k618    agecat    agecat        wc
----------+-------------------------------------------------------------
        1 |    0.578      .238      1.35      .385      .219      .282
        2 |    0.159         2         0         0         0         0
        3 |    0.394         2         0         0         0         1
        4 |    0.748         0      1.77         1         0         1
        5 |    0.631         0         0         0         1         1

          |       1.
          |       hc       lwg       inc
----------+---------------------------
        1 |     .392       1.1      20.1
        2 |        0        .75        10
        3 |        1      1.62      16.6
        4 |        1      1.32      26.3
        5 |        1      1.38      27.9

Specified values where .n indicates no values specified with at()

          |  No at()
----------+---------
  Current |       .n
```


#6.5 Test if predictions are equal

```
. mlincom 2 - 3 // young lower - young higher

             |   lincom    pvalue        ll        ul
-------------+----------------------------------------
           1 |   -0.235     0.000    -0.340    -0.129
```

A young mother in a low income family without college is significantly less likely to be in the labor force than a similar mother in a college educated family (p<.001).

Tables of predicted probabilities

1. The profiles suggest young children and the wife's education are important
2. This table summarizes the effects of these variables

Number of Young Children	Did Not Attend College	Attended College	Difference	
0	.60	.77	.17	
1	.28	.46	.18	
2	.09	.17	.09	< due to rounding
3	.02	.05	.03	

3. Where do these numbers come from?

Curves behind the table of probabilities

1. Let **Θ** be the linear combination of all variables except **k5** and **wc**
2. The model is

$$\Pr(y=1\,|\,\mathbf{x}) = \Lambda\left(\beta_0 + \beta_{k5}k5 + \beta_{wc}wc + \Theta\right)$$
$$= \Lambda\left(\beta_0^* + \beta_{k5}k5 + \beta_{wc}wc\right)$$

3. If **wc=0**

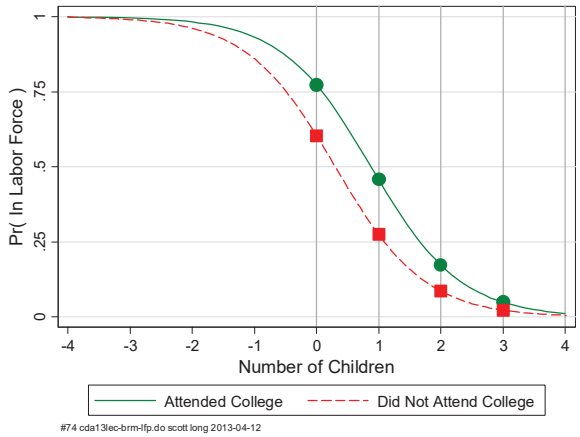
$$\Pr(y=1\,|\,\mathbf{x}, wc=0) = \Lambda\left(\beta_0^* + \beta_{k5}k5\right)$$

4. If **wc=1**

$$\Pr(y=1\,|\,\mathbf{x}, wc=1) = \Lambda\left(\beta_0^* + \beta_{k5}k5 + \beta_{wc}\right)$$
$$= \Lambda\left(\left[\beta_0^* + \beta_{wc}\right] + \beta_{k5}k5\right)$$
$$= \Lambda\left(\beta_0^{**} + \beta_{k5}k5\right)$$

5. These are *parallel curves* as shown on the next page.

# Young Children	Not College	Attended College	Difference
0	.60	.77	.17
1	.28	.46	.18
2	.09	.17	.09
3	.02	.05	.03



#7.1 `mtable` for two variables

```
. mtable, atmeans at(wc=(0 1) k5=(0 1 2 3))

Expression: Pr(lfp), predict()

          |       k5        wc     Pr(y)
----------+------------------------------
        1 |        0         0     0.604
        2 |        0         1     0.772
        3 |        1         0     0.275
        4 |        1         1     0.457
        5 |        2         0     0.086
        6 |        2         1     0.173
        7 |        3         0     0.023
        8 |        3         1     0.049

Specified values of covariates

          |                    2.        3.        1.
          |     k618      agecat    agecat        hc       lwg       inc
----------+-------------------------------------------------------------
  Current |     1.35        .385      .219      .392       1.1      20.1
```


#7.2 `mtable` for a nicer table

1. **`mtable`** stacks predictions from previous **`mtable`** results.
2. **`clear`** means we want a new table, starting from nothing.
3. **`right`** means place new estimates to the right.
4. **`atvars(_none)`** means that no atvars should be added to the table.
5. **`dydx(wc)`** requests a discrete change in **`i.wc`**.

```
1]  . qui mtable, atmeans at(wc=(0) k5=(0 1 2 3)) atvars(k5) ///
    >     clear estname(NoCol)
2]  . qui mtable, atmeans at(wc=(1) k5=(0 1 2 3)) atvars(_none) ///
    >     right estname(College)
3]  . mtable, atmeans dydx(wc) at(k5=(0 1 2 3)) atvars(_none) ///
    >     right estname(Diff) stats(est p)

          |       k5     NoCol   College      Diff         p
----------+-------------------------------------------------
        1 |        0     0.604     0.772     0.168     0.000
        2 |        1     0.275     0.457     0.182     0.001
        3 |        2     0.086     0.173     0.087     0.013
        4 |        3     0.023     0.049     0.027     0.085
```


Local and global means

1. We held other variables at the global means, which might be unrealistic
   - If you have 3 young children, you are unlikely to be in the oldest age group
2. Local means hold other variables at levels "around" or local to other variables being examined held constant
   - For example, the mean age for those with 3 young children
3. Compute predictions with local means using *if* and *atmeans*
   a. Create a selection variable that defines the group of interest.
   b. Tell **`mtable`** to select only these cases.
   c. Use **`atmeans`** to compute means within this group.
4. One of the regressors can be the selection variable or other variables can be used

#7.3 local means for tables using if

1. Select cases where **k5** is 0 and use **atmeans**

```
. mtable if k5==0, estname(k5_0) at(wc=(0 1) k5=0) atvars(1.wc) atmeans ///
>     clear

       1.
      wc      k5 0
-----------------
       0     0.583        <= prediction for k5==0 and wc==0
       1     0.757        <= prediction for k5==0 and wc==1

                            2.         3.         1.
       k5       k618     agecat     agecat         hc         lwg         inc
-------------------------------------------------------------------------------
    0.000      1.279      0.436      0.269      0.358      1.107     19.987
```


2. Adding predictions for **k5=1**

```
. mtable if k5==1, estname(k5_1) at(wc=(0 1) k5=1) atvars(_none) ///
>           atmeans right
```

3. Adding predictions for **k5=2** and **k5=3**

```
. mtable if k5==2, estname(k5_2) at(wc=(0 1) k5=2) atvars(_none) atmeans ///
>     right
. mtable if k5==3, estname(k5_3) at(wc=(0 1) k5=3) atvars(_none) atmeans ///
>     right

       1.
      wc      k5 0      k5 1      k5 2      k5 3
-------------------------------------------------
       0     0.583     0.337     0.154     0.017
       1     0.757     0.530     0.288     0.037
```

4. Compute discrete changes for **wc** for each level of **k5**
   - Does the probability for a given number of children differ by **wc**?

#7.4 DC for wc

1. **dydx(***var***)** tells **margins** and **m*** commands to compute marginal effects for *var*.

2. If *var* is a factor variable, it computes a discrete change; else a marginal change

```
mtable if k5==0, dydx(wc) stat(est p) atmeans clear long ///
    roweqnm(DCwc) coleqnm(k5_0)
mtable if k5==1, dydx(wc) stat(est p) atmeans right long coleqnm(k5_1)
mtable if k5==2, dydx(wc) stat(est p) atmeans right long coleqnm(k5_2)
mtable if k5==3, dydx(wc) stat(est p) atmeans right long coleqnm(k5_3)
```

3. Results in

```
Expression: Pr(lfp), predict()

           | k5_0      k5_1      k5_2      k5_3
           |  d Pr(y)   d Pr(y)   d Pr(y)   d Pr(y)
 ----------+-------------------------------------
 DCwc      |
   d Pr(y) |    0.173     0.193     0.134     0.020
         p |    0.000     0.000     0.003     0.070

Specified values of covariates
<snip>
```

Comparing results for global and local means

	wc=0	wc=1	Change	pvalue
global				
k5=0	0.60	0.77	0.17	0.00
k5=1	0.27	0.46	0.18	0.00
k5=2	0.09	0.17	0.09	0.01
k5=3	0.02	0.05	0.03	0.09
local				
k5=0	0.58	0.76	0.17	0.00
k5=1	0.34	0.53	0.19	0.00
k5=2	0.15	0.29	0.13	0.00
k5=3	0.02	0.04	0.02	0.07

1. Trends are similar
2. Biggest differences are for one and two children
3. Which predictions make the most sense to you?
4. It is important to determine how sensitive results are to the levels of the other variables

Plotting predictions

1. For continuous variables, graphs can be effective
   - To know if a graph is effective, you have to create it
2. **`marginsplot`** is easy, but does not allow you to combine results from multiple models and in Stata 13 does not work with multiple outcomes
3. I prefer **`mgen`** to generate variables to plot with **`graph`**

Tool: graph options

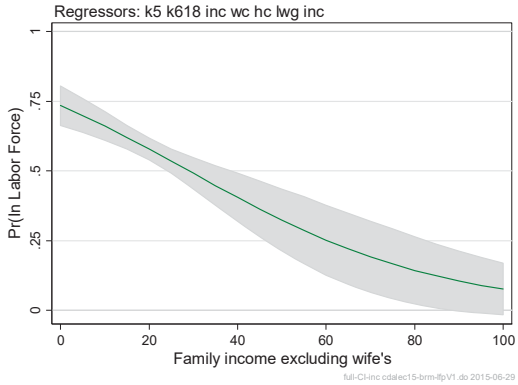
1. Graphs have many options that make the command difficult to read
2. It is hard to keep options consistent across graphs
3. I put key options into locals.
   a. **`local ylab "0(.25)1., grid gmin gmax"`**
   b. Then **`` `ylab' ``** means **`0(.25)1., grid gmin gmax`**

4. Here are the options used for the graphs that follow:
   - They help you create consistent graphs
   - They make it easier to debug graph commands

```
local inc_rng  "0(5)100"
local xlab_inc "0(20)100"
local ylab     "0(.25)1., grid gmin gmax"
local ylabdc   "0(.1).4, grid gmin gmax"
local ytitle   "Pr(In Labor Force)"
local lineprob "msym(i) lcol(green) lpat(solid)"
local linelow  "msym(i) lcol(gold)  lpat(dash)"
local line1    "msym(i) lcol(red)   lpat(dash)"
local line0    "msym(i) lcol(blue)  lpat(solid)"
local line30   "msym(Oh) msiz(*1.1) mcol(red)   lcol(red)"
local line40   "msym(Sh) msiz(*0.9) mcol(green) lcol(green)"
local line50   "msym(Th) msiz(*0.9) mcol(blue)  lcol(blue)"
```

Introduction to lowess plots

1. This is the probability plot for income
2. Is it substantively reasonable?



#8.1 Start analysis with a lowess plot

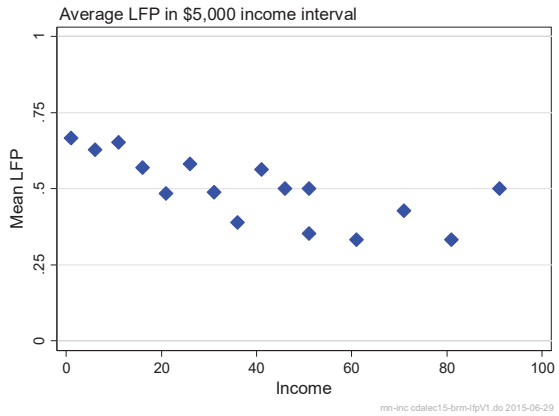
1. Since a *lowess* plot is non-parametric, it does not constrain the shape of the relationship between a regressor and the outcome
2. **A lowess is a valuable first step in evaluating how a regressor is related to the outcome**

Intuition behind a lowess plot

1. Compute mean LFP within income intervals of 5:

```
. sum lfp if inc>=0 & inc<5

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         lfp |        12    .6666667     .492366          0          1

<snip>

. sum lfp if inc>=35 & inc<40

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         lfp |        18    .3888889    .5016313          0          1
```


2. Plotting the means by income



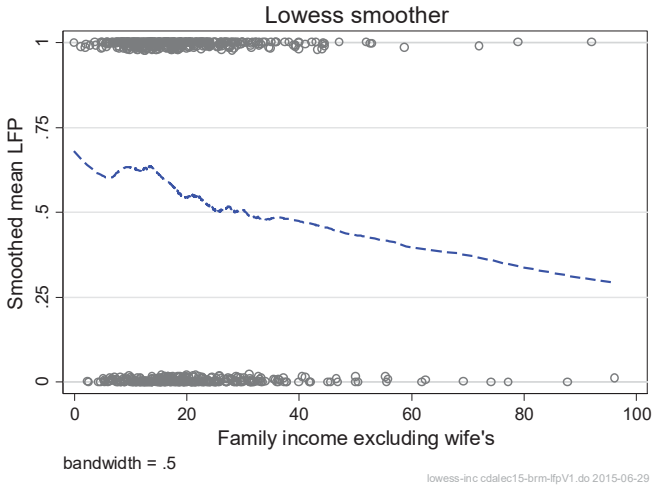
- The "noise" above 50 due to the small N's

3. A lowess plot is a sophisticated way to do this that uses "sliding" intervals.

```
. sort inc
. lowess lfp inc, jitter(3) gen(lowesslfp) xlab(`xlab_inc') ///
      bwidth(.5) ytitle(Smoothed mean LFP) ylab(`ylab') ///
      ylin(0 1, lcol(gs13)) msym(oh) lineopt(lcol(green) ///
      lwid(medthick))
```

4. The **`gen(lowesslfp)`** option saves the predictions to a variable.

Graph on next page...



Lowess and logit curves

1. Sometimes lowess curves do not look like something from a logit model
   a. Since lowess does not control for other variables, the curve might not look "logit-like"
   b. The process might not fit the standard specification of the logit model. Examples of this are given in Part 8.
2. Next we plot the predicted probabilities by income from a logit model.

#8.2 predictions from logit

1. Estimate model with only income

```
. logit lfp inc
  <snip>
. estimates store blminconly
```

2. **mgen** computes predictions at specific values of income and saves predictions in variables to plot
   - **stub(***stub-name***)**: Variables generated begin with *stub-name*.
   - **predname(***pred-name***)**: The default name of prediction is name returned by **mtable**. If you want a different name, use the **predname()**.
   - **predlabel(***label***)**: The label used for the prediction. This is useful for making graphs where the label is used for titles and legends.
   - **at(***var-name*=<range>**)**: Values of *var-name* are saved as <*stub-name*><*var-name*>.

3. Generated variables: <*stub-name*><*stat-name*>

```
    <stat-name>  Description
    ----------------------------------------------------------------
    predname     Estimate such as probabilities, dydx, etc.
    ll           Lower level bound of confidence interval
    ul           Upper level bound of confidence interval
    pval         p-value for test margin is 0
    se           Standard error of margin
    z            z-value of test prediction is 0
```

4. Predict outcome as income increases from 0 to 100 by 5:

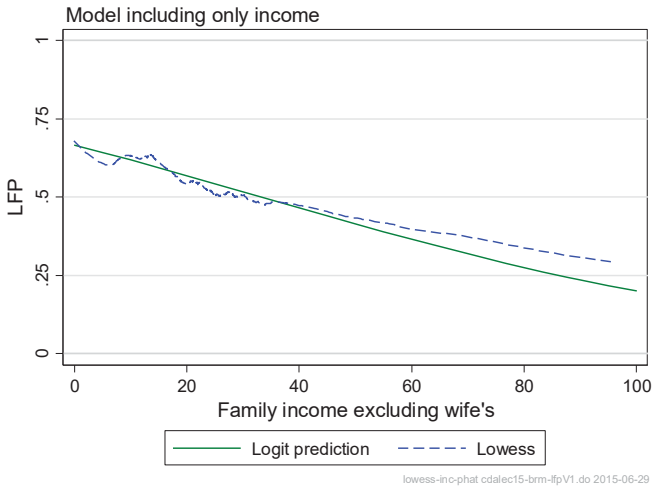
```
. mgen, at(inc=(0(5)100)) atmeans stub(PLT) predlabel(Logit prediction)

Predictions from: margins, at(inc=(0(5)100)) atmeans predict(pr)

Variable   Obs Unique      Mean       Min       Max  Label
------------------------------------------------------------------------------
PLTpr1      21     21  .4223433  .2008354  .6669906  Logit prediction
PLTll1      21     21   .320794  .0336831  .6007513  95% lower limit
PLTul1      21     21  .5238926  .3679877  .7332299  95% upper limit
PLTinc      21     21        50         0       100  Family income excluding .
------------------------------------------------------------------------------

* inc is the only covariate, so there is no list of other covariates
```


5. Variables beginning with **PLT** are created by **mgen**:

```
. format %9.3g lfp inc PLTpr PLTll PLTul PLTinc
. list lfp inc PLTpr PLTll PLTul PLTinc in 1/25, clean nolabel
```

	Observed Variables		mgen variables			
	lfp	inc	PLTpr1	PLTll1	PLTul1	PLTinc
1.	1	-.029	.667	.601	.733	0
2.	1	1.2	.644	.588	.699	5
3.	0	1.5	.619	.573	.666	10
4.	1	2.13	.595	.556	.633	15
5.	1	2.2	.569	.534	.605	20
<snip>						
15.	1	5	.319	.176	.462	70
16.	1	5.12	.297	.146	.448	75
17.	1	5.12	.276	.119	.433	80
18.	1	5.32	.255	.0938	.417	85
19.	0	5.33	.236	.0714	.401	90
20.	1	5.49	.218	.0514	.385	95
21.	0	5.55	.201	.0337	.368	100
22.	0	6	.	.	.	.
23.	0	6	.	.	.	.
24.	1	6.02	.	.	.	.
25.	1	6.25	.	.	.	.

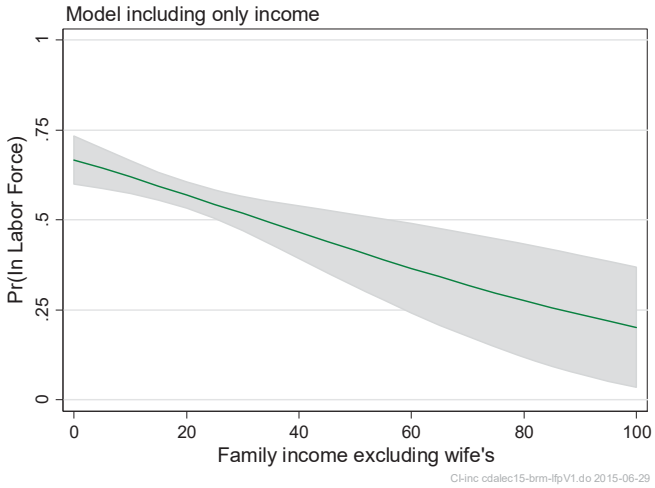
6. Plotting **PLTpr** against **PLTinc** with lowess included



#8.2 adding a CI around the predictions

1. The means **lfp** for $5,000 intervals of income were unstable at higher incomes due to small N's
2. The logit model uses all the observations to fit a curve for all values of income
3. The curve is smooth, but the confidence in the prediction varies
4. The CI reflects the lower certainty for larger values of income

```
. graph twoway ///
>     (rarea PLTul PLTll PLTinc, color(gs13)) /// shaded CI
>     (connected PLTpr PLTinc, `lineprob'), /// line for prob
>     subtitle("Model including only income", position(11)) ///
>     ytitle("`ytitle'") ylab(`ylab') xtitle("`xtitle'") ///
>     legend(off)
```

Plot income in full model

1. Now consider the full model

   logit lfp k5 k618 i.agecat i.wc i.hc lwg inc

2. Compute predictions holding other variables at their means

$$\Pr\left(y=1\,|\,\mathbf{x}^{*}, AGE\right)$$

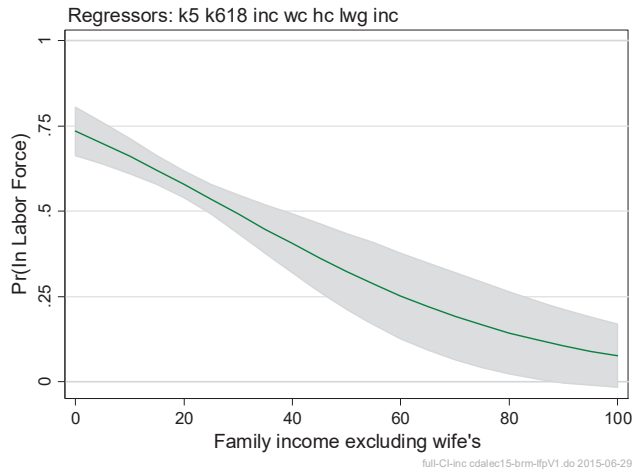
```
. estimates restore blm // full logit model
. mgen, at(inc=(`inc_rng')) atmeans stub(PLT)
Predictions from: margins, at(inc=(0(5)100)) atmeans predict(pr)
```

Variable	Obs	Unique	Mean	Min	Max	Label
PLTpr1	21	21	.3583955	.0768617	.7349035	pr(y=in LF) from margins
PLTll1	21	21	.2680128	-.0156624	.6641427	95% lower limit
PLTul1	21	21	.4487782	.1693859	.8056643	95% upper limit
PLTinc	21	21	50	0	100	Family income excludi...

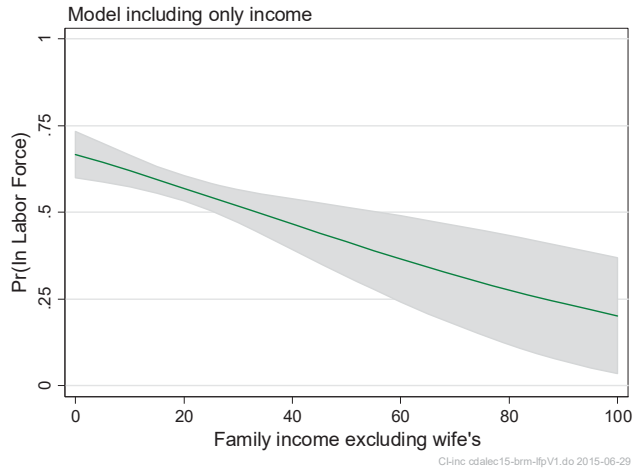
Specified values of covariates

k5	k618	2. agecat	3. agecat	1. wc	1. hc	lwg
.2377158	1.353254	.3851262	.2191235	.2815405	.3917663	1.097115

Full model: discussion follows next graph



Income only model: discussion follows

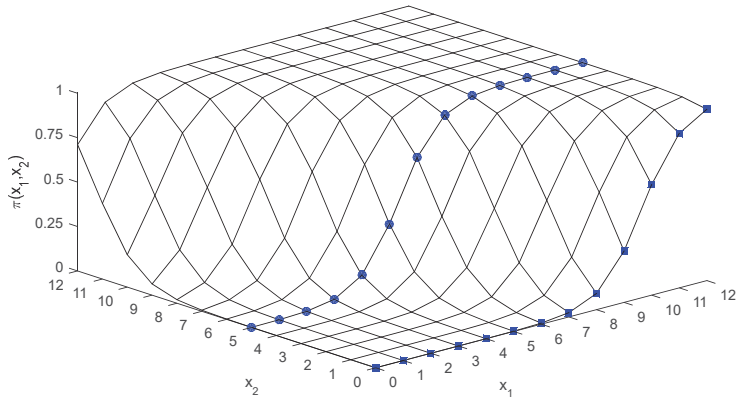


Adding controls affects the curve in three ways

1. The coefficients for **inc** differ; this affects the slope of the curves
   a. Income only: -.0207569
   b. Full model: -.0350542
2. The intercepts differ which move the curves left and right
   a. Income only: .6946054
   b. Full model: 1.013999
3. The levels of the other variables shift the curve left and right, essentially changing the intercept

Recall our earlier graphs...

Visually: How x2 affects the "effect" of x1



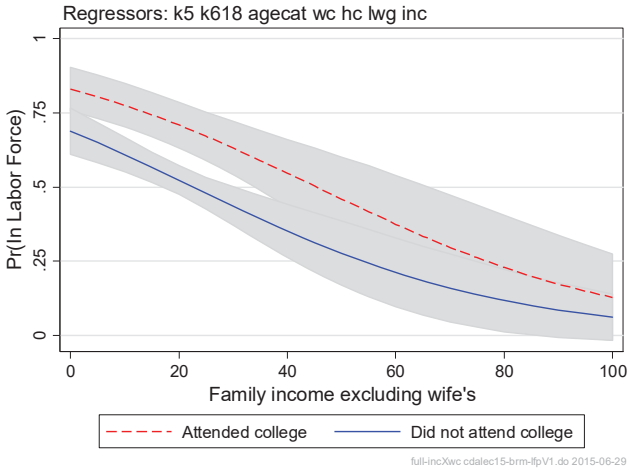
#8.3 predictions for income by wife's college

1. To show interactions, I plot curves at different values of other variables.
   - For example, plot Pr(LFP) by **inc** for each level of **wc**.
2. Let x* be the fixed values for all variable except **age** and **wc**.
3. Compute

$$\Pr\left(y=1 \mid \mathbf{x}^*, WC=0, INC\right) \text{ and } \Pr\left(y=1 \mid \mathbf{x}^*, WC=1, INC\right)$$

```
. mgen, at(inc=(`inc_rng') wc=0) atmeans stub(PLTwc0) ///
>     predlabel(Did not attend college)

Predictions from: margins, at(inc=(0(5)100) wc=0) atmeans predict(pr)

Variable   Obs Unique      Mean        Min       Max  Label
------------------------------------------------------------------------------
PLTwc0pr1   21     21  .3177494   .0623648  .6889161  Did not attend college
PLTwc0ll1   21     21  .2309727  -.0151898  .6107004  95% lower limit
PLTwc0ul1   21     21   .404526   .1399194  .7671317  95% upper limit
PLTwc0inc   21     21        50          0       100  Family income exclud...
------------------------------------------------------------------------------

<snip values of covariates>

. mgen, at(inc=(`inc_rng') wc=1) atmeans stub(PLTwc1) ///
>     predlabel(Attended college)

Predictions from: margins, at(inc=(0(5)100) wc=1) atmeans predict(pr)

Variable   Obs Unique      Mean        Min       Max  Label
------------------------------------------------------------------------------
PLTwc1pr1   21     21  .4684761   .1286839  .8310055  Attended college
PLTwc1ll1   21     21  .3404584  -.0173306  .7575446  95% lower limit
PLTwc1ul1   21     21  .5964938   .2746983  .9044663  95% upper limit
PLTwc1inc   21     21        50          0       100  Family income exclud...
------------------------------------------------------------------------------

<snip values of covariates>
```


Income by wife's education



#8.4 DC(wc|inc): are the curves significantly different

1. Do women who go to college have higher rates of LFP for all levels of income?
2. The figure shows two curves with their CIs.
   a. Red curve: CI [LB Pr(y=1|wc=1, inc); UB Pr(y=1|wc=1, inc)]
   b. Blue curve: CI [LB Pr(y=1|wc=0, inc); UB Pr(y=1|wc=0, inc)]
3. If the CI's *do not overlap*, the predictions are *significantly different*
4. If the CI's *overlap*, significance is *unknown*
5. Here's how to do this for the two curves

Testing differences in predictions

1. We want to test

   H_0: DC(wc|inc) = 0

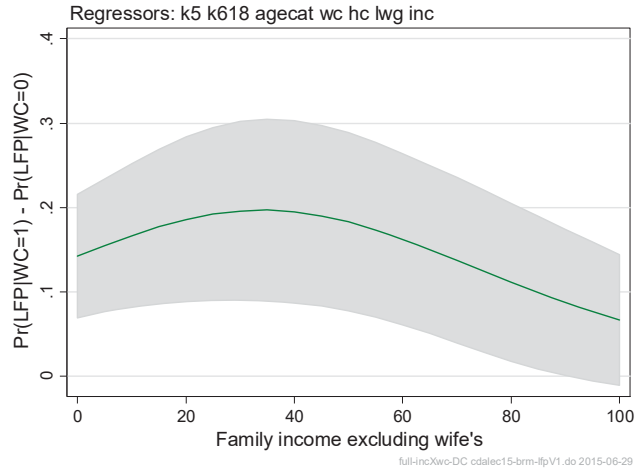
2. We compute

   [Lower bound DC(wc|inc), Upper bound DC(wc|inc)]

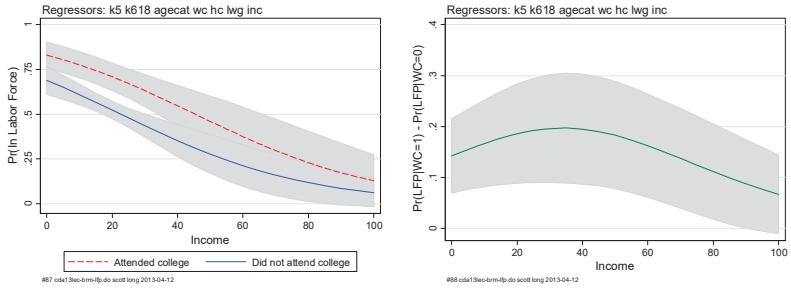
3. Since **wc** is a factor variable **i.wc** , **mgen** computes DC with the **dydx(wc)**

```
. mgen, dydx(wc) at(inc=(`inc_rng')) atmeans stub(PLTdc) ///
>     predlabel(DC of wc by income)

Predictions from: margins, dydx(wc) at(inc=(0(5)100)) atmeans predict(pr)

Variable    Obs Unique      Mean        Min       Max  Label
------------------------------------------------------------------------------
PLTdcd_pr1   21     21  .1507267    .066319  .1967745  DC of wc by income
PLTdcll1     21     21  .0556941  -.0111785  .0895455  95% lower limit
PLTdcul1     21     21  .2457593   .1438166  .3049388  95% upper limit
PLTdcinc     21     21        50          0       100  Family income excl...
------------------------------------------------------------------------------

<snip values of covariates>
```


DC of wc by income



Comparing overlapping CI's to tests of DC:



1. Overlapping CIs do *not* indicted non-significant differences
2. For two curves (left graph), I do not find plotting the CI useful

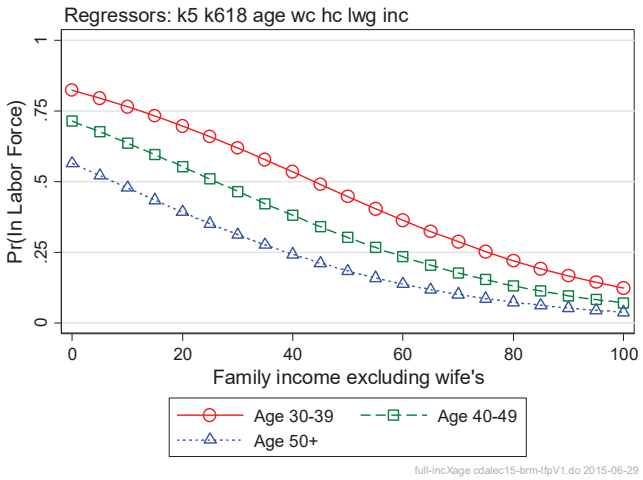
* Predictions by inc for various age levels

1. Here we look at the effect of income by age group:

```
. estimates restore blm
. mgen, at(inc=(`inc_rng') agecat=1) atmeans stub(p30) predlabel(Age 30-39)

Predictions from: margins, at(inc=(0(5)100) agecat=1) atmeans predict(pr)

Variable   Obs Unique      Mean        Min       Max  Label
------------------------------------------------------------------------------
p30pr1      21     21  .4583632   .1230226  .8236541  Age 30-39
p30ll1      21     21  .3331298  -.0230338  .7624032  95% lower limit
p30ul1      21     21  .5835967    .269079  .8849049  95% upper limit
p30inc      21     21        50          0       100  Family income exclud...
------------------------------------------------------------------------------

Specified values of covariates

                                        1.         1.
       k5       k618     agecat         wc         hc        lwg
-------------------------------------------------------------------
 .2377158   1.353254          1   .2815405   .3917663   1.097115

. mgen, at(inc=(`inc_rng') agecat=2) atmeans stub(p40) predlabel(Age 40-49)
  <snip>
. mgen, at(inc=(`inc_rng') agecat=3) atmeans stub(p50) predlabel(Age 50+)
  <snip>
```

We find…

The effect of income on LFP by age category



Graphs for discovery versus presentation

2. *You need a graph to decide if you need a graph!*

1. If a graph is simple, you probably don't need it in a paper
   - you need it to decide if you don't need it
2. You need tools to create graphs quickly and must organize them efficiently or you won't do it
   a. Use templates to speed up the process of making graphs
   b. Use a file viewer to quickly examine graphs

Interpretation with odds ratios (OR)

Odds ratios are a *common* and often *unsatisfactory* method of interpretation

What is an odds ratio?

Probability and odds at x and x+1

Probability: $\Pr(y=1\mid x)$ $\Pr(y=1\mid x+1)$

Odds: $\Omega(x)=\dfrac{\Pr(y=1\mid x)}{\Pr(y=0\mid x)}$ $\Omega(x+1)=\dfrac{\Pr(y=1\mid x+1)}{\Pr(y=0\mid x+1)}$

The OR is the ratio

Odds ratios: $OR(x\rightarrow x+1)=\dfrac{\Omega(x+1)}{\Omega(x)}$

For a unit increase in x, the odds increase by a factor of *OR holding other variables constant*.

To fully understand this, we start with the log odds or logit

Linear in the log of the odds

1. The logit is the log of the odds
2. The logit model is linear in the logit

$$\ln\left[\frac{\Pr(y=1\mid\mathbf{x})}{1-\Pr(y=1\mid\mathbf{x})}\right]=\ln\Omega(\mathbf{x})=\mathbf{x\beta}=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3$$

3. For a unit change in x_k, the logit is expected to change by β_k, holding other variables constant.
4. *Substantively, what does a change of β_k logits mean?*
   - The logit of LFP decreases by 1.30
5. To understand the change in logit, we transform it to odds

Change logit to odds and compute odds ratio (ORs)

1. Take the exponential of the logit with a focus on x_3:

$$\begin{aligned}\Omega(\mathbf{x})&=\exp\left[\ln\Omega(\mathbf{x})\right]=\exp(\mathbf{x\beta})\\&=e^{\beta_0+\beta_1x+\beta_2x_2+\beta_3x_3}\\&=e^{\beta_0}e^{\beta_1x_1}e^{\beta_2x_2}e^{\beta_3x_3}=\Omega(\mathbf{x},x_3)\end{aligned}$$

2. Let x_3 change by 1

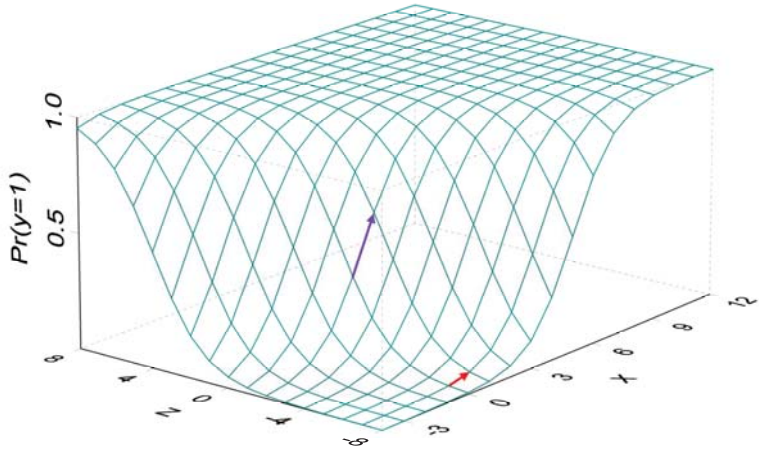
$$\begin{aligned}\Omega(\mathbf{x},x_3+1)&=e^{\beta_0}e^{\beta_1x_1}e^{\beta_2x_2}e^{\beta_3(x_3+1)}\\&=e^{\beta_0}e^{\beta_1x_1}e^{\beta_2x_2}e^{\beta_3x_3}e^{\beta_3}\end{aligned}$$

3. The odds ratio

$$\frac{\text{Ending }\Omega}{\text{Starting }\Omega}=\frac{\Omega(\mathbf{x},x_3+1)}{\Omega(\mathbf{x},x_3)}=\frac{e^{\beta_0}e^{\beta_1x_1}e^{\beta_2x_2}e^{\beta_3x_3}e^{\beta_3}}{e^{\beta_0}e^{\beta_1x_1}e^{\beta_2x_2}e^{\beta_3x_3}}=e^{\beta_3}$$

4. *The OR does not depend on the level of other variables*

A change of 1 in x has the same OR everywhere



#9 Logit estimates

```
. logit lfp k5 k618 i.agecat i.wc i.hc lwg inc

Logistic regression                                     Number of obs   =        753
                                                        LR chi2(8)      =     124.30
                                                        Prob > chi2     =     0.0000
Log likelihood = -452.72367                             Pseudo R2       =     0.1207

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          k5 |  -1.391567   .1919279    -7.25   0.000    -1.767739   -1.015395
        k618 |  -.0656678    .068314    -0.96   0.336    -.1995607    .0682251
             |
      agecat |
           2 |  -.6267601    .208723    -3.00   0.003     -1.03585   -.2176705
           3 |  -1.279078   .2597827    -4.92   0.000    -1.788242   -.7699128
             |
        1.wc |   .7977136   .2291814     3.48   0.001     .3485263    1.246901
        1.hc |   .1358895   .2054464     0.66   0.508     -.266778    .5385569
         lwg |   .6099096   .1507975     4.04   0.000      .314352    .9054672
         inc |  -.0350542   .0082718    -4.24   0.000    -.0512666   -.0188418
       _cons |   1.013999   .2860488     3.54   0.000     .4533539    1.574645
------------------------------------------------------------------------------
```


#9 ORs with `listcoef`: interpretation on next page

```
. listcoef, constant help

logit (N=753): Factor Change in Odds

  Odds of: 1InLF vs 0NotInLF

------------------------------------------------------------------------------
         lfp |      b         z     P>|z|     e^b    e^bStdX      SDofX
-------------+----------------------------------------------------------------
          k5 |  -1.39157   -7.250   0.000   0.2487     0.4823     0.5240
        k618 |  -0.06567   -0.961   0.336   0.9364     0.9170     1.3199
   2.agecat |  -0.62676   -3.003   0.003   0.5343     0.7370     0.4869
   3.agecat |  -1.27908   -4.924   0.000   0.2783     0.5889     0.4139
        1.wc |   0.79771    3.481   0.001   2.2205     1.4319     0.4500
        1.hc |   0.13589    0.661   0.508   1.1456     1.0686     0.4885
         lwg |   0.60991    4.045   0.000   1.8403     1.4310     0.5876
         inc |  -0.03505   -4.238   0.000   0.9656     0.6651    11.6348
       _cons |   1.01400    3.545   0.000
------------------------------------------------------------------------------
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
     e^b = exp(b) = factor change in odds for unit increase in X
 e^bStdX = exp(b*SD of X) = change in odds for SD increase in X
```

Odds ratio: factor change in the odds

1. For a *unit change* in x_k the odds are expected to change by a factor of $\exp(\beta_k)$, holding other variables constant.
   a. For $\exp(\beta_k)>1$, the odds are $\exp(\beta_k)$ times larger.

      By attending college her odds of LFP are 2.22 times larger, holding other variables constant.

   b. For $\exp(\beta_k)<1$, the odds are $\exp(\beta_k)$ times smaller.

      For an additional young child, the odds of LFP are .25 times smaller, ...

2. For a *standard deviation change* in x_k, the odds are expected to change by a factor of $\exp(s_k\beta_k)$, holding other variables constant.

   For a standard deviation increase in the log of wages the odds of LFP are 1.43 times larger, ...

Percentage change in the odds

1. If the odds change by a *factor of 2*, they are *100% larger*.
2. If the odds change by a *factor of .5*, they are *50% smaller*.
3. In general, **%change = 100*(OR-1)**.

   100% = 100*(2-1) Double odds, is 100% increase

   -50% = 100*(.5-1) Halve odds, is 50% decrease

4. For example
   a. By attending college her odds of LFP are 124 percent larger, holding other variables constant.
   b. For an additional young child, the odds of LFP are 77 percent smaller, ...
   c. For a standard deviation increase in the log of wages the odds of LFP are 43 percent larger, ...
5. To compute these: `listcoef, percent`

Interpreting odds ratios (ORs)

1. OR is a *multiplicative* coefficient
   a. Positive effects are greater than one
   b. Negative effects are between zero and one
2. Magnitudes of positive and negative ORs are compared by taking the inverse of the negative effect (or vice versa)
   a. A positive OR=2 has the same magnitude as a "negative" OR=1/2.
   b. An OR=1/10 is larger than OR=2.
3. The effect on the odds of the event *not* occurring is the inverse of the OR of the event occurring

   Being ten years older makes the odds of not being in the labor force 1.9 (=1/.52) times greater, holding other variables constant.

OR compared to Pr(y) for groups

1. Two logit models are estimated

```
logit tenure pub phdyr if female==1
logit tenure pub phdyr if female==0
```

where $\exp\left(\hat{\beta}_{pub}^{Women}\right) = \exp\left(\hat{\beta}_{pub}^{Men}\right) = 2.$

2. Suppose the base probabilities and odds

   $p_M = .500$ → $\Omega_M = .500/(1-.500) = 1.000$

   $p_W = .050$ → $\Omega_W = .050/(1-.050) = 0.053$

3. How does doubling the odds change the probability?

   $2*\Omega_M = 2.000$ → $p_M = 2.000/(2.000+1) = .667$

   $2*\Omega_W = 0.105$ → $p_W = 0.105/(0.105+1) = .095$

4. Then,

   Δ pM/ Δ pub = .167 = (.667 - .500)

   Δ pW/ Δ pub = .045 = (.095 - .050)

Overview of binary LHS

Why so much time on BRM

1. The BRM is the foundation for many models for ordinal, nominal, and count variables
2. A deep understanding of BRM makes models easier to understand

Key points

1. Interpretation requires understanding nonlinearity
2. No single method of interpretation is always best
   - Try alternative methods to find which one works best.
3. There are subtle ways in which the BRM differs from the LRM that will be explored as the class progresses
   - Be careful about taking what you know about LRM and appying it to BRM

Part 4: Hypothesis testing

Read and run

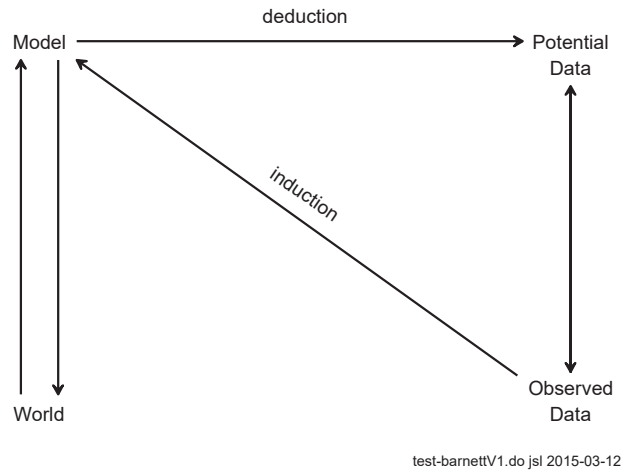
Long & Freese	Chapter 3.2
cdalec*.do	cdalec17-test-lfp.do

Overview

Hypothesis testing is critical for the effective use of regression models

1. Review of the theory of hypothesis testing
2. Testing a single coefficients
3. Simultaneously testing multiple coefficients

Barnett's model of inference



Test of a single coefficients

1. If H_0: $\beta_k = \beta_k^*$ is true, the ML estimator is

$$\hat{\beta}_k \overset{a}{\sim} \text{Normal}\left(\beta_k^*, \text{Var}\left(\hat{\beta}_k\right)\right)$$

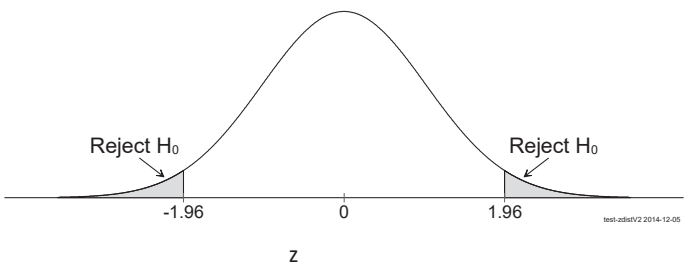
2. Two types of errors are possible when testing H_0: β=0

	Decision	
H_0: β=0	**Accept H_0**	**Reject H_0**
In fact β=0	No error	**Type I: Pr(reject true)=α** Size of test (the shaded tail).
In fact β≠0	**Type II: accept false** Power of test.	No error

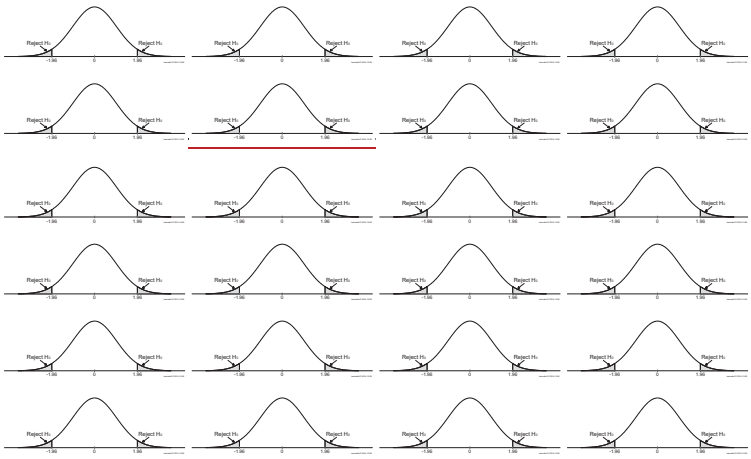
3. Consider testing H_0: $\beta_k = 0$ using $\hat{\beta}_k$ with $\hat{\sigma}_{\hat{\beta}_k}$:

$$z = \frac{\hat{\beta}_k - 0}{\hat{\sigma}_{\hat{\beta}_k}}$$

If H_0 is true, then the sampling distribution is



Twenty tests when β=0



#11 *z*-test of β with logit (-test-lfp.do)

```
. logit lfp k5 k618 i.agecat i.wc i.hc lwg inc, nolog

Logistic regression                                     Number of obs   =        753
                                                        LR chi2(8)      =     124.30
                                                        Prob > chi2     =     0.0000
Log likelihood = -452.72367                             Pseudo R2       =     0.1207

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          k5 |  -1.391567   .1919279    -7.25   0.000    -1.767739   -1.015395
        k618 |  -.0656678    .068314    -0.96   0.336    -.1995607    .0682251
             |
      agecat |
           2 |  -.6267601    .208723    -3.00   0.003     -1.03585   -.2176705
           3 |  -1.279078   .2597827    -4.92   0.000    -1.788242   -.7699128
             |
          1.wc |   .7977136   .2291814     3.48   0.001     .3485263    1.246901
          1.hc |   .1358895   .2054464     0.66   0.508     -.266778    .5385569
           lwg |   .6099096   .1507975     4.04   0.000      .314352    .9054672
           inc |  -.0350542   .0082718    -4.24   0.000    -.0512666   -.0188418
         _cons |   1.013999   .2860488     3.54   0.000     .4533539    1.574645
------------------------------------------------------------------------------
```


Interpreting z-test for a single coefficient

```
------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          k5 |  -1.391567   .1919279    -7.25   0.000    -1.767739   -1.015395
        k618 |  -.0656678    .068314    -0.96   0.336    -.1995607    .0682251
------------------------------------------------------------------------------
```

1. Having young children has a significant effect on the probability of working (z=-7.25, p<0.01 for a two-tailed test).
2. The effect of having young children is significant (p<.01).
3. The effect of having older children is not significant (z=-.96, p=.34).

Note: Unless it is clear from the context in which the result is presented, you should indicate if it is a one-tailed or two-tailed test.

Hypothesis for multiple coefficients

1. Consider the model

```
logit lfp k5 k618 i.agecat i.wc i.hc lwg inc
```

2. What if we wanted to test
   - Kids have no impact on LFP
   - Education has no impact
3. We cannot do this with the z-values from **`logit`**
4. Consider *algebraic* statements and *probabilistic* statements

Algebraic relationships among hypothesis

1. These hypotheses are algebraic statements

   H_A: $\beta_X = 0$ <= income has no effect

   H_B: $\beta_Z = 0$ <= wealth has no effect

   H_C: $\beta_X = \beta_Z$ <= income & wealth have equal effects

   H_D: $\beta_X = \beta_Z = 0$ <= income & wealth have no effects

2. *If* H_A and H_B are *true*, then H_C and H_D *must* be true
   - $\beta_X = 0$ & $\beta_Z = 0$ algebraically imply $\beta_X = \beta_Z = 0$

Conclusions from hypothesis tests

1. Consider *conclusions* from tests of four hypotheses

   H_A: $\beta_X = 0$ ➔ evidence this *might* be true

   H_B: $\beta_Z = 0$ ➔ evidence this *might* be true

   H_C: $\beta_X = \beta_Z$ ➔ ?

   H_D: $\beta_X = \beta_Z = 0$ ➔ ?

2. Accepting H_A and H_B does not imply you will accept H_C or H_D!
   - Who stole my wallet?
3. More formally, consider the formula from the LRM

$$y = \beta_0 + \beta_x x + \beta_z z + \varepsilon$$

$$Var\left(\hat{\beta}_X\right) = \frac{\sigma_\varepsilon^2}{N\sigma_X^2\left(1-\rho_{XZ}^2\right)}$$

Wald tests of joint hypotheses

1. ML theory shows that:

$$\hat{\boldsymbol{\beta}} \overset{a}{\sim} \text{Normal}\left(\boldsymbol{\beta}, \mathbf{Var}(\hat{\boldsymbol{\beta}})\right)$$

2. With three coefficients:

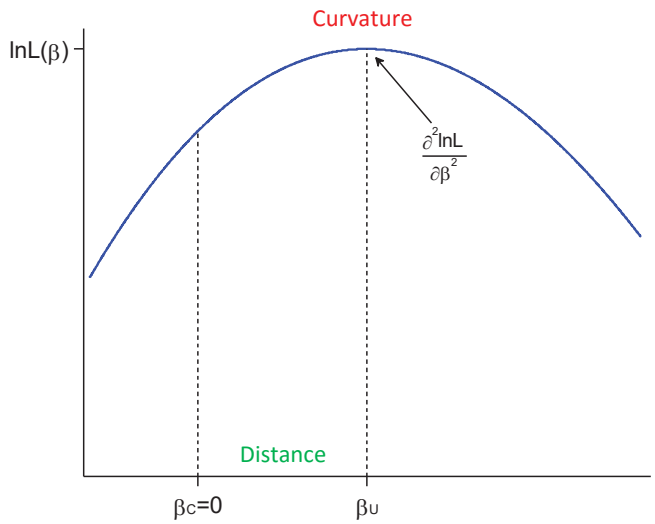
$$\text{Var}\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_X\\ \hat{\beta}_Z\end{pmatrix} = \begin{pmatrix}\sigma^2_{\hat{\beta}_0} & \sigma_{\hat{\beta}_0,\hat{\beta}_X} & \sigma_{\hat{\beta}_0,\hat{\beta}_Z}\\ \sigma_{\hat{\beta}_X,\hat{\beta}_0} & \sigma^2_{\hat{\beta}_X} & \sigma_{\hat{\beta}_X,\hat{\beta}_Z}\\ \sigma_{\hat{\beta}_Z,\hat{\beta}_0} & \sigma_{\hat{\beta}_Z,\hat{\beta}_X} & \sigma^2_{\hat{\beta}_Z}\end{pmatrix}$$

3. $\sigma_{\hat{\beta}_X,\hat{\beta}_Z}$ tell you how the "regression plane rocks".

4. The Wald test measures:
   a. How far estimates are from hypothesized values.
   b. How flat the likelihood functions is.

Graphically...

Wald test and the log likelihood function



Wald test overview

1. The Wald test estimates model *without* constraints
2. Hypothesis H_0: β=0 imposes a constraint on the coefficient
3. The Wald test evaluates
   a. *Distance* from the unconstrained estimate to the constraint
   b. *Curvature* of ln *L* at the constraint as indicated by $\frac{\partial^2 \ln L}{\partial \beta^2}$
4. The flatter the curve, the less significance

Shown on next page...

Wald test of linear constraints

1. Consider linear constraints $\mathbf{Q\beta = 0}$
   a. $\boldsymbol{\beta}$ is vector of parameters
   b. $\mathbf{Q}$ is matrix that combine the β's
2. Examples:
   a. $\mathbf{Q\beta} = \beta_1 - \beta_2 = 0$
   b. $\mathbf{Q\beta} = \beta_1 = 0$
   c. $\mathbf{Q\beta} = \beta_1 = \beta_2 = 0$
3. The Wald statistic equals

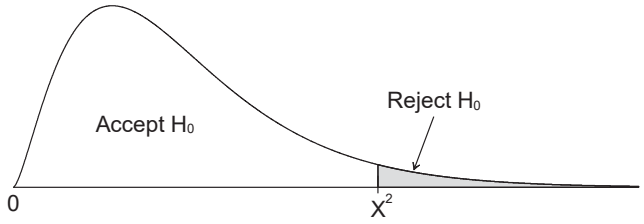
$$W = \left[\mathbf{Q}\hat{\boldsymbol{\beta}} - \mathbf{0}\right]' \left[\mathbf{Q}Var\left(\hat{\boldsymbol{\beta}}\right)\mathbf{Q}'\right]^{-1} \left[\mathbf{Q}\hat{\boldsymbol{\beta}} - \mathbf{0}\right] \sim \chi^2$$

[Distance] [Curvature] [Distance]

4. See Long 1997 for details

Sampling distribution of the Wald test

If H_0 is true, as N increases the sampling distributions of *W* converges to the chi-square distribution with *degrees of freedom* equal to the *number of constraints* being tested



Examples follow...

Wald tests using `test`

The model is:

```
logit lfp k5 k618 i.agecat i.wc i.hc lwg inc
estimates store blm
```

#13 H_0: $\beta_{k5} = 0$

```
. test k5

 ( 1)  [lfp]k5 = 0

           chi2(  1) =   52.57
         Prob > chi2 =    0.0000
```

The effect of having young children on entering the labor force is significant at the .01 level (X_1^2=52.6).

Note

Chi-square 52.57 equals the z-value squared -7.25*-7.25.

How do you know the names of coefficients?

```
. logit, coeflegend

Logistic regression                                     Number of obs   =        753
                                                        LR chi2(8)      =     124.30
                                                        Prob > chi2     =     0.0000
Log likelihood = -452.72367                             Pseudo R2       =     0.1207

------------------------------------------------------------------------------
         lfp |      Coef.  Legend
-------------+----------------------------------------------------------------
          k5 |  -1.391567  _b[k5]
        k618 |  -.0656678  _b[k618]
             |
      agecat |
      40-49  |  -.6267601  _b[2.agecat]
        50+  |  -1.279078  _b[3.agecat]
             |
          wc |
    college  |   .7977136  _b[1.wc]
             |
          hc |
    college  |   .1358895  _b[1.hc]
         lwg |   .6099096  _b[lwg]
         inc |  -.0350542  _b[inc]
       _cons |   1.013999  _b[_cons]
------------------------------------------------------------------------------
```


#14 H_0: $\beta_{wc} = \beta_{hc} = 0$

```
. test 1.wc 1.hc // not test wc hc

 ( 1)  [lfp]1.wc = 0
 ( 2)  [lfp]1.hc = 0

           chi2(  2) =   17.83
         Prob > chi2 =    0.000
```

We can reject the hypothesis that the effects of the husband's and the wife's education are simultaneously zero (X_2^2=17.83, p<.01).

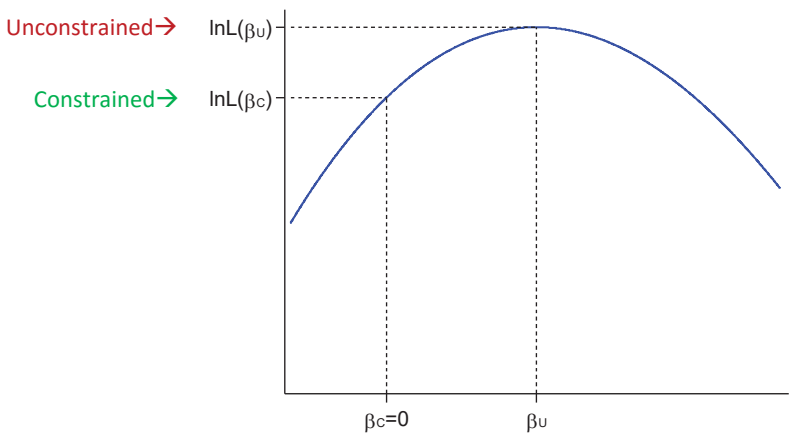
#15 H_0: $\beta_{wc} = \beta_{hc}$

```
. test 1.wc = 1.hc

 ( 1)  [lfp]1.wc - [lfp]1.hc = 0

           chi2(  1) =    3.24
         Prob > chi2 =    0.0719
```

The hypothesis that the effects of husband's and wife's education are equal is rejected marginally at the .05 level (X_1^2=3.24, p=.07).

LR test of *nested* models

The LR test is an alternative to the Wald test.



lr test-wald-lr-lmV2.do 2015-06-10

Nested models

1. What is a constrained model?

   Constrained model = Unconstrained model + constraints.

2. Let M_C be the constrained model.

3. Let M_U be the unconstrained model.

4. M_C is **nested** in M_U.

5. Consider these models:

   M1: $\Pr(y=1|\mathbf{x}) = \Lambda(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$

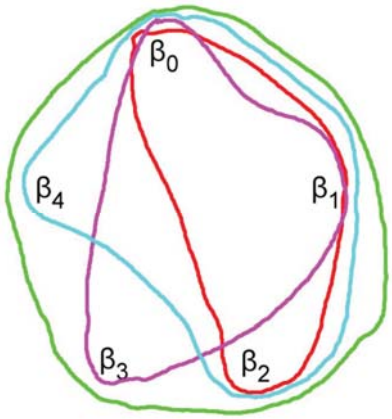
   M2: $\Pr(y=1|\mathbf{x}) = \Lambda(\beta_0 + \beta_1 x_1 + \beta_3 x_3)$

   M3: $\Pr(y=1|\mathbf{x}) = \Lambda(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_4 x_4)$

   M4: $\Pr(y=1|\mathbf{x}) = \Lambda(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4)$

6. We can show how these are nested...

M1: $\beta_0, \beta_1, \beta_2$ **M2**: $\beta_0, \beta_1, \beta_3$ **M3**: $\beta_0, \beta_1, \beta_2, \beta_4$ **M4**: $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$



Computation of LR

1. M_C with L_C=L(M_C) is nested in M_U with L_U=L(M_U).

2. $\boldsymbol{\beta}_C$ is created from $\boldsymbol{\beta}_U$ by *imposing constraints*.

3. We test the hypothesis:

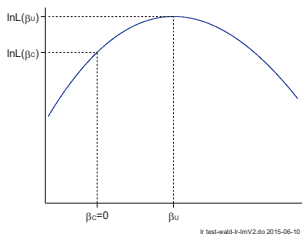
   H_0: The constraints imposed on $\boldsymbol{\beta}_U$ are true.

4. The *LR test statistic* equals:

   $$G^2(M_C \mid M_U) = 2\ln L_U - 2\ln L_C$$

5. Under general conditions, if H_0 is true:

   $$G^2 \sim \chi^2(\text{number of constraints})$$



#21-23 H_0: $\beta_{k5} = 0$

The full model

```
. logit lfp k5 k618 i.agecat i.wc i.hc lwg inc

Iteration 0:   log likelihood =  -514.8732
<snip>
Iteration 4:   log likelihood = -452.72367

Logistic regression                               Number of obs   =        753
                                                  LR chi2(8)      =     124.30
                                                  Prob > chi2     =     0.0000
Log likelihood = -452.72367                       Pseudo R2       =     0.1207
<snip>

. estimates store full
```

Restricted model

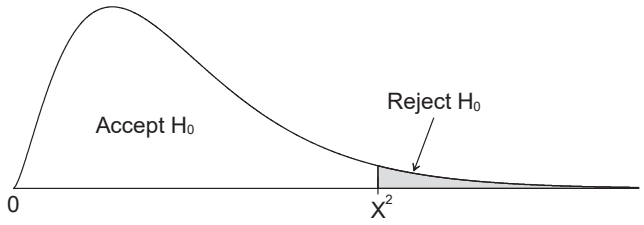
```
. logit lfp    k618 i.agecat i.wc i.hc lwg inc, nolog
  <snip>
. estimates store dropk5
```


LR test

```
. lrtest full dropk5

Likelihood-ratio test                             LR chi2(1)  =     62.55
(Assumption: dropk5 nested in full)               Prob > chi2 =    0.0000
```

$$G^2\left(M_{[K5]} \mid M_U\right) = G^2\left(M_U\right) - G^2\left(M_{[K5]}\right) = 62.55$$

Having young children is significant at the .01 level (LRX^2(1)=62.55).



#29 Comparing LR and Wald results

1. LR and Wald are asymptotically equivalent, differ in finite samples

		Wald		LR		
	df	chi2	pval	chi2	pval	Ratio
k5_0	1.000	52.569	0.000	62.554	0.000	0.840
wc_hc_0	2.000	17.832	0.000	18.684	0.000	0.954
wc_hc	1.000	3.239	0.072	3.264	0.071	0.992
agecat_0	2.000	24.273	0.000	25.417	0.000	0.955
all	8.000	95.897	0.000	124.299	0.000	0.772

2. Which test is used is often determined by convenience and convention
3. LR test requires estimation of two models, but subtraction is easy
4. Wald test requires estimation of single model, but the computation requires matrix manipulations
5. Statisticians generally prefer the LR test

Practical issues when computing the LR test

1. The *same sample* must be the same for all model
   a. Since ML excludes cases with missing data, the sample size can change when the variables in a model change
   b. Construct a data set that excludes every observation that has missing values for any of the variables used in any of the models being tested
   c. Use **keep** or **drop** to select observations that have no missing values for variables in the model
2. Dropping categories of factor variable
   a. Start with

```
logit y female 2.edcat 3.edcat 4.edcat
```

   b. You *cannot* drop **2.edcat** with

```
logit y female          3.edcat 4.edcat
```

   c. How should you do this?

Summary on testing

1. LR and Wald tests can be used with other models using MLE
2. Testing multiple coefficients is often critical for your work
3. Sometimes researchers use only the default tests from the estimation command
   - They test things they aren't interested in
   - They don't test things they are interested in
4. Never "add" the results of two or more tests!

Part 5: Complex sampling

Primary source: Heeringa, S., West, B.T., & Berglund, P.A. (2010). *Applied survey data analysis*. Boca Raton, FL: Chapman Hall/CRC. (HWB) **Read and run**

Long & Freese	Pages 100-103; help svy and read introduction to SVY manual
cdalec*.do	cdalec17-svy-hrs.do

Overview

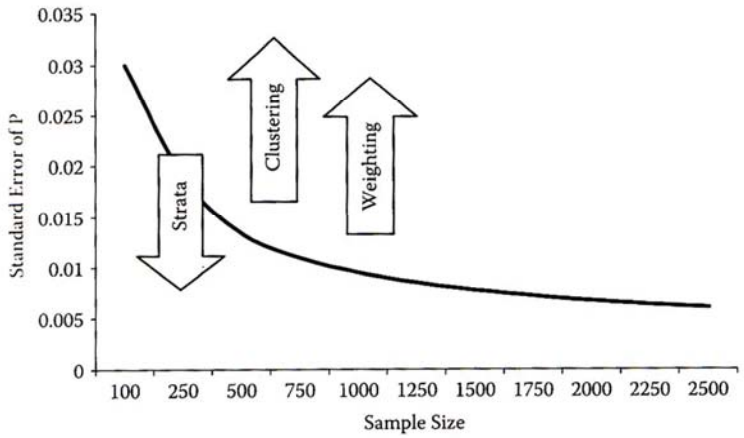
1. Standard software assumes data come from a *simple random sample* (**SRS**)
   a. Each person in the population has the same probability of selection
   b. The probability of one person being selected does not affect the probability of another person being selected.
   c. It is conceptually simple, but impractical.

2. Most major datasets use a *complex sampling designs*
    a. *Clustering*: clusters are sampled; all cases in cluster are included
    b. *Stratification*: strata are chosen, not sampled; sampling occurs within strata
    c. *Sampling weights*: different cases represent different proportions of the population
3. Complex sampling can
    a. Reduce costs
    b. Reduce or increase sampling variability
    c. Increase the representation of subpopulations
4. If you do not adjust for complex sampling
    a. Variances of estimates are usually underestimated
    b. Estimates might be biased
5. I review key concepts and **svy** syntax

Complex sampling designs (HWB 2010)

1. We want to make the standard error (SE) of our estimate small
2. Four aspects of sampling affect the SE
    a. N
    b. Clustering
    c. Stratification
    d. Weights
3. This graph from Heeringa et al. summarizes how these affect the SE

Sample and the size of the Standard Error



1. Each sampling complication changes the "effective N" in the sample (HWB 34)

Design	Estimator	$\bar{y}$	$se(\bar{y})$	Effective n
SRS	$\bar{y}_{SRS}$	40.77	2.41	32.0
Clustered	$\bar{y}_{CL}$	40.77	3.66	13.9
Stratified	$\bar{y}_{ST}$	40.77	2.04	44.4
Stratified, clustered	$\bar{y}_{CL,ST}$	40.77	2.76	24.4

Practical steps for using complex samples

See HWB (10, 13, 115)

1. *Review documentation*. Check the web site for best practices
2. *Identify the variables for survey adjustment*.
   - This can be difficult
3. *Plot survey weights against variables of interest.* Variability in the weights can affect sampling variability of descriptive statistics
4. Create an *analysis dataset* with analysis variables and survey design variables
   - Review the descriptive statistics
5. *Examine the documentation to understand nonresponse issues*
   - Check the web site for information on handling missing data
   - Contact the data producer if you have questions

Using Stata for survey data

1. Stata's **svy** commands provide design-based estimates for complex sampling
2. There are *many* subtle points involving the survey commands. Here I provide only an overview. For details *Stata Survey Data*
3. Using **svy** commands involves two steps
   a. **svyset** to describe the design
   b. **svy:** for commands such as **svy: logit**
   c. Interpretation is largely unchanged from non-svy analysis

HRS: Health and Retirement Study (-svy-hrs.do)

1. The University of Michigan Health and Retirement Study with 22K Americans over 50 every two years. Large, longitudinal study of LFP and health transitions later in life
2. My example examines

```
arthritis  1=arthritis 0=no arthritis
```

3. Regressors

```
female     Is female?
age        Age at 2006 interview
ed11less   Ed years <= 11?
ed12       Ed years = 12?
ed1315     Ed years 13-15?
ed16plus   Ed years 16 or more?
```

4. The variables the describe the complex sample are:

```
secu     sampling error computation unit
kwgtr    2006 weight: respondent level
stratum  stratum id
```

5. *In practice it can be hard to be sure which variables to use*

#2 Declaring the survey design

1. The design is specified

```
. svyset secu             /// clusters
>     [pweight=kwgtr],    /// weights
>     strata(stratum)     /// stratum
>     vce(linearized) singleunit(missing) // method of compute
SE's

      pweight: kwgtr
          VCE: linearized
  Single unit: missing
     Strata 1: stratum
         SU 1: secu
        FPC 1: <zero>
```

2. The output means:

`vce(linearized)` : linearization for estimating standard errors.

`singleunit(missing)` : stratum with single sampling unit is missing.

#6-9 Modeling arthritis

```
// #6 logits without survey

logit arthritis age i.female i.ed4cat
estimates store nosvy
predict nosvyphat
label var nosvyphat "nosvy phat"

// #7 non-svy with weights and cluster

logit arthritis age i.female i.ed4cat ///
    [pweight=kwgtr], cluster(secu)
estimates store wtclstr
predict wtclstrphat
label var wtclstrphat "wtclstr phat"

// #8 logits with full survey adjustments

svyset secu [pweight=kwgtr], ///
    strata(stratum) vce(linearized) singleunit(missing)
svy: logit arthritis age i.female i.ed4cat
estimates store svy
predict svyphat
label var svyphat "svy phat"
```

```
. // #9 tables of estimated coefficients

. estimates table nosvy wtclstr svy, ///
>     b(%9.3f) t(%9.2f) stats(N) eform

-------------------------------------------------
    Variable |   nosvy      wtclstr       svy
-------------+-----------------------------------
         age |     1.046       1.049       1.049
             |     29.57      910.60       21.92
      female |
           1 |     1.759       1.779       1.779
             |     17.68       12.10       12.99
             |
    ed11less |     1.162       1.206       1.206
             |      3.50        2.57        3.16
      ed1315 |     0.961       0.937       0.937
             |     -0.92       -0.94       -1.21
    ed16plus |     0.703       0.638       0.638
             |     -8.20      -11.47       -8.54
        _cons |     0.054       0.046       0.046
             |    -26.60     -226.92      -19.54
-------------+-----------------------------------
           N |     18341       16862       18375
-------------------------------------------------
                                    legend: b/t
```


Sources for complex sample

Heeringa, S., West, B. T., & Berglund, P. A. (2010). Applied survey data analysis. Boca Raton, FL: Chapman & Hall/CRC. [HWB]

Korn, E. L., & Graubard, B. I. (1999). Analysis of health surveys. New York: Wiley. [KG]

Levy, P. S., & Lemeshow, S. (1999). Sampling of populations : methods and applications (3rd ed.). New York: Wiley. [LL]

StataCorp Stata Survey Data Reference Manual. StataCorp LP: College Station, TX. [Stata]

* Part 6: Internal fit 2017-03-08

Read and run

Long & Freese Pages 206-218

cdalec*.do cdalec17-fitinternal-lfp.do

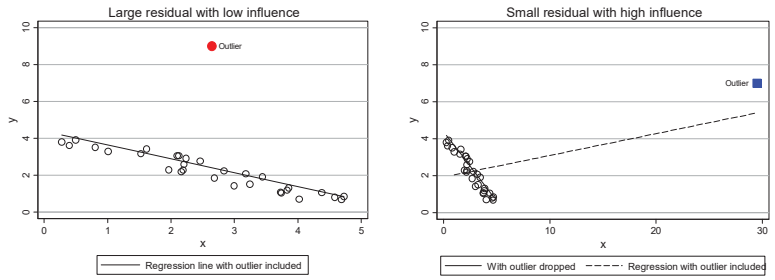
Overview

1. Internal fit consider how individual observations fit the overall model
2. This involves three related concepts
   a. *Residual*: distance between model predictions and the observed values
   b. *Outlier*: observation that is far from predicted value
   c. *Influential observation*: observation that strongly affects estimated coefficients
3. These concepts are shown in these two figures:

Graph on next page...

Panel A: Outlier that is not an influential observation

Panel B: Influential observation that is not an outlier



Residuals for binary outcomes

1. For a binary model

$$\pi_i = \Pr(y_i = 1 \mid \mathbf{x}_i)$$

2. Deviations $y_i - \pi_i$ are *heteroscedastic*

$$Var(y_i - \pi_i \mid \mathbf{x}_i) = \pi_i(1-\pi_i)$$

3. *Pearson residuals* adjust deviations for heteroscedasticity

$$r_i = \frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}}$$

4. *Standardized Pearson residuals* correct for estimation of the variance in the Pearson residual; results are usually similar

$$r_i^{Std} = \frac{r_i}{\sqrt{Var(r_i)}} = \frac{r_i}{\sqrt{1-\left(\hat{\pi}_i(1-\hat{\pi}_i)\mathbf{x}_i Var(\hat{\boldsymbol{\beta}})\mathbf{x}_i'\right)}}$$

#10-12 Index plots for residuals (-fitinternal-lfp.do)

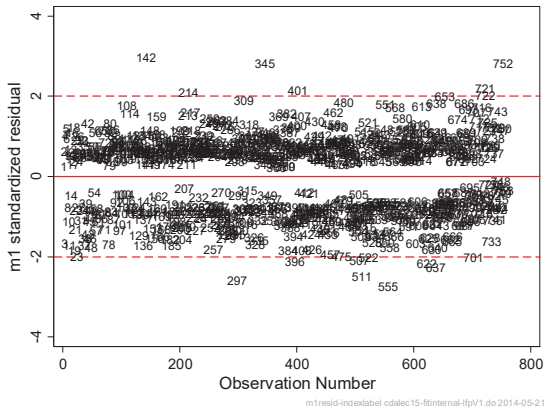
1. *Index plots* show residuals against the case or index number
2. Create a variable with the index number

```
. use binlfp4, clear
. sort inc, stable // keep covariate sets in same relative order
. generate index = _n
. label var index "Observation Number"
```

3. Estimate model and compute residuals and influence

```
. // #11 estimate model
. logit lfp k5 k618 i.agecat i.wc i.hc lwg inc, nolog
  <snip>
. estimates store M1logit

. // #12 standardized residuals
. predict m1resid, rs
. label var m1resid "m1 standardized residual"
```

#14 Index plot of residuals with index numbers

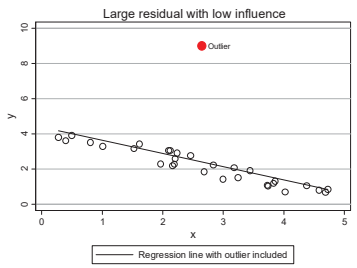


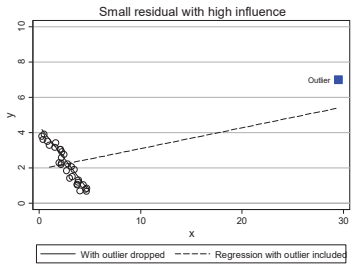

```
graph twoway scatter m1resid index,
   msym(none) mlab(index) mlabpos(0) ...
```


Influence for binary outcomes

Panel A: Outlier that is not an influential observation.

Panel B: Influential observation that is not an outlier.





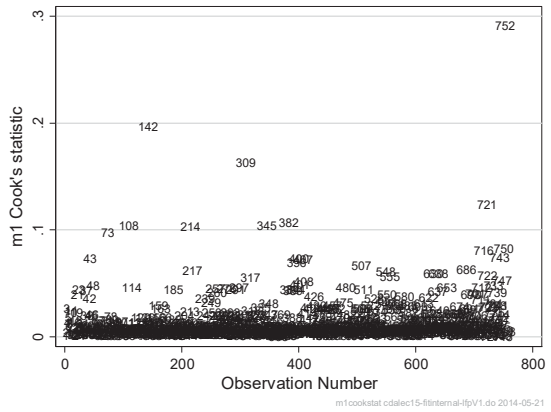
Overview follows…

1. *Large residuals* indicate that an observation does not fit well
2. *Influence* reflects the impact an observation has on the $\hat{\beta}$s
3. Large residuals *in the middle* do not have a large influence
   - Think of a see-saw
4. Extreme observations can influence estimates without being outliers
5. To determine influential or *high-leverage* observations, compute change in $\hat{\boldsymbol{\beta}}$ when dropping each observation
6. Computing influence requires estimating N logits. Pregibon's approximation uses a single estimate of the model.
7. *Cook's distance* summarizes the effects of removing each observation
8. There is *no critical value* for significantly influential observations

#18 computing Cook's statistic

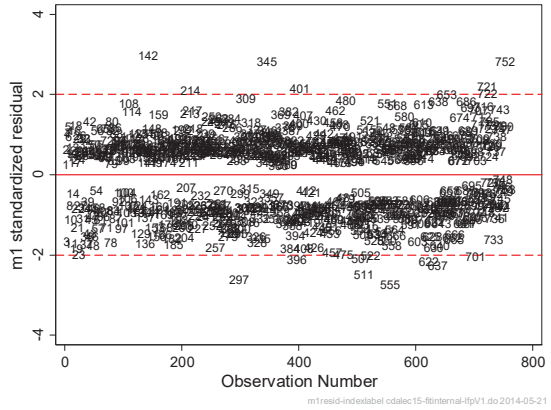
```
predict   M1cookstat, dbeta
label var M1cookstat "Cook's statistic"
```

#18 index plot of influential observations (resid next page)



Compared to the residuals...

Residuals for BRM



Examining outliers, residuals, and influence

1. I find this most useful for finding errors in the data
   - This is an *essential* step in your work
2. Outliers, residuals, and influential cases *could* suggests how the model should be modified
   - I have rarely seen this happen

Part 7: Scalar measures of fit

Read and run

Long & Freese Chapter 3.3

cdalec*.do cdalec17-internalfit-lfp.do

Overview

I consider two types of scalar numbers that characterize a model

1. *Information criteria* such as BIC and AIC are valuable for model selection
2. Briefly, *Pseudo R^2s*

Information criteria

1. Two information criteria are commonly used to select models
   - **AIC**: Akaike's information criterion
   - **BIC**: Bayesian information criterion
2. These measures quantify the tradeoff between
   a. *Fit* of model to data.
   b. *Complexity* of model
   c. More complex models fit better at the cost of more parameters
3. IC are computed as

   IC = **- Fit** + **Complexity**

   = - 2ln*L* + Function of N and # of parameters
   - o Fit is negative; more negative is a better fit
   - o Complexity is positive so more positive is worse fit
4. *Model with the smaller IC is preferred*

Computing IC measures

1. Define
   N = number of observations
   k = number of parameters
   lnL = log likelihood
2. Then

   IC = **fit** + **complexity**

   AIC = **-2ln*L*** + **2*k** // smaller penalty

   BIC = **-2 ln*L*** + **ln(N)*k** // larger penalty
3. *BIC prefers more parsimonious models than AIC*

Comparing models

1. Estimate multiple models
2. Select model with the smallest IC

Example

1. Consider models M1 and M2
   a. ΔBIC = BIC1 - BIC2
   b. If ΔBIC > 0 choose M2 (BIC1 > BIC2)
   c. If ΔBIC < 0 choose M1 (BIC1 < BIC2)
2. While BIC is *not* a statistical test, Raftery suggests degrees of evidence

Absolute ΔBIC	*Strength of Evidence*
0 - 2	Weak
2 - 6	Positive
6 - 10	Strong
>10	Very strong

Software variations

1. BIC in Stata

$$\text{BIC} = \left[-2\ln(likelihood)\right] + \left[\ln N * k\right]$$

where k is the number of parameters

2. BIC'

$$\text{BIC}' = \left[-G^2(M)\right] + \left[df_k' \ln N\right]$$

G^2=LR chi-squared and df_k' =# of regressors (not parameters)

3. BIC deviance or BIC in Raftery's notation

$$\text{BIC}^D = [D] - [df \ln N]$$

Deviance D with *df* = N - (# of parameters):

4. Critically,

$$\text{BIC}_1 - \text{BIC}_2 = \text{BIC}_1' - \text{BIC}_2' = \text{BIC}_1^D - \text{BIC}_2^D$$

Comparing models with IC (-fitexternal-lfp.do)

#10 adding inc-squared and dropping k618 & hc

```
. sysuse binlfp4, clear
. logit lfp k5 k618 i.agecat i.wc i.hc lwg inc, nolog
  <snip>
. estimates store m1

. estat ic

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC          BIC
-------------+---------------------------------------------------------------
          m1 |    753   -514.8732   -452.7237      9     923.4473     965.0639
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note

. qui fitstat, ic save

. logit lfp k5      i.agecat i.wc      lwg c.inc##c.inc, nolog
  <snip>
. estimates store m2

. estat ic
  <snip>
```

```
. estimates table m1 m2, stats(N bic) b(%9.3f) t(%6.2f)

    Variable |    m1          m2
-------------+----------------------
          k5 |    -1.392      -1.385
             |     -7.25       -7.27
        k618 |    -0.066
             |     -0.96
   agecat  2 |    -0.627      -0.585
             |     -3.00       -2.87
           3 |    -1.279      -1.186
             |     -4.92       -5.08
          wc |     0.798       0.904
             |      3.48        4.36
          hc |     0.136
             |      0.66
         lwg |     0.610       0.631
             |      4.04        4.19
         inc |    -0.035      -0.065
             |     -4.24       -3.47
 c.inc#c.inc |                  0.000
             |                   1.88
-------------+----------------------
           N |       753         753
         bic |   965.064     956.484
-------------------------------------
                           legend: b/t
```


`fitstat` *for IC measures*

1. SPost **`fitstat`** command compares BIC and AIC statistics

```
. logit lfp k5 k618 i.agecat i.wc i.hc lwg inc, nolog
  <snip>
. quietly fitstat, ic save

. logit lfp k5      i.agecat i.wc      lwg c.inc##c.inc, nolog
  <snip>

. fitstat, ic diff

                           |     Current        Saved   Difference
-------------------------+-------------------------------------
AIC                        |
                       AIC |     919.491      923.447       -3.956
          (divided by N) |       1.221        1.226       -0.005
-------------------------+-------------------------------------
BIC                        |
         BIC (df=8/9/-1) |     956.484      965.064       -8.580
 BIC (based on deviance) |   -4031.438    -4022.857       -8.580
    BIC' (based on LRX2) |     -79.887      -71.307       -8.580

Difference of    8.580 in BIC provides strong support for current model.
```

2. There is strong support for the model that adds income-squared and drops k618 and hc

* Pseudo R^2's

1. It would be *great* to have a single number to summarize model fit
2. Such a measure would aid in comparing competing models
   - Within a substantive area, measures of fit might provide a rough index of whether a model is adequate
   - If prior models of LFP routinely have values of .4 for a given measure, you expect analyses with a different sample or with revised measures of the variables to have a similar value for that measure.
3. Long (1997) warns

   I am unaware of convincing evidence that selecting a model that maximizes the value of a given measure of fit results in a model that is *optimal in any sense other than the model having a larger value of that measure*.
4. Still, these measures are commonly used in the literature and you should use the measure that is commonly used in your field. But, do not over-interpret it!

#21 Pseudo R^2's in BLM

```
. logit lfp k5 k618 i.agecat i.wc i.hc lwg inc, nolog

Logistic regression                                     Number of obs   =        753
                                                        LR chi2(8)      =     124.30
                                                        Prob > chi2     =     0.0000
Log likelihood = -452.72367                             Pseudo R2       =     0.1207
<snip>
. estimates store m1
. qui fitstat, save

. logit lfp k5      i.agecat i.wc      lwg c.inc##c.inc, nolog
<snip>
. estimates store m2

. fitstat, diff
```



```
. fitstat, diff
                         |     Current        Saved   Difference
-------------------------+-----------------------------------
Log-likelihood           |
                   Model |    -451.746     -452.724        0.978
            Intercept-only |    -514.873     -514.873        0.000
-------------------------+-----------------------------------
Chi-square               |
         D (df=745/744/1) |     903.491      905.447       -1.956
           LR (df=7/8/-1) |     126.255      124.299        1.956
                 p-value |       0.000        0.000        1.000
-------------------------+-----------------------------------
R2                       |
               McFadden |       0.123        0.121        0.002
     McFadden (adjusted) |       0.107        0.103        0.004
      McKelvey & Zavoina |       0.216        0.215        0.001
            Cox-Snell/ML |       0.154        0.152        0.002
  Cragg-Uhler/Nagelkerke |       0.207        0.204        0.003
                   Efron |       0.156        0.153        0.003
               Tjur's D |       0.156        0.153        0.003
                  Count |       0.684        0.676        0.008
         Count (adjusted) |       0.268        0.249        0.018
-------------------------+-----------------------------------
IC                       |
                    AIC |     919.491      923.447       -3.956
         AIC divided by N |       1.221        1.226       -0.005
          BIC (df=8/9/-1) |     956.484      965.064       -8.580
:::
```



```
-------------------------+-----------------------------------
Variance of              |
                      e |       3.290        3.290        0.000
                 y-star |       4.195        4.192        0.003

Note: Likelihood-ratio test assumes current model nested in saved model.

Difference of    8.580 in BIC provides strong support for current model.
```

Overview of fit

1. IC measures can be valuable for selecting models that are not nested
   - Do not over use it
   - Think about your model
2. Scalar measures of fit might be required by referees, but are often of little value

* Part 8: Nonlinearities on the RHS

Read and run

Long & Freese Pages 301-302

cdalec*.do cdalec*-brmnonlin-hrs.do

- The do file is complex with commands for three outcomes
- Each outcome has these sections

```
#1 labels, retrieve mean
#2 lowess
#3 logit with only age
#5 M1: age with controls
#6 M2: age + age^2
#7 M3: age + age^2 + age^3
#8 table of estimated coefficients
#9 Models without survey estimation to compare B
#10 svy: dotplot of predictions
#12 predictions by age for women with HS degree
#13 predictions with CI for each of the 3 models
#16 gender differences at 65
```


Overview

1. Assume **xβ** does not have power terms
2. As x_k increases, either the probability:
   - Always increases with x_k approaching 1.0
   - Always decreases with x_k approaching 0.0
3. Substantively,
   - Should it only increase or only decreases?
   - Should the maximum be 1? The minimum 0?

Adding nonlinearities to a nonlinear model

1. Consider model where x is age with other controls

$$\Pr(y=1\,|\,\mathbf{x}) = \Lambda\left(\beta_0 + \beta_1 x + \beta_2 x^2 + \cdots\right)$$

2. x and x^2 are *linked* since you when x changes x^2 must change

   If x=1, then x^2=1

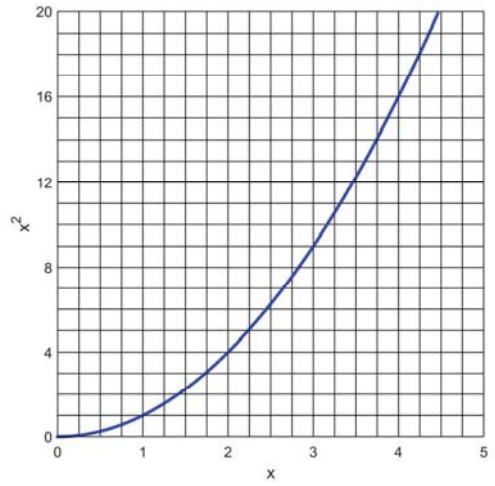
   If x=2, then x^2=4

   If x=3, then x^2=9

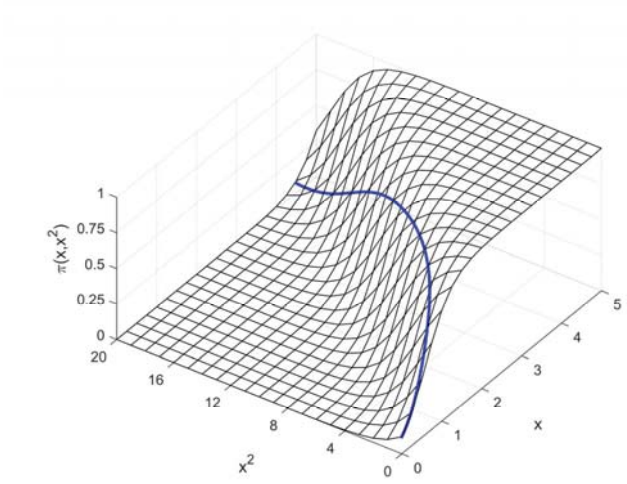
3. With polynomials on the RHS, you are not limited to probabilities that uniformly increase or uniformly decrease with x.

4. To see this, consider this graph.

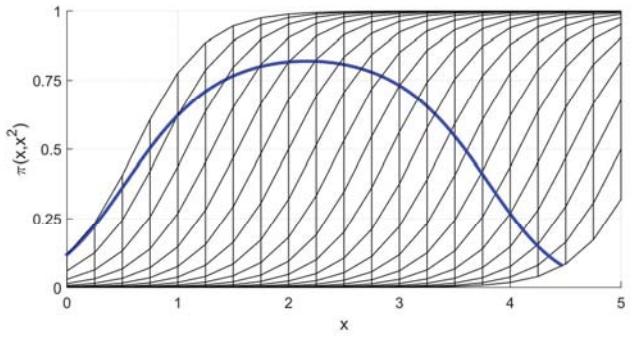
Top view of logit with x and x^2



Front view of logit with x and x^2



Side view of logit with x and x^2



Lowess for assessing nonlinearities

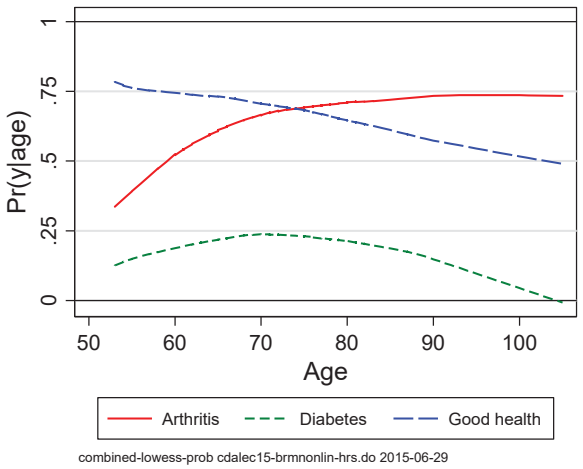
1. *Lowess plots* show a moving average of y as x changes
2. When a regressor is potentially important, a lowess is an essential first step
3. Outcomes from the *HRS: The Health and Retirement Stud*.

```
arthritis  1=arthritis 0=no arthritis
diabetes   1=has diabetes 0=no diabetes
goodhlth   Is health good?
```

4. Age has a *qualitatively different effect* on each outcome

Graph follows...

Lowess graphs for health outcomes (-brmnonlin-hrs.do)



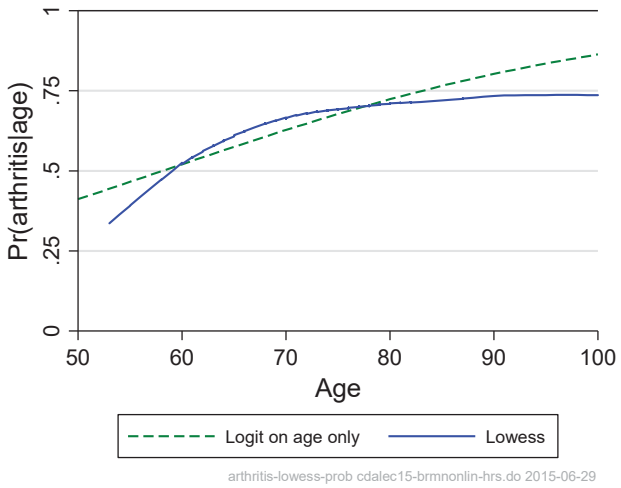
Summary of lowess curves

1. **Arthritis** increases then levels off at .75, not at 1.0.
2. **Diabetes** increases till 65 and then decreases.
3. **Good health** steadily decreases, consistent with a logit on age.

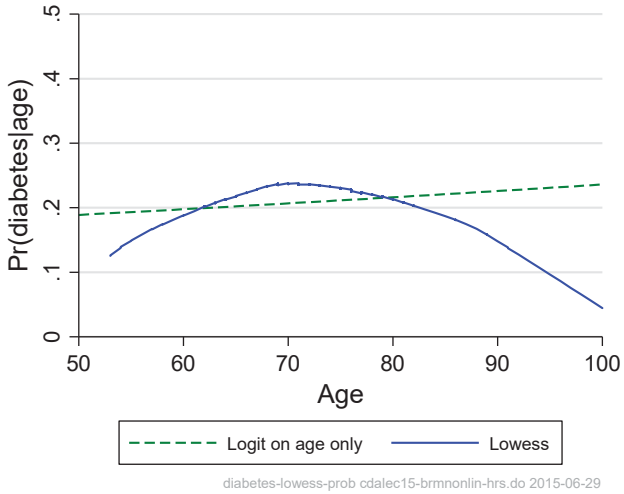
Could these data be generated by a BLM?

1. The relationship between age and the probability could be "logit-like" if we add controls
2. To explore this, I start with a logit on age as the only regressor
3. Then models are estimated that add controls and powers of age

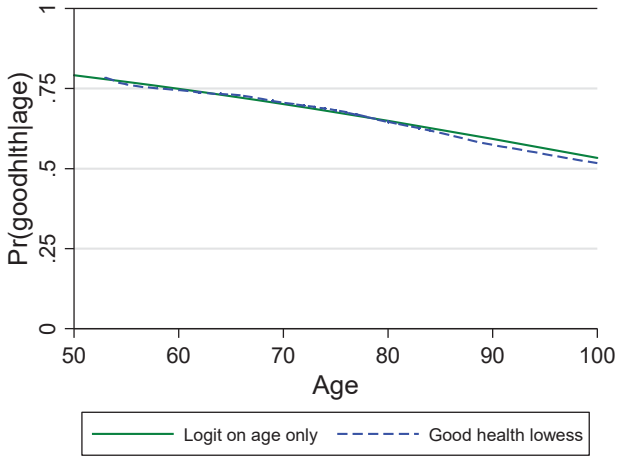
Arthritis: Lowess and BLM on age



Diabetes: Lowess and BLM on age



Good health: Lowess and BLM on age



goodhlth-lowess-prob cdalec15-brmnonlin-hrs.do 2015-06-29

Logit models for arthritis

1. Models are estimated with age, age-squared and age-cubed plus controls
2. Predicted probabilities at observed values are plotted
3. Tests of the effect of age are made
4. IC measures are computed
5. Predictions against age are made with **mgen**

Estimation, test, and predict

```
//  #A5 aM1: age

 1> svy: logit arthritis age i.female i.ed4cat
 2> estimates store aM1
 3> predict aM1pred
 4> label var aM1pred "M1: age"
 5> test age

//  #A6 M2: age + age^2

 6> svy: logit arthritis c.age##c.age i.female i.ed4cat
 7> estimates store aM2
 8> predict aM2pred
 9> label var aM2pred "M2: +age-squared"
10> test age c.age#c.age

//  #A7 aM3: age + age^2 + age^3

11> svy: logit arthritis c.age c.age#c.age c.age#c.age#c.age ///
12>     i.female i.ed4cat
13> estimates store aM3
14> predict aM3pred
15> label var aM3pred "M3: +age-cubed"
16> test age c.age#c.age c.age#c.age#c.age
```

arthritis: estimates

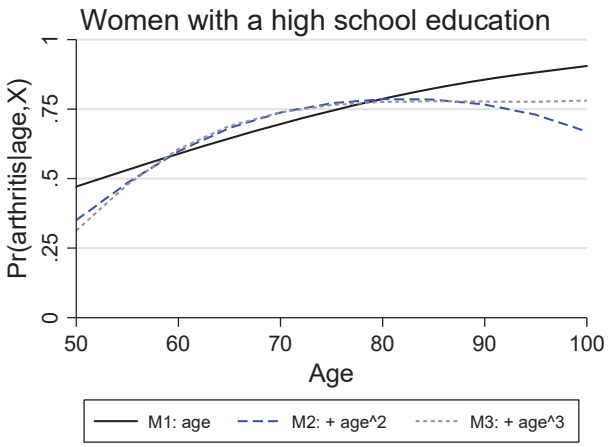
Variable	aM1	aM2	aM3
female			
female	1.77543	1.80948	1.81087
	12.97	13.08	13.10
ed4cat			
12 years	0.82788	0.82101	0.82109
	-3.12	-3.27	-3.27
13-15 years	0.77455	0.79218	0.79310
	-3.82	-3.46	-3.44
16+ years	0.52825	0.53507	0.53543
	-9.64	-9.69	-9.67
age	1.04844	1.35998	2.28835
	21.51	12.06	3.17
c.age#c.age		0.99813	0.99076
		-10.54	-2.55
c.age#c.age#			
c.age			1.00003
			2.07
_cons	0.05711	0.00001	0.00000
	-16.39	-12.98	-3.86

legend: b/t

Predictions with mgen

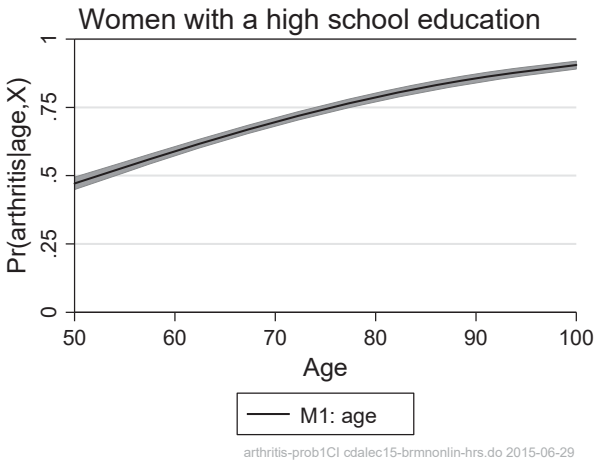
```
//  #A11 predictions for women in high school

estimates restore aM1
mgen, at(age=(50(5)100) female=1 ed4cat=2) atmeans stub(aM1)

estimates restore aM2
mgen, at(age=(50(5)100) female=1 ed4cat=2) atmeans stub(aM2)

estimates restore aM3
mgen, at(age=(50(5)100) female=1 ed4cat=2) atmeans stub(aM3)
```


Arthritis: Predictions for women with high school degrees



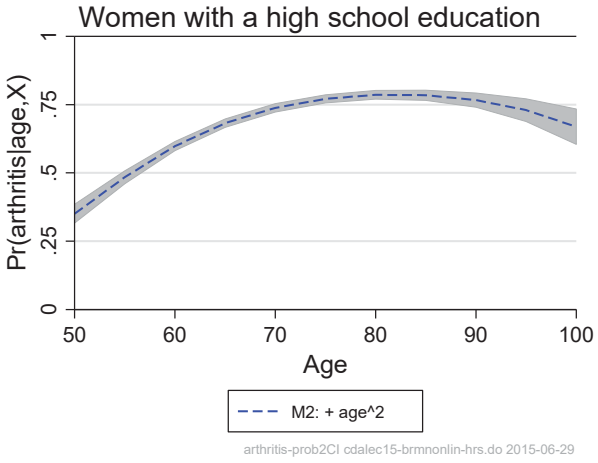
arthritis-prob123 cdalec15-brmnonlin-hrs.do 2015-06-29

Arthritis: M1 with CI



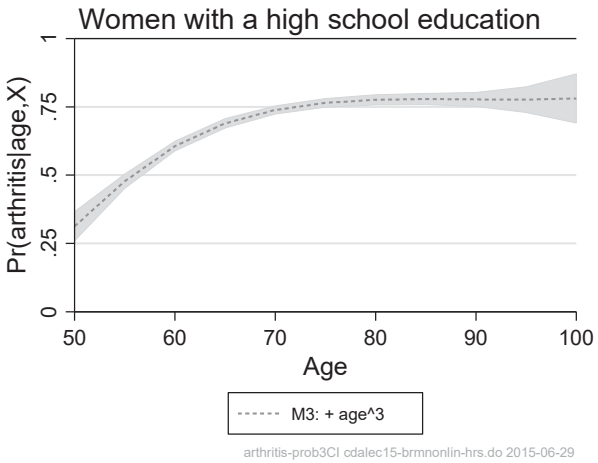
arthritis-prob1CI cdalec15-brmnonlin-hrs.do 2015-06-29

Arthritis: M2 with CI



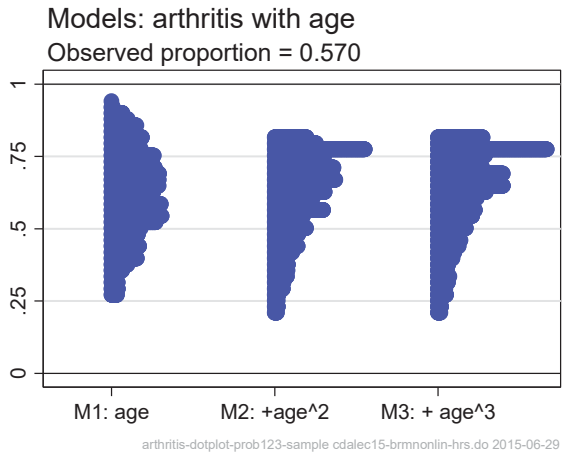
arthritis-prob2CI cdalec15-brmnonlin-hrs.do 2015-06-29

Arthritis: M3 with CI



arthritis-prob3CI cdalec15-brmnonlin-hrs.do 2015-06-29

Arthritis: Dot plot of predictions across models



Arthritis: Which model would you choose?

1. What does medical research say about the incidence of arthritis with age?
2. Are the higher order terms significant?

	F	p	df_r	df
M1age1	462.854	0.000	56.000	1.000
M2age2	111.134	0.000	56.000	1.000
M2age12	261.602	0.000	55.000	2.000
M3age3	4.294	0.043	56.000	1.000
M3age23	54.470	0.000	55.000	2.000
M3age123	168.888	0.000	54.000	3.000

3. Based on simplicity, which do you prefer?
4. Based on BIC and AIC for *non-svy models*

	AIC	BIC
aM1	21247.04	21293.42
aM2	21093.76	21147.87
aM3	21086.76	21148.60

Arthritis: gender differences at age 65

1. Suppose our concern was the movement into Medicare around age 65
2. Do men and women differ in how frequently they have arthritis?
3. We want to compute and test

   Pr(arthritis | women, age=65, x*) - Pr(arthritis | men, age=65, x*)
4. Where should we hold other variables? The global mean doesn't make sense?
5. We will restrict the sample to those with high school degrees ages 60 to 70 and compute the average discrete change for female in this subsample
6. The results (slightly edited)

```
. estimates restore aM1
. mchange female if ed4cat==2 & age>=60 & age<=70, brief

svy logit: Changes in Pr(y) | Number of obs = 2208

Expression: Pr(arthritis), predict(pr)

               |    Change    p-value
---------------+----------------------
 female vs male |     0.139      0.000

. estimates restore aM2
. mchange female if ed4cat==2 & age>=60 & age<=70, brief

               |    Change    p-value
---------------+----------------------
 female vs male |     0.139      0.000

. estimates restore aM3
. mchange female if ed4cat==2 & age>=60 & age<=70, brief

               |    Change    p-value
---------------+----------------------
 female vs male |     0.139      0.000
```

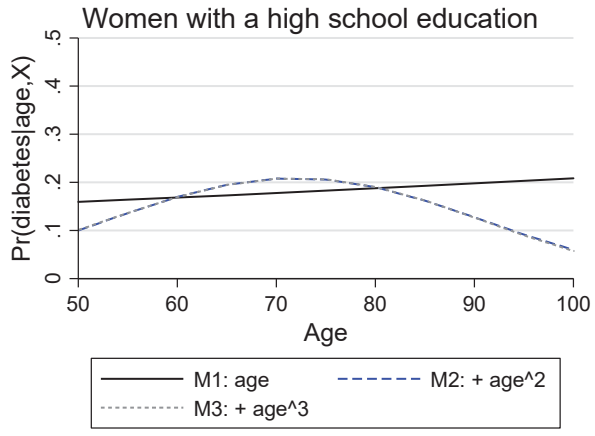
Among high school graduates ages 60 to 70, women on average have a probability of arthritis that is .14 larger than that of men (p<.001).

Logit models for diabetes

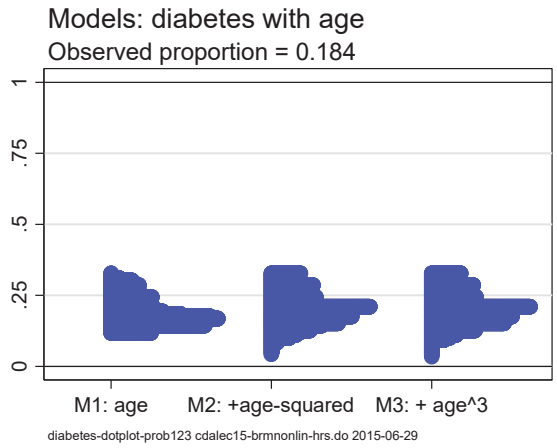
Variable	dM1	dM2	dM3
female			
female	0.80854	0.81816	0.81815
	-4.24	-3.99	-3.99
ed4cat			
12 years	0.66281	0.65679	0.65678
	-8.01	-8.18	-8.18
13-15 years	0.54123	0.55383	0.55378
	-9.66	-9.10	-9.09
16+ years	0.44993	0.45797	0.45794
	-12.83	-12.68	-12.69
age	1.00656	1.29691	1.25235
	3.14	8.22	0.66
c.age#c.age		0.99819	0.99869
		-8.14	-0.28
c.age#c.age#			
c.age			1.00000
			-0.11
_cons	0.25513	0.00004	0.00010
	-8.95	-8.98	-1.15

Note: Exponentiated coefficients; t statistics

Diabetes: Predictions for women with high school degrees



Diabetes: Dot plot of predictions across models



Diabetes: Which model would you choose?

1. What does medical research say about the incidence of diabetes with age?
2. Are the higher order terms significant?

	F	p	df_r	df
M1age1	9.864	0.003	56.000	1.000
M2age2	66.222	0.000	56.000	1.000
M2age12	33.479	0.000	55.000	2.000
M3age3	0.011	0.915	56.000	1.000
M3age23	38.202	0.000	55.000	2.000
M3age123	25.701	0.000	54.000	3.000

3. Based on simplicity, which do you prefer?
4. Based on BIC and AIC for *non-svy models*:

	AIC	BIC
dM1	16881.38	16927.77
dM2	16774.23	16828.34
dM3	16775.96	16837.81

Diabetes: Gender differences at age 65

The command:

```
mchange female if ed4cat==2 & age>=60 & age<=70, brief
```

is run for each of the models:

	Change	p-value
Model 1: female vs male	-0.033	0.000
Model 2: female vs male	-0.033	0.000
Model 3: female vs male	-0.033	0.000

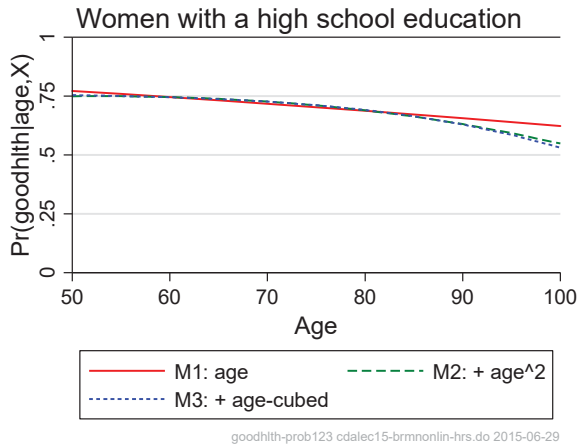
Among high school graduates ages 60 to 70, women on average have a probability of diabetes that is .03 smaller than that of men (p<.001).

Logit models for good health

Variable	gM1	gM2	gM3
female			
female	0.96264	0.96545	0.96540
	-0.91	-0.84	-0.84
ed4cat			
12 years	2.67666	2.67569	2.67557
	19.76	19.75	19.74
13-15 years	3.69172	3.72017	3.71935
	22.41	22.00	21.92
16+ years	6.33682	6.37862	6.37744
	25.74	25.46	25.42
age	0.98577	1.04372	0.97299
	-5.10	1.41	-0.10
c.age#c.age		0.99959	1.00059
		-1.97	0.16
c.age#c.age#			
c.age			1.00000
			-0.27
_cons	2.68731	0.37568	1.91674
	4.90	-0.88	0.10

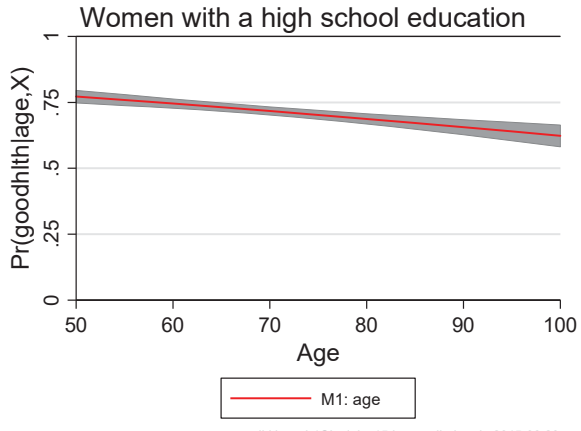
Note: Exponentiated coefficients; t statistics

Good Health: Predictions for women with HS degrees



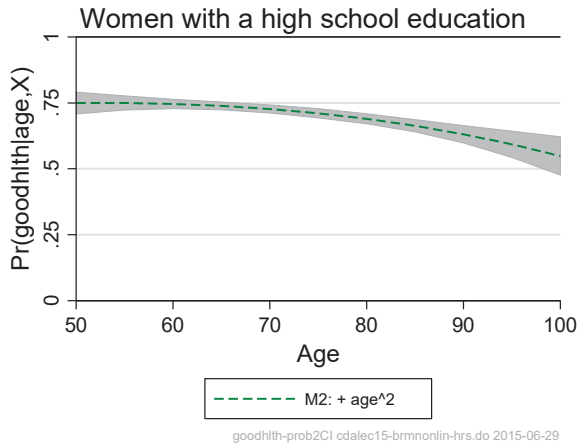
goodhlth-prob123 cdalec15-brmnonlin-hrs.do 2015-06-29

Good Health: M1

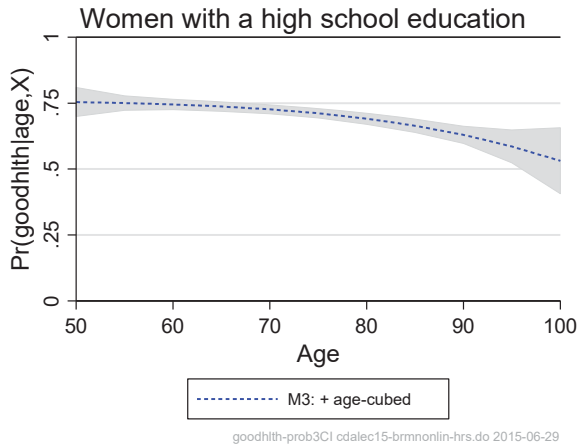


goodhlth-prob1CI cdalec15-brmnonlin-hrs.do 2015-06-29

Good Health: M2



Good Health: M3



Good Health: Dot plot of predictions across models

What causes the dual mode?

Good Health: Which model would you choose?

1. What does medical research say about the incidence of diabetes with age?
2. Are the higher order terms significant?

	F	p	df_r	df
M1age1	9.864	0.003	56.000	1.000
M2age2	3.891	0.054	56.000	1.000
M2age12	25.570	0.000	55.000	2.000
M3age3	0.075	0.785	56.000	1.000
M3age23	2.273	0.113	55.000	2.000
M3age123	18.421	0.000	54.000	3.000

3. Based on simplicity, which do you prefer?
4. Based on BIC and AIC for *non-svy models*:

	AIC	BIC
gM1	18999.61	19045.99
gM2	18989.54	19043.66
gM3	18989.99	19051.83

Good Health: Gender differences at age 65

Commands not shown

Among high school graduates who are 60 to 70, there is difference by gender in reporting good health (p>.10).

Summary of nonlinearities on the RHS

1. Just like the LRM, you should consider nonlinearities on the RHS
2. This can lead to:
   - Effects that do not plateau at 1
   - Effects that change direction
   - Predictions that are more linear
3. I start with a lowess

Part 10: Nominal outcomes 2017-03-10

Read and run

Long & Freese	Chapter 8 sections on MNLM
cdalec*.do	cdalec17-nrm-nomocc-.do; cdalec17-nrm-partyid-.do; cdalec17-nrm-ordwarm-.do

Overview

1. What does it mean for an outcome to be nominal or ordinal?
2. Introduce MNLM as a set of binary logits
3. Address challenges of interpretation
4. Briefly consider models related to the MNLM

Level of measurement

1. *S.S. Stevens* (1946) introduced terms *nominal* and *ordinal*
   a. *Nominal scales* have numbers assigned to categories as labels with *no ordering implied* by the numbers
   b. *Ordinal scales* have numbers indicating rank ordering on *one* attribute.
2. Hotly debated and critiqued when it was proposed, his taxonomy is now firmly established
3. *Ordinal models* assume his definition of ordinal. Is this a problem?

The bias-efficiency trade-off

Bias and efficiency is LRM

```
regress y x1 x2 x3    // true model

regress y x1          // bias

regress y x1 x2 x3 x4 // inefficiency
```

Effects of assuming wrong level of measurement

		True Level			
		Nominal	**Ordinal**	**Interval**	**Ratio**
Assumed	**N**	**OK**	Inefficient	Inefficient	Inefficient
Level	**O**	*Biased*	**OK**	Inefficient	Inefficient
for	**I**	*Biased*	*Biased*	**OK**	Inefficient
Analysis	**R**	*Biased*	*Biased*	*Biased*	**OK**

1. If outcome is ordinal and model is nominal, it is inefficient, but often safe choice.
2. MNLM can even be used for *interval outcomes* to explore nonlinearities
3. I start with nominal models rather than ordinal to give you the tools for assessing the implications of assuming ordinal

Review of BLM

MNLM is a *simple* extension of the BLM

The latent variable model

$$y^* = \alpha + \beta x + \varepsilon$$

The probability model

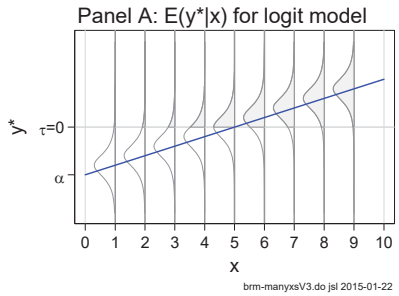
$$\Pr(y = 1 \mid x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$$

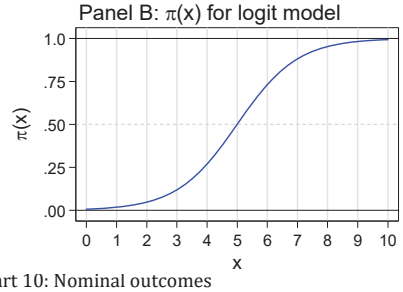
The logit model

$$\text{logit}\left[\frac{\Pr(y = 1 \mid \mathbf{x})}{\Pr(y = 0 \mid \mathbf{x})}\right] = \alpha + \beta x$$

The link between y* and Pr(y)

Graph on next page...





BLM with new notation

1. The logit model is *linear in the logit* for outcome A versus B

$$\ln\left[\frac{\Pr(y=A\,|\,\mathbf{x})}{\Pr(y=B\,|\,\mathbf{x})}\right]=\ln\Omega(\mathbf{x})$$
$$=\beta_{0,A|B}+\beta_{1,A|B}x_1+\beta_{2,A|B}x_2+\beta_{3,A|B}x_3$$

2. The model is *multiplicative in the odds*

$$\Omega(\mathbf{x})=\exp\left[\ln\Omega(\mathbf{x})\right]$$
$$=e^{\beta_{0,A|B}}e^{\beta_{1,A|B}x_1}e^{\beta_{2,A|B}x_2}e^{\beta_{3,A|B}x_3}$$
$$=\Omega(\mathbf{x},x_2)$$

3. The *odds ratio*

$$\frac{\Omega(\mathbf{x},x_2+1)}{\Omega(\mathbf{x},x_2)}=\frac{e^{\beta_{0,A|B}}e^{\beta_{1,A|B}x_1}e^{\beta_{2,A|B}(x_2+1)}e^{\beta_{3,A|B}x_3}}{e^{\beta_{0,A|B}}e^{\beta_{1,A|B}x_1}e^{\beta_{2,A|B}x_2}e^{\beta_{3,A|B}x_3}}$$
$$=e^{\beta_{2,A|B}}$$

Interpreting odds ratios

1. For a unit increase in x_2 the odds are expected to change by a factor of $\exp(\beta_{2,A|B})$, holding other variables constant.

   The odds of tenure are 1.12 times larger for women than comparable men.

2. For a standard deviation increase in x_k the odds are expected to change by a factor of $\exp(s_k\beta_{k,A|B})$, holding other variables constant.

   Increasing the number of published articles by a standard deviation increases the odds of tenure by a factor of 1.23, holding other variables constant.

Properties of OR

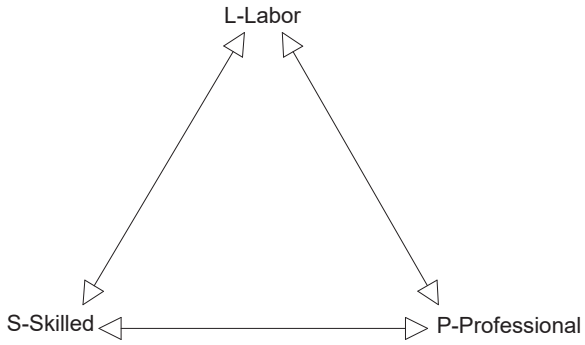
1. OR does *not* depend on the level of any variables in the model
2. OR does *not* correspond to a constant change in $\Pr(y=1|\mathbf{x})$
3. The substantive meaning of OR depends on values of all variables

Introduction to the MNLM

1. MNLM *simultaneously estimates BLMs* for all comparisons among outcomes
2. Interpretation is complicated by large number of parameters
   - With 5 outcomes, there are 10 binary logits
   - With 10 outcomes, there are 45 binary logits
3. We start with a simple model and gradually make it more complicated

MNLM with three outcomes

Categories L, S, and P with N_L, N_S, and N_P observations.



MNLM is a set of BLM that are simultaneously estimated.

1. ***L vs S*** using N_L+N_S observations

$$\ln\left[\frac{\Pr(L\mid Ed)}{\Pr(S\mid Ed)}\right]=\beta_{0,L\mid S}+\beta_{1,L\mid S}Ed$$

2. ***S vs P*** using N_S+N_P observations

$$\ln\left[\frac{\Pr(S\mid Educ)}{\Pr(P\mid Educ)}\right]=\beta_{0,S\mid P}+\beta_{1,S\mid P}Ed$$

3. ***L vs P*** using N_L+N_P observations

$$\ln\left[\frac{\Pr(L\mid Educ)}{\Pr(P\mid Educ)}\right]=\beta_{0,L\mid P}+\beta_{1,L\mid P}Ed$$

Redundancy among the BLMs

1. Since *ln(a/b)=ln a - ln b* we can add 0 as

$$\ln\left[\frac{\Pr(S\mid Ed)}{\Pr(S\mid Ed)}\right]=0=\ln\Pr(S\mid Ed)-\ln\Pr(S\mid Ed)$$

2. Then

$$\begin{aligned}\ln\left[\frac{\Pr(L\mid Ed)}{\Pr(P\mid Ed)}\right]&=\ln\Pr(L\mid Ed)-\ln\Pr(P\mid Ed)+\left[\ln\Pr(S\mid Ed)-\ln\Pr(S\mid Ed)\right]\\&=\left[\ln\Pr(L\mid Ed)-\ln\Pr(S\mid Ed)\right]+\left[\ln\Pr(S\mid Ed)-\ln\Pr(P\mid Ed)\right]\\&=\ln\left[\frac{\Pr(L\mid Ed)}{\Pr(S\mid Ed)}\right]+\ln\left[\frac{\Pr(S\mid Ed)}{\Pr(P\mid Ed)}\right]\end{aligned}$$

3. Since

$$\ln\left[\frac{\Pr(L\mid Ed)}{\Pr(P\mid Ed)}\right]=\ln\left[\frac{\Pr(L\mid Ed)}{\Pr(S\mid Ed)}\right]+\ln\left[\frac{\Pr(S\mid Ed)}{\Pr(P\mid Ed)}\right]$$

4. Then

$$\left(\beta_{0,L|S}+\beta_{1,L|S}Ed\right)+\left(\beta_{0,S|P}+\beta_{1,S|P}Ed\right)=\left(\beta_{0,L|P}+\beta_{1,L|P}Ed\right)$$

5. And

$$\beta_{L|P}=\beta_{L|S}+\beta_{S|P};\ \beta_{L|S}=\beta_{L|P}-\beta_{S|P};\ \beta_{S|P}=\beta_{L|P}-\beta_{L|S};\ etc.$$

Numeric example of link among odds

1. Frequencies of events

   Labor = 10 Skilled = 20 Prof = 30

2. Odds of events

   Labor/Skilled = 10/20

   Skilled/Prof = 20/30

   Labor/Prof = 10/30

3. Link among odds

   (Labor/Skilled) * (Skilled/Prof) = (Labor/Prof)

   (10/20) * (20/30) = (10/30)

4. Link among logits

   ln(Labor/Skilled) + ln(Skilled/Prof) = ln(Labor/Prof)

   ln(10/20) + ln(20/30) = ln(10/30)

A minimal set of coefficients

1. Because of mathematical links among odds, some coefficients are "extra"

   $$\beta_{L|P}=\beta_{L|S}+\beta_{S|P}$$

   From any two of L/S, S/P, and L/P, you can compute the third

2. In general, with J outcomes you only need J-1 comparisons
3. Each set of J-1 comparisons is a *minimal set*
4. Different software computes different minimal sets
5. For example, here are the **`mlogit`** coefficients for education...

#12 Minimal sets for 3 category MNLM (-nrm-nomocc.do)

1. Here are all comparisons in a model with 3 outcomes

```
mlogit (N=337): Factor change in the odds of occlsp

Variable: ed (sd=2.946)
------------------------------------------------------------------------
                     |        b       z   P>|z|     e^b  e^bStdX
---------------------+--------------------------------------------------
1Labor   vs 2Skilled |  -0.1711  -2.900   0.004   0.843    0.604
1Labor   vs 3Prof    |  -0.7433  -8.773   0.000   0.476    0.112
2Skilled vs 1Labor   |   0.1711   2.900   0.004   1.187    1.655
2Skilled vs 3Prof    |  -0.5722  -7.651   0.000   0.564    0.185
3Prof    vs 1Labor   |   0.7433   8.773   0.000   2.103    8.936
3Prof    vs 2Skilled |   0.5722   7.651   0.000   1.772    5.398
------------------------------------------------------------------------
```

2. Notice links among coefficients

(1L|3P)	=	(1L|2S)	+	(2S|3P)
-0.74332	=	-0.17109	+	-0.57223

3. MNLM forces these constraints to hold when it is estimating the model

Comparing MNLM to a set of BLMs

1. These equalities are necessary relationships among *population* parameters
2. They do not hold exactly with *sample* estimates from separate BLMs.
3. MNLM estimates *J-1* BLM simultaneously
   a. This enforces the logical relationship among the parameters
   b. It uses the data efficiently
4. Here's an example with real data.

`labrskil` = 1 if labor, 0 if skilled, else missing.

`profskil` = 1 if professional, 0 if skilled, else missing.

`labrprof` = 1 if labor, 0 if professional, else missing.

`occlsp` = 1 if labor, 2 if skilled, and 3 if professional.

Comparing BLM and MNLM estimates

#11 Three binary logits

```
. logit labrskil ed

  Odds of: 1Labor vs 0Skilled (N=225)

   labrskil |      b         z     P>|z|    e^b    e^bStdX      SDofX
------------+----------------------------------------------------------
         ed |  -0.18398   -2.989   0.003   0.8320   0.6485     2.3536

. logit profskil ed

  Odds of: 1Prof vs 0Skilled (N=237)

   profskil |      b         z     P>|z|    e^b    e^bStdX      SDofX
------------+----------------------------------------------------------
         ed |   0.56026    7.186   0.000   1.7511   4.9101     2.8403

. logit labrprof ed

  Odds of: 1Labor vs 0Prof (N=212)

   labrprof |      b         z     P>|z|    e^b    e^bStdX      SDofX
------------+----------------------------------------------------------
         ed |  -0.69037   -7.115   0.000   0.5014   0.1065     3.2443
```

#12 One MNLM

```
. mlogit occlsp ed

mlogit (N=337): Factor Change in the Odds of occlsp

Variable: ed (sd=2.9464271)

Odds comparing    |
Alternative 1     |
to Alternative 2  |      b         z     P>|z|
------------------+----------------------------
1Labor vs 2Skilled|  -0.1711   -2.900   0.004
1Labor vs 3Prof   |  -0.7433   -8.773   0.000
3Prof  vs 2Skilled|   0.5722    7.651   0.000
-----------------------------------------------
```

Compared to the BLM results

```
                  |      b         z     P>|z|
------------------+----------------------------
         labrskil |  -0.1840   -2.989   0.003
         labrprof |  -0.6904   -7.115   0.000
         profskil |   0.5603    7.186   0.000
-----------------------------------------------
```


Estimating MNLM with five outcomes (-nrm-nomocc.do)

#21 Descriptive statistics

```
occ    Occupation
white  Race: 1=white 0=nonwhite
ed     Years of education
exper  Years of work experience

-> tabulation of occ

 Occupation |      Freq.     Percent        Cum.
------------+-----------------------------------
     Menial |         31        9.20        9.20
    BlueCol |         69       20.47       29.67
      Craft |         84       24.93       54.60
   WhiteCol |         41       12.17       66.77
       Prof |        112       33.23      100.00
------------+-----------------------------------
      Total |        337      100.00

. sum white ed exper

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       white |       337    .9169139    .2764227          0          1
          ed |       337    13.09496    2.946427          3         20
       exper |       337    20.50148    13.95936          2         66
```


#22 Output from mlogit using base(1)

1. In **`mlogit`** option **`base(#)`** sets the base category.

2. Estimates shown for each category compared to base category Menial

```
. mlogit occ i.white ed exper, base(1) nolog

Multinomial logistic regression                 Number of obs   =        337
                                                LR chi2(12)     =     166.09
                                                Prob > chi2     =     0.0000
Log likelihood = -426.80048                     Pseudo R2       =     0.1629

         occ |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Menial       |  (base outcome)
-------------+----------------------------------------------------------------
BlueCol      |
     1.white |   1.236504   .7244352     1.71   0.088    -.1833631    2.656371
          ed |  -.0994247   .1022812    -0.97   0.331    -.2998922    .1010428
       exper |   .0047212   .0173984     0.27   0.786    -.0293789    .0388214
       _cons |   .7412336    1.51954     0.49   0.626     -2.23701    3.719477
-------------+----------------------------------------------------------------
Craft        |
     1.white |   .4723436   .6043097     0.78   0.434    -.7120817    1.656769
          ed |   .0938154    .097555     0.96   0.336    -.0973888    .2850197
       exper |   .0276838   .0166737     1.66   0.097     -.004996    .0603636
       _cons |  -1.091353   1.450218    -0.75   0.452    -3.933728    1.751022
-------------+----------------------------------------------------------------
```

```
WhiteCol     |
     1.white |   1.571385   .9027216     1.74   0.082    -.1979166    3.340687
          ed |   .3531577   .1172786     3.01   0.003     .1232959    .5830194
       exper |   .0345959   .0188294     1.84   0.066     -.002309    .0715007
       _cons |  -6.238608   1.899094    -3.29   0.001    -9.960764   -2.516453
-------------+----------------------------------------------------------------
Prof         |
     1.white |   1.774306   .7550543     2.35   0.019     .2944273    3.254186
          ed |   .7788519   .1146293     6.79   0.000     .5541826    1.003521
       exper |   .0356509    .018037     1.98   0.048      .000299    .0710028
       _cons |  -11.51833   1.849356    -6.23   0.000       -15.143   -7.893659
------------------------------------------------------------------------------
```

3. A table in a paper might look like this...

Logit Coefficients					
Comparison		*Constant*	*WHITE*	*ED*	*EXP*
$B \mid M$	β	0.741	1.237	-0.099	0.0047
	z	0.49	1.71	-0.97	0.27
$C \mid M$	β	-1.091	0.472	0.094	0.0277
	z	-0.75	0.78	0.96	1.66
$W \mid M$	β	-6.239	1.571	0.353	0.0346
	z	-3.29	1.74	3.01	1.84
$P \mid M$	β	-11.518	1.774	0.779	0.0357
	z	-6.23	2.35	6.79	1.98

4. This table corresponds to these equations

$$\ln\Omega_{B|M}(\mathbf{x}_i) = \beta_{0,B|M} + \beta_{1,B|M}WHITE + \beta_{2,B|M}ED + \beta_{3,B|M}EXP$$
$$\ln\Omega_{C|M}(\mathbf{x}_i) = \beta_{0,C|M} + \beta_{1,C|M}WHITE + \beta_{2,C|M}ED + \beta_{3,C|M}EXP \quad \ldots$$
$$\ln\Omega_{W|M}(\mathbf{x}_i) = \beta_{0,W|M} + \beta_{1,W|M}WHITE + \beta_{2,W|M}ED + \beta_{3,W|M}EXP$$
$$\ln\Omega_{P|M}(\mathbf{x}_i) = \beta_{0,P|M} + \beta_{1,P|M}WHITE + \beta_{2,P|M}ED + \beta_{3,P|M}EXP$$

5. Another minimal set

$$\ln\Omega_{B|P}(\mathbf{x}_i) = \beta_{0,B|P} + \beta_{1,B|P}WHITE + \beta_{2,B|P}ED + \beta_{3,B|P}EXP$$
$$\ln\Omega_{C|P}(\mathbf{x}_i) = \beta_{0,C|P} + \beta_{1,C|P}WHITE + \beta_{2,C|P}ED + \beta_{3,C|P}EXP$$
$$\ln\Omega_{W|P}(\mathbf{x}_i) = \beta_{0,W|P} + \beta_{1,W|P}WHITE + \beta_{2,W|P}ED + \beta_{3,W|P}EXP$$
$$\ln\Omega_{M|P}(\mathbf{x}_i) = \beta_{0,M|P} + \beta_{1,M|P}WHITE + \beta_{2,M|P}ED + \beta_{3,M|P}EXP$$

6. And so on

7. Which minimal set should you use?

#22 Examining all ORs (extracted output)

1. Do not judge statistical significance using tests from a minimal set

Base BlueCol: 0 significant coefficients

```
                          e^b    P>|z|
WhiteCol vs BlueCol   1.3978    0.720
Prof     vs BlueCol   1.7122    0.501
Craft    vs BlueCol   0.4657    0.227
Menial   vs BlueCol   0.2904    0.088
```

Base Craft: 1 significant coefficient

```
                          e^b    P>|z|
BlueCol  vs Craft     2.1472    0.227
WhiteCol vs Craft     3.0013    0.179
Prof     vs Craft     3.6765    0.044
Menial   vs Craft     0.6235    0.434
```


Base Menial: 1 significant coefficient

```
                          e^b    P>|z|
Craft    vs Menial    1.6037    0.434
BlueCol  vs Menial    3.4436    0.088
WhiteCol vs Menial    4.8133    0.082
Prof     vs Menial    5.8962    0.019
```

Base Prof: 2 significant coefficients

```
                          e^b    P>|z|
WhiteCol vs Prof      0.8163    0.815
BlueCol  vs Prof      0.5840    0.501
Craft    vs Prof      0.2720    0.044
Menial   vs Prof      0.1696    0.019
```

Base WhiteCol: 0 significant coefficients

```
                          e^b    P>|z|
Prof     vs WhiteCol  1.2250    0.815
BlueCol  vs WhiteCol  0.7154    0.720
Craft    vs WhiteCol  0.3332    0.179
Menial   vs WhiteCol  0.2078    0.082
```

2. Looking at all ORs can be overwhelming....

```
. listcoef
```



```
mlogit (N=337): Factor change in the odds of occ

Variable: 1.white (sd=0.276)
------------------------------------------------------------------------
                        |       b        z    P>|z|     e^b e^bStdX
------------------------+-----------------------------------------------
Menial       vs BlueCol | -1.2365   -1.707    0.088   0.290   0.710
Menial       vs Craft   | -0.4723   -0.782    0.434   0.624   0.878
Menial       vs WhiteCol| -1.5714   -1.741    0.082   0.208   0.648
Menial       vs Prof    | -1.7743   -2.350    0.019   0.170   0.612
BlueCol      vs Menial  |  1.2365    1.707    0.088   3.444   1.407
BlueCol      vs Craft   |  0.7642    1.208    0.227   2.147   1.235
BlueCol      vs WhiteCol| -0.3349   -0.359    0.720   0.715   0.912
BlueCol      vs Prof    | -0.5378   -0.673    0.501   0.584   0.862
Craft        vs Menial  |  0.4723    0.782    0.434   1.604   1.139
Craft        vs BlueCol | -0.7642   -1.208    0.227   0.466   0.810
Craft        vs WhiteCol| -1.0990   -1.343    0.179   0.333   0.738
Craft        vs Prof    | -1.3020   -2.011    0.044   0.272   0.698
WhiteCol     vs Menial  |  1.5714    1.741    0.082   4.813   1.544
WhiteCol     vs BlueCol |  0.3349    0.359    0.720   1.398   1.097
WhiteCol     vs Craft   |  1.0990    1.343    0.179   3.001   1.355
WhiteCol     vs Prof    | -0.2029   -0.233    0.815   0.816   0.945
Prof         vs Menial  |  1.7743    2.350    0.019   5.896   1.633
Prof         vs BlueCol |  0.5378    0.673    0.501   1.712   1.160
Prof         vs Craft   |  1.3020    2.011    0.044   3.677   1.433
Prof         vs WhiteCol|  0.2029    0.233    0.815   1.225   1.058
------------------------------------------------------------------------
```

```
Variable: ed (sd=2.946)
------------------------------------------------------------------------------
                        |       b        z    P>|z|     e^b e^bStdX
------------------------+-----------------------------------------------------
Menial       vs BlueCol |  0.0994    0.972    0.331   1.105   1.340
Menial       vs Craft   | -0.0938   -0.962    0.336   0.910   0.758
Menial       vs WhiteCol| -0.3532   -3.011    0.003   0.702   0.353
Menial       vs Prof    | -0.7789   -6.795    0.000   0.459   0.101
BlueCol      vs Menial  | -0.0994   -0.972    0.331   0.905   0.746
BlueCol      vs Craft   | -0.1932   -2.494    0.013   0.824   0.566
BlueCol      vs WhiteCol| -0.4526   -4.425    0.000   0.636   0.264
BlueCol      vs Prof    | -0.8783   -8.735    0.000   0.415   0.075
Craft        vs Menial  |  0.0938    0.962    0.336   1.098   1.318
Craft        vs BlueCol |  0.1932    2.494    0.013   1.213   1.767
Craft        vs WhiteCol| -0.2593   -2.773    0.006   0.772   0.466
Craft        vs Prof    | -0.6850   -7.671    0.000   0.504   0.133
WhiteCol     vs Menial  |  0.3532    3.011    0.003   1.424   2.831
WhiteCol     vs BlueCol |  0.4526    4.425    0.000   1.572   3.794
WhiteCol     vs Craft   |  0.2593    2.773    0.006   1.296   2.147
WhiteCol     vs Prof    | -0.4257   -4.616    0.000   0.653   0.285
Prof         vs Menial  |  0.7789    6.795    0.000   2.179   9.923
Prof         vs BlueCol |  0.8783    8.735    0.000   2.407  13.300
Prof         vs Craft   |  0.6850    7.671    0.000   1.984   7.526
Prof         vs WhiteCol|  0.4257    4.616    0.000   1.531   3.505
------------------------------------------------------------------------------
```



```
Variable: exper (sd=13.959)
------------------------------------------------------------------------------
                        |       b        z    P>|z|     e^b e^bStdX
------------------------+-----------------------------------------------------
Menial       vs BlueCol | -0.0047   -0.271    0.786   0.995   0.936
Menial       vs Craft   | -0.0277   -1.660    0.097   0.973   0.679
Menial       vs WhiteCol| -0.0346   -1.837    0.066   0.966   0.617
Menial       vs Prof    | -0.0357   -1.977    0.048   0.965   0.608
BlueCol      vs Menial  |  0.0047    0.271    0.786   1.005   1.068
BlueCol      vs Craft   | -0.0230   -1.829    0.067   0.977   0.726
BlueCol      vs WhiteCol| -0.0299   -1.954    0.051   0.971   0.659
BlueCol      vs Prof    | -0.0309   -2.147    0.032   0.970   0.649
Craft        vs Menial  |  0.0277    1.660    0.097   1.028   1.472
Craft        vs BlueCol |  0.0230    1.829    0.067   1.023   1.378
Craft        vs WhiteCol| -0.0069   -0.495    0.621   0.993   0.908
Craft        vs Prof    | -0.0080   -0.627    0.531   0.992   0.895
WhiteCol     vs Menial  |  0.0346    1.837    0.066   1.035   1.621
WhiteCol     vs BlueCol |  0.0299    1.954    0.051   1.030   1.517
WhiteCol     vs Craft   |  0.0069    0.495    0.621   1.007   1.101
WhiteCol     vs Prof    | -0.0011   -0.073    0.941   0.999   0.985
Prof         vs Menial  |  0.0357    1.977    0.048   1.036   1.645
Prof         vs BlueCol |  0.0309    2.147    0.032   1.031   1.540
Prof         vs Craft   |  0.0080    0.627    0.531   1.008   1.118
Prof         vs WhiteCol|  0.0011    0.073    0.941   1.001   1.015
------------------------------------------------------------------------------
```

Out task is to make sense out of all of these numbers

Roadmap

1. The MNLM as a probability model
2. Estimation
3. Omnibus tests of each regressor
4. Tests if categories can be combined
5. Methods of interpretation
    a. Odds ratios
    b. Predicted probabilities

MNLM as a Probability Model

1. We motivated MNLM as a set of binary logits
2. Outcomes are logs of odds Log(Pr_A/Pr_B)
3. We can solve the equations in terms of these probabilities
4. If *y* has J categories 1 to J, let $\Pr(y = m \mid \mathbf{x})$ be the probability of *m* given **x**:

$$\Pr(y_i = m \mid \mathbf{x}_i) = \frac{\exp(\mathbf{x}_i\boldsymbol{\beta}_{m|J})}{\sum_{j=1}^{J}\exp(\mathbf{x}_i\boldsymbol{\beta}_{j|J})}$$

5. You get the same values regardless of the base J you use

ML estimation

1. p_i is the probability of observing the value of *y* actually observed for person i
   a. If y=1 for person i, $p_i = Pr(y=1|x_i)$
   b. If y=2 for person i, $p_i = Pr(y=2|x_i)$
   c. If y=3 for person i, $p_i = Pr(y=3|x_i)$
   d. Etc.
2. If the observations are independent, the likelihood equation is

$$L(\boldsymbol{\beta}_{2|J},\ldots,\boldsymbol{\beta}_{J|J} \mid \mathbf{y},\mathbf{X}) = \prod_{i=1}^{N} p_i$$

3. Solving for the parameters works well even with small samples

Software Issues

Different programs estimate different minimal sets of coefficients

Testing

Contrasts

1. Tests in standard output are for the minimal set of coefficients.
   a. Your program computes $\beta_{k,m|J}$
   b. Another program computes $\beta_{k,m|L}$
2. Stata's **`baseoutcome()`** sets the reference category:

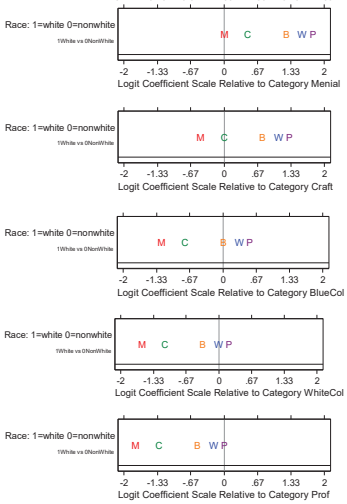
```
mlogit occ white ed exper, baseoutcome(1)
mlogit occ white ed exper, baseoutcome(2)
```

3. From the minimal set you can compute comparisons of other categories, which are called *contrasts*
4. **`listcoef`** computes all contrasts automatically. For example,...

#22 All contrast for white

```
. listcoef white
mlogit (N=337): Factor change in the odds of occ

Variable: 1.white (sd=0.276)
-------------------------------------------------------------------------
                        |       b         z    P>|z|      e^b e^bStdX
------------------------+------------------------------------------------
Menial       vs BlueCol | -1.2365   -1.707    0.088    0.290   0.710
Menial       vs Craft   | -0.4723   -0.782    0.434    0.624   0.878
Menial       vs WhiteCol| -1.5714   -1.741    0.082    0.208   0.648
Menial       vs Prof    | -1.7743   -2.350    0.019    0.170   0.612
BlueCol      vs Menial  |  1.2365    1.707    0.088    3.444   1.407
BlueCol      vs Craft   |  0.7642    1.208    0.227    2.147   1.235
BlueCol      vs WhiteCol| -0.3349   -0.359    0.720    0.715   0.912
BlueCol      vs Prof    | -0.5378   -0.673    0.501    0.584   0.862
Craft        vs Menial  |  0.4723    0.782    0.434    1.604   1.139
Craft        vs BlueCol | -0.7642   -1.208    0.227    0.466   0.810
Craft        vs WhiteCol| -1.0990   -1.343    0.179    0.333   0.738
Craft        vs Prof    | -1.3020   -2.011    0.044    0.272   0.698
WhiteCol     vs Menial  |  1.5714    1.741    0.082    4.813   1.544
WhiteCol     vs BlueCol |  0.3349    0.359    0.720    1.398   1.097
WhiteCol     vs Craft   |  1.0990    1.343    0.179    3.001   1.355
WhiteCol     vs Prof    | -0.2029   -0.233    0.815    0.816   0.945
Prof         vs Menial  |  1.7743    2.350    0.019    5.896   1.633
Prof         vs BlueCol |  0.5378    0.673    0.501    1.712   1.160
Prof         vs Craft   |  1.3020    2.011    0.044    3.677   1.433
Prof         vs WhiteCol|  0.2029    0.233    0.815    1.225   1.058
-------------------------------------------------------------------------
```


Graphs with different base categories



Testing that a variable has no effect

1. The hypothesis that x_k has no effect involves J-1 coefficients

$$H_0: \beta_{k,B|M} = \beta_{k,C|M} = \beta_{k,W|M} = \beta_{k,P|M} = 0$$

2. This is *not* equivalent to combined tests of individual coefficients

$$H_0: \beta_{k,B|M} = 0 \qquad H_0: \beta_{k,C|M} = 0$$

$$H_0: \beta_{k,W|M} = 0 \qquad H_0: \beta_{k,P|M} = 0$$

3. The Wald statistic for $H_0: \boldsymbol{\beta}_k = \mathbf{0}$: is:

$$W_k = \hat{\boldsymbol{\beta}}_k' Var\left(\hat{\boldsymbol{\beta}}_k\right)^{-1} \hat{\boldsymbol{\beta}}_k$$

where $W_k \sim \chi^2_{J-1}$ if H_0 is true

4. This is computed with **`test`** or the SPost **`mlogtest`**

#24 Wald tests using test

```
. test 1.white // test will work with svy estimates too
. * 1.white is the variable created from i.white!

 ( 1)  [Menial]o.white = 0
 ( 2)  [BlueCol]white = 0
 ( 3)  [Craft]white = 0
 ( 4)  [WhiteCol]white = 0
 ( 5)  [Prof]white = 0
       Constraint 1 dropped

           chi2(  4) =    8.15
         Prob > chi2 =    0.0863

. test ed
  <snip>

. test exper
  <snip?
```

← This is $\beta_{white,M|M}$

#24 Wald tests using mlogtest

```
. mlogtest, wald

Wald tests for independent variables (N=337)

  Ho: All coefficients associated with given variable(s) are 0

                |      chi2    df   P>chi2
----------------+------------------------
        1.white |     8.149     4    0.086
             ed |    84.968     4    0.000
          exper |     7.995     4    0.092
```

1. The effect of race is not significant at the .05 level (G^2=8.15, df=4).
2. The effect of education is significant at the .01 level.
3. The effect of experience is significant at the .10 level but not at the .05 level.

#25 LR test using mlogtest

```
. estimates restore full
(results full are active now)

. mlogtest, lr

Likelihood-ratio tests for independent variables (N=337)

                |      chi2    df   P>chi2
----------------+------------------------
        1.white |     8.095     4    0.088
             ed |   156.937     4    0.000
          exper |     8.561     4    0.073
```

1. We conclude:

   The hypothesis that being white does not affect occupational attainment can be rejected at the .10 level, but not at the .05 level (LRX^2=8.10, df=4).

2. More simply:

   The effect of race is significant at the .09 level.

Comparing Wald and LR Tests

	LR			Wald		
	G^2	df	p	W	df	p
WHITE	8.10	4	0.09	8.15	4	0.09
ED	156.94	4	<0.01	84.97	4	<0.01
EXPER	8.56	4	0.09	7.99	4	0.09

1. I compute both tests for didactic purposes; in practice, only compute one
2. If you do test all coefficients, you might be misled by the minimal set
3. Testing that *all* coefficients for a variable are simultaneously zero might *not* be appropriate for your substantive goals
   - When we plot OR's this will be easy to see

* Testing that outcomes can be combined

1. If no regressor significantly affects the odds of P vs W, we say
   - P and W are *indistinguishable*.
2. The hypothesis that P and W are indistinguishable is

$$H_0: \beta_{1,P|W} = \beta_{2,P|W} = \beta_{3,P|W} = 0$$

3. Tests of indistinguishably can be computed for all pairs of outcomes

Example on next page...

Tests for combining can lead to inconsistencies

1. An LR test supports combining M and B (p=.251), suggesting

   M can be combined with B **M = B**

2. An LR test supports combining M and C (p=.337), suggesting

   M can be combined with C **M = C**

3. Algebraically, this suggests

   B can be combined with C **B = C**

4. An LR test **rejects** the hypothesis that B and C (p=.003) can be combined:

   B cannot be combined with C **B ≠ C**

5. **Tests of hypothesis are not algebraic statements.**
6. If you decide to combine categories
   a. Estimate the model with the new outcome
   b. Compute tests of indistinguishability for the new model

#26 Wald tests for combining outcomes

mlogtest to test for indistinguishably

```
. mlogtest, combine

Wald tests for combining alternatives (N=337)

Ho: All coefficients except intercepts associated with a given pair
    of alternatives are 0 (i.e., alternatives can be combined)

                |      chi2    df    P>chi2
----------------+--------------------------
Menial & BlueCol |     3.994     3    0.262
  Menial & Craft |     3.203     3    0.361
Menial & White~l |    11.951     3    0.008
   Menial & Prof |    48.190     3    0.000
 BlueCol & Craft |     8.441     3    0.038
BlueCol & Whit~l |    20.055     3    0.000
  BlueCol & Prof |    76.393     3    0.000
Craft & WhiteCol |     8.892     3    0.031
    Craft & Prof |    60.583     3    0.000
 WhiteCol & Prof |    22.203     3    0.000
```


Specification searches

1. Tests for combining categories and tests that all coefficients for a variable are zero can be used in a *specification search*
2. Be careful not to *over-fit* your data
3. Examine individual coefficients before revising your model
4. Think substantively about changes to your model
5. In models constructed from tests using the same data, significance levels are invalid.
   - Consider randomly dividing the sample into an *exploration subsample* and a *verification subsample*

Overview of interpretation

1. The MNLM has *many* parameters
2. Do *not*
   a. Present a minimal set of parameters without interpretation
   b. Include stars for significance of individual coefficients
   c. Ignore coefficients not in the minimal set are ignored
   d. Ignore direction of effects and overall significance
3. All coefficients can be interpreted
   a. *Odds ratios* for all contrasts can tell part of the story
   b. *Predicted probabilities* and *marginal effects* provide substantive insights
4. We start with marginal effects

Using probabilities for interpretation

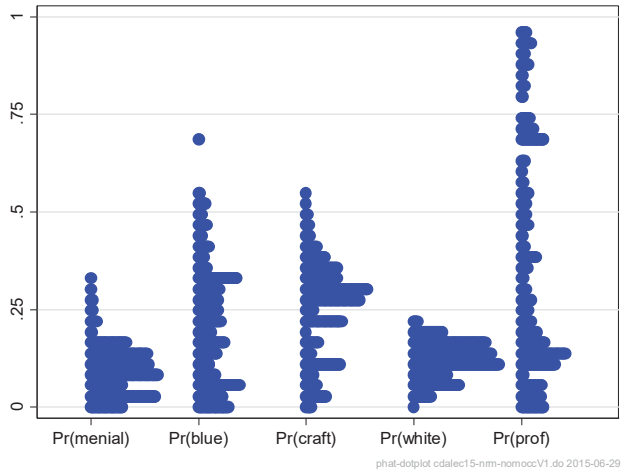
1. Check within-sample variation in predicted probabilities with **`predict`**
2. Examine the "effect" of x_k with marginal effects, tables, and graphs
3. Some methods are shown in this chapter; others in the next chapter
   - Methods for ordinal models generally work for nominal model

Example: occupation type (-nrm-nomocc.do)

#31 Distribution of predictions

```
. predict prmenial prblue prcraft prwhite prprof
(option pr assumed; predicted probabilities)

. label var prmenial "Pr(menial)"
. label var prblue   "Pr(blue)"
. label var prcraft  "Pr(craft)"
. label var prwhite  "Pr(white)"
. label var prprof   "Pr(prof)"
. local graphnm "`pgm'-phat-dotplot"

. dotplot prmenial prblue prcraft prwhite prprof, ///
>     ylab(0(.25)1, grid gmin gmax)
```

Plot on next page...



- The biggest effects are likely to occur with professional outcomes
- What causes the range in predictions? Does it concern you?
- Are there outliers to examine?

Marginal change & discrete change

While OR's are often the first thing people examine, marginal effects are usually more informative.

Marginal change

1. The marginal change (MC) is

$$\frac{\partial \Pr(y=m\,|\,\mathbf{x})}{\partial x_k} = \Pr(y=m\,|\,\mathbf{x})\left[\beta_{k,m|J} - \sum_{j=1}^{J}\beta_{k,j|J}\Pr(y=j\,|\,\mathbf{x})\right]$$

2. The formula combines many coefficients and the since of the MC does not need to be the same sign as the $\beta_{k,m|J}$
3. The sign of the MC can change as x_k changes

Discrete change

1. The discrete change (DC) for x_k

$$\frac{\Delta \Pr(y=m\,|\,\mathbf{x}^*)}{\Delta x_k} = \Pr(y=m\,|\,\mathbf{x}^*, \text{End } x_k) - \Pr(y=m\,|\,\mathbf{x}^*, \text{Start } x_k)$$

2. This can be interpreted as

   If x_k changes from the **start value** to the **end value**, the predicted probability of outcome *m* changes by $\Delta \Pr(y=m\,|\,\mathbf{x}^*)/\Delta x_k$, holding other variables at the specified values.

The DC for x_k is affected by

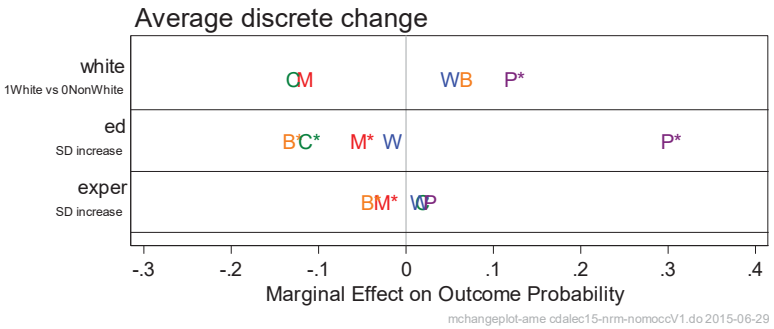
1. All coefficients for x_k
2. The amount of change in x_k
3. The starting value of x_k
4. The values of other regressors

#32 average marginal effects (AMEs)

```
. mchange, amount(one sd) brief

mlogit: Changes in Pr(y) | Number of obs = 337

Expression: Pr(occ), predict(outcome())

                     |    Menial    BlueCol      Craft    WhiteCol       Prof
---------------------+-----------------------------------------------------
white                |
 1White vs 0NonWhite |    -0.116      0.069     -0.129      0.052      0.124
             p-value |     0.143      0.315      0.164      0.327      0.054
ed                   |
                +1   |    -0.017     -0.050     -0.033      0.002      0.099
             p-value |     0.000      0.000      0.000      0.662      0.000
                +SD  |    -0.050     -0.129     -0.111     -0.014      0.304
             p-value |     0.000      0.000      0.000      0.333      0.000
exper                |
                +1   |    -0.002     -0.003      0.002      0.001      0.002
             p-value |     0.127      0.051      0.338      0.329      0.187
                +SD  |    -0.023     -0.040      0.019      0.017      0.027
             p-value |     0.078      0.031      0.407      0.385      0.228
```

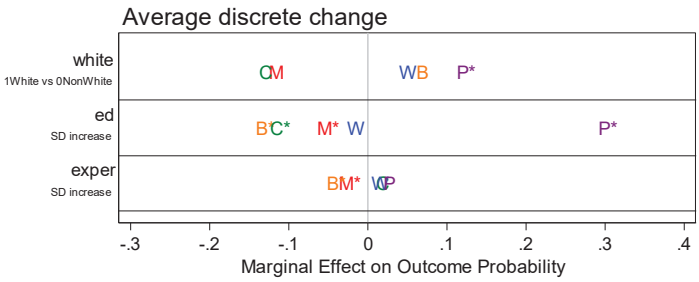
Graph using mchangeplot…

```
mchangeplot, amount(sd sd) min(-.3) max(.4) gap(.1) mcol(rainbow) aspect(.3)
title(Average discrete change, position(11)) leftmargin(4) sig(.10)
```



On average, being white decreases the probability of a menial job by .12 and a craft job by .13, while increasing the probabilities of blue collar jobs by .07, white color jobs by .05, and professional jobs by .12. However, only the change in professional jobs is statistically significant at the .05 level.

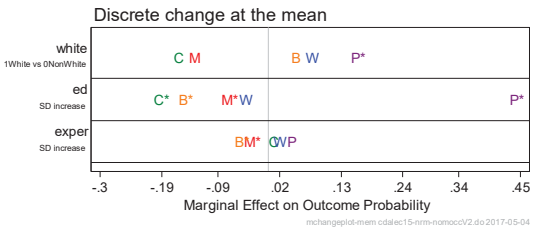
More interpretation



- The effects of a standard deviation increase in education are largest, with an increase of about .3 for professional occupations.
- The effects of race are substantial, with blacks on average being less likely to enter blue collar, white collar, or professional jobs.
- The changes due to a standard deviation increase in experience are much smaller and show that experience increases the probabilities of more highly skilled occupations.

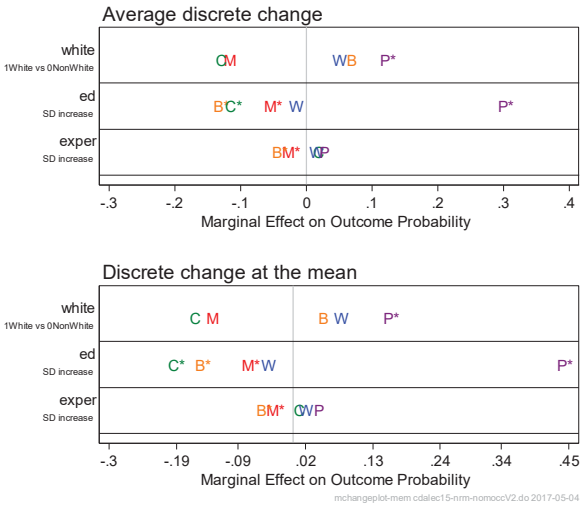
#32 Marginal effect at the mean

```
mchange, atmeans amount(one sd)
```



- For someone who is average on all characteristics, the effects of a standard deviation increase in education are largest, with an increase of nearly .45 for professional occupations.
- The effects of race are substantial. Blacks with average education and experience are less likely to enter blue collar, white collar, or professional jobs than comparable white respondents.

AME compared to MEM



Odds ratios

1. Discrete change (DC) is useful, but partial reflection of the process
2. DC's do not indicate the dynamics among the outcomes
   - For example, a decrease in education increases blue collar and craft jobs, but how does it affect craft jobs relative to blue collar jobs?
3. The dynamics between categories is reflected by the OR.
4. Consider the odds of outcome m versus n, highlighting x_2

$$\Omega_{m|n}(\mathbf{x}, x_2) = e^{\beta_{0,m|n}}\, e^{\beta_{1,m|n}x_1} e^{\beta_{2,m|n}x_2} e^{\beta_{3,m|n}x_3}$$

5. If x_2 is changed by 1, then

$$\Omega_{m|n}(\mathbf{x}, x_2+1) = e^{\beta_{0,m|n}}\, e^{\beta_{1,m|n}x_1} e^{\beta_{2,m|n}x_2} e^{\beta_{2,m|n}} e^{\beta_{3,m|n}x_3}$$

6. The OR for x_2 for a unit change is

$$\frac{\Omega_{m|n}(\mathbf{x}, x_2+1)}{\Omega_{m|n}(\mathbf{x}, x_2)} = \frac{e^{\beta_{0,m|n}}\, e^{\beta_{1,m|n}x_1} e^{\beta_{2,m|n}x_2} e^{\beta_{2,m|n}} e^{\beta_{3,m|n}x_3}}{e^{\beta_{0,m|n}}\, e^{\beta_{1,m|n}x_1} e^{\beta_{2,m|n}x_2} e^{\beta_{3,m|n}x_3}} = e^{\beta_{2,m|n}}$$

7. Interpretation
   a. For a unit increase in $\mathbf{x_k}$ the odds of **m** versus **n** change by a factor of **exp($\beta_{k,m|n}$)**, holding other variables *constant*.
   b. For a standard deviation change in x_k, the odds are expected to change by a factor of exp($s_k \beta_{k,m|n}$), holding other variables constant.
8. *Unlike the DC or MC, the OR does not depend on the level of any of the variables*
   - The only requirement is that one variable changes while the others do not

Challenges interpreting ORs

1. The meaning of an OR depends on the outcome probability
   - The outcome probability depends on all parameters and values of all variables
2. You cannot use OR's if you have linked variables such as *age* and *age-squared*
   - Unless you computed them with advanced methods
3. There can be a *lot* of ORs to interpret
   - We address this problem first

ORs for white

Factor Change			Outcome n				
in the Odds of m vs n			M	B	C	W	P
Outcome	M	Menial	- - -	0.29	0.62	0.21	0.17
m	B	Blue Collar	3.44	- - -	2.15	0.72	0.58
	C	Craft	1.60	0.47	- - -	0.33	0.27
	W	White Collar	4.81	1.40	3.00	- - -	0.82
	P	Professional	5.90	1.71	3.68	1.23	- - -

With listcoef...

#41 Examining all ORs with listcoef

```
Variable: 1.white (sd=0.276)
------------------------------------------------------------------------
                        |       b        z    P>|z|     e^b e^bStdX
------------------------+-----------------------------------------------
Menial       vs BlueCol | -1.2365   -1.707    0.088   0.290   0.710
Menial       vs Craft   | -0.4723   -0.782    0.434   0.624   0.878
Menial       vs WhiteCol| -1.5714   -1.741    0.082   0.208   0.648
Menial       vs Prof    | -1.7743   -2.350    0.019   0.170   0.612
BlueCol      vs Menial  |  1.2365    1.707    0.088   3.444   1.407
BlueCol      vs Craft   |  0.7642    1.208    0.227   2.147   1.235
BlueCol      vs WhiteCol| -0.3349   -0.359    0.720   0.715   0.912
BlueCol      vs Prof    | -0.5378   -0.673    0.501   0.584   0.862
Craft        vs Menial  |  0.4723    0.782    0.434   1.604   1.139
Craft        vs BlueCol | -0.7642   -1.208    0.227   0.466   0.810
Craft        vs WhiteCol| -1.0990   -1.343    0.179   0.333   0.738
Craft        vs Prof    | -1.3020   -2.011    0.044   0.272   0.698
WhiteCol     vs Menial  |  1.5714    1.741    0.082   4.813   1.544
WhiteCol     vs BlueCol |  0.3349    0.359    0.720   1.398   1.097
WhiteCol     vs Craft   |  1.0990    1.343    0.179   3.001   1.355
WhiteCol     vs Prof    | -0.2029   -0.233    0.815   0.816   0.945
Prof         vs Menial  |  1.7743    2.350    0.019   5.896   1.633
Prof         vs BlueCol |  0.5378    0.673    0.501   1.712   1.160
Prof         vs Craft   |  1.3020    2.011    0.044   3.677   1.433
Prof         vs WhiteCol|  0.2029    0.233    0.815   1.225   1.058
------------------------------------------------------------------------
```

Plotting ORs for BLM

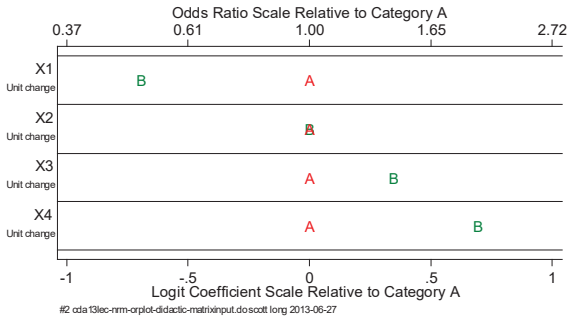
1. An *odds ratio plot* is an easy way to see complex patterns in the estimates
2. Consider a BLM with test coefficients

x	$\beta_{B\|A}$	$\exp(\beta_{B\|A})$	p
x_1	-0.693	0.500	0.02
x_2	0.000	1.000	0.99
x_3	0.347	1.414	0.11
x_4	0.693	2.000	0.04

3. Think of the *OR as the distance* between outcomes A and B

Graph on next page...

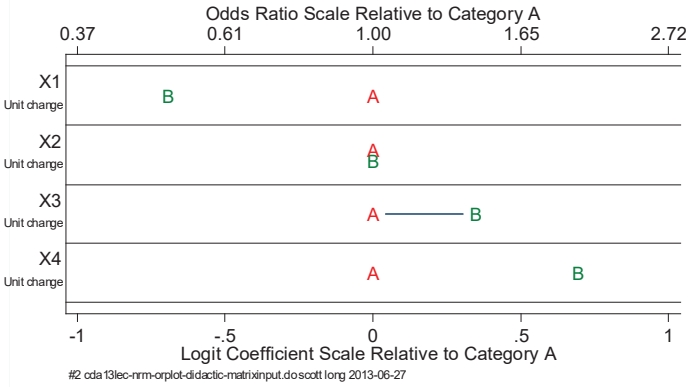
OR plot for binary logit



x	$\beta_{B\|A}$	$\exp(\beta_{B\|A})$
x_1	-0.693	0.500
x_2	0.000	1.000
x_3	0.347	1.414
x_4	0.693	2.000

Lack of significance indicated by connecting line

1. Lack of significance is shown by a connecting line
2. If a coefficient is *not significant*, the two outcomes are *tied together*

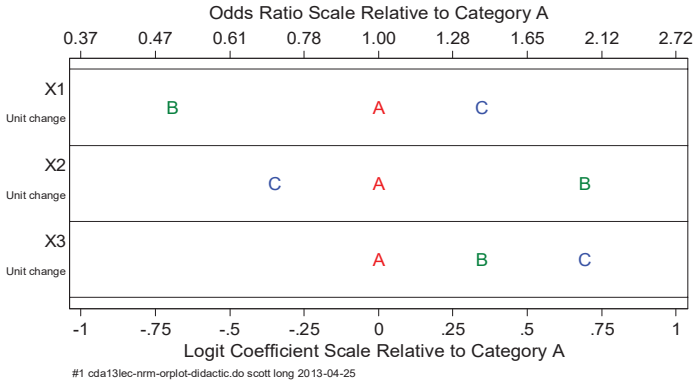


OR plot for three categories

1. Consider a hypothetical model with three outcomes:

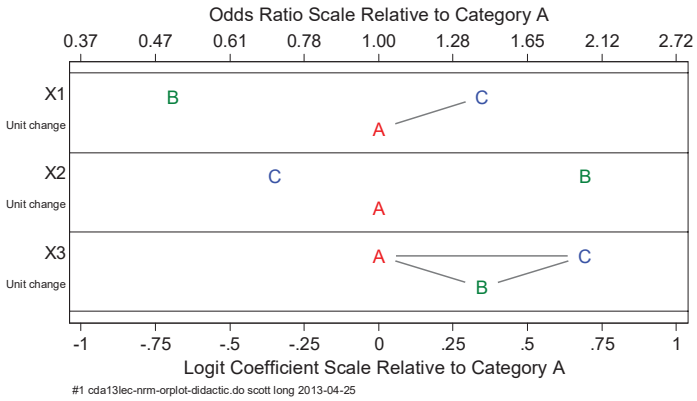
Comparison		x1	x2	x3
B|A	$\beta_{B\|A}$	-.693	0.693	0.347
	$\exp(\beta_{B\|A})$	0.500	2.000	1.414
	p	0.04	0.01	0.42
C|A	$\beta_{C\|A}$	0.347	-.347	0.693
	$\exp(\beta_{C\|A})$	1.414	0.707	2.000
	p	0.21	0.04	0.37
C|B	$\beta_{C\|B}$	1.040	-1.040	0.346
	$\exp(\beta_{C\|B})$	2.828	0.354	1.414
	p	0.02	0.03	0.21

2. Plotting *relative to A*

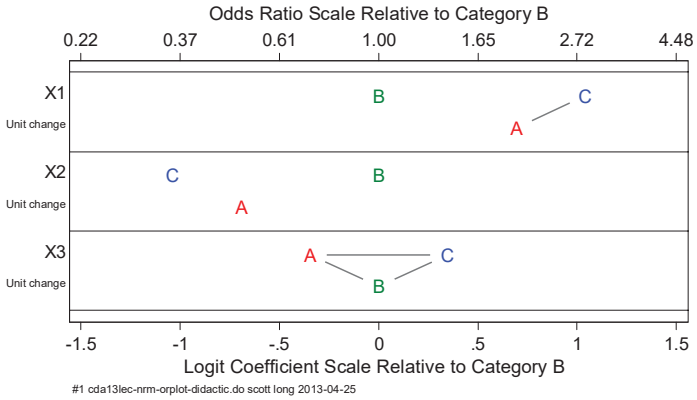


3. Consider the implicit constraints for x_1:
   a. Distance: B -> A = .693
   b. Distance: A -> C = .347
   c. Distance: B -> C = 1.040 = .693 + .347

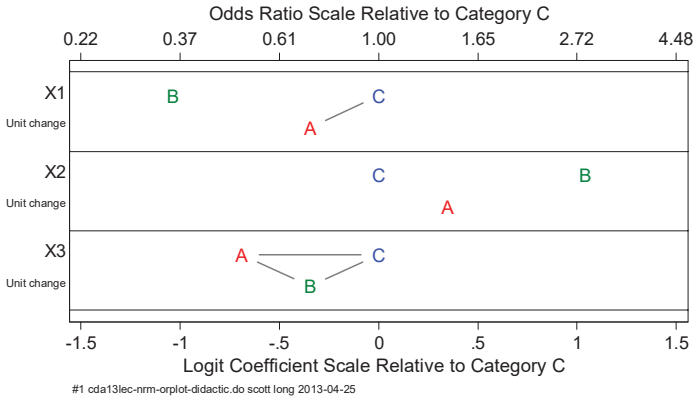
4. Indicate *non-significance with a connecting line*



5. Plotting *relative to B* shows the same information



6. Plotting *relative to C* shows the same information



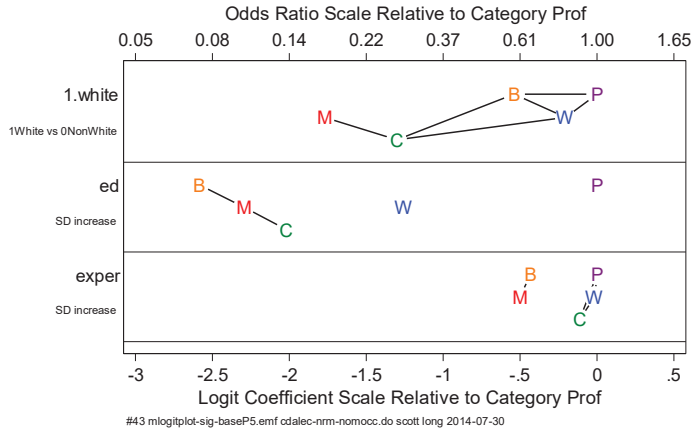
#41-47 OR plot for occupational attainment

1. Next we plot the three pages of OR's shown above

```
Variable: 1.white (sd=0.276)
                              b        z    P>|z|      e^b e^bStdX
Menial       vs BlueCol  | -1.2365   -1.707    0.088   0.290   0.710
Menial       vs Craft    | -0.4723   -0.782    0.434   0.624   0.878
Menial       vs WhiteCol | -1.5714   -1.741    0.082   0.208   0.648
Menial       vs Prof     | -1.7743   -2.350    0.019   0.170   0.612
BlueCol      vs Menial   |  1.2365    1.707    0.088   3.444   1.407
BlueCol      vs Craft    |  0.7642    1.208    0.227   2.147   1.235
BlueCol      vs WhiteCol | -0.3349   -0.359    0.720   0.715   0.912
BlueCol      vs Prof     | -0.5378   -0.673    0.501   0.584   0.862
Craft        vs Menial   |  0.4723    0.782    0.434   1.604   1.139
Craft        vs BlueCol  | -0.7642   -1.208    0.227   0.466   0.810
Craft        vs WhiteCol | -1.0990   -1.343    0.179   0.333   0.738
Craft        vs Prof     | -1.3020   -2.011    0.044   0.272   0.698
WhiteCol     vs Menial   |  1.5714    1.741    0.082   4.813   1.544
WhiteCol     vs BlueCol  |  0.3349    0.359    0.720   1.398   1.097
WhiteCol     vs Craft    |  1.0990    1.343    0.179   3.001   1.355
WhiteCol     vs Prof     | -0.2029   -0.233    0.815   0.816   0.945
Prof         vs Menial   |  1.7743    2.350    0.019   5.896   1.633
Prof         vs BlueCol  |  0.5378    0.673    0.501   1.712   1.160
Prof         vs Craft    |  1.3020    2.011    0.044   3.677   1.433
Prof         vs WhiteCol |  0.2029    0.233    0.815   1.225   1.058
```

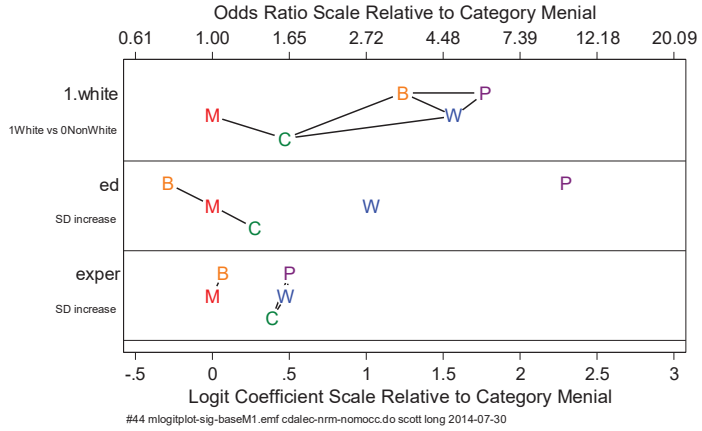
And so on...

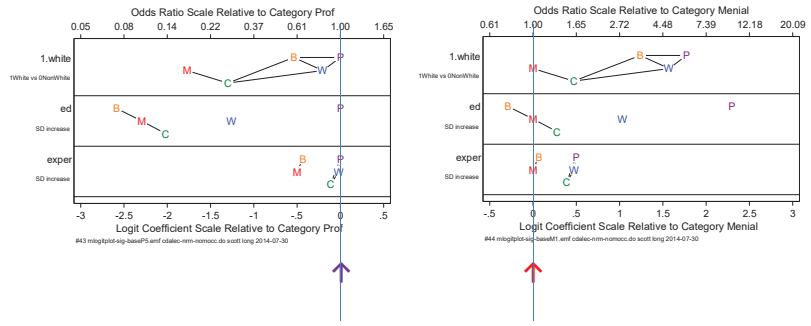
#43 OR plot for occupational attainment - base P



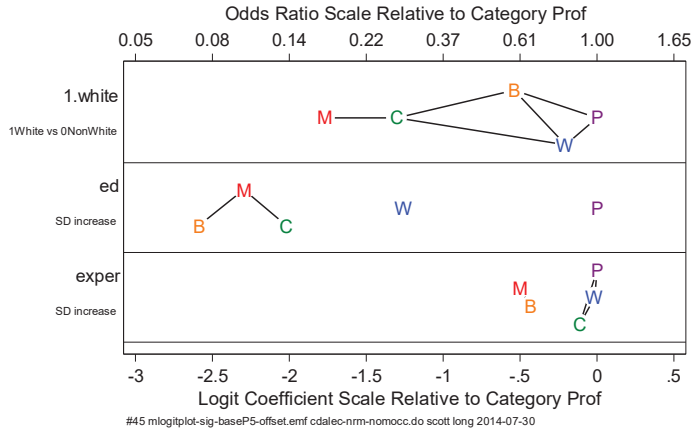
mlogitplot command discussed below…

#44 Change the base category to M



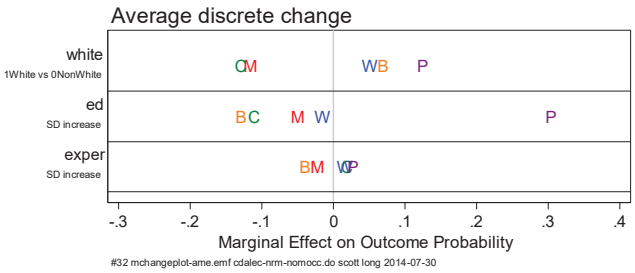


#45 Change the vertical offsets for clarity



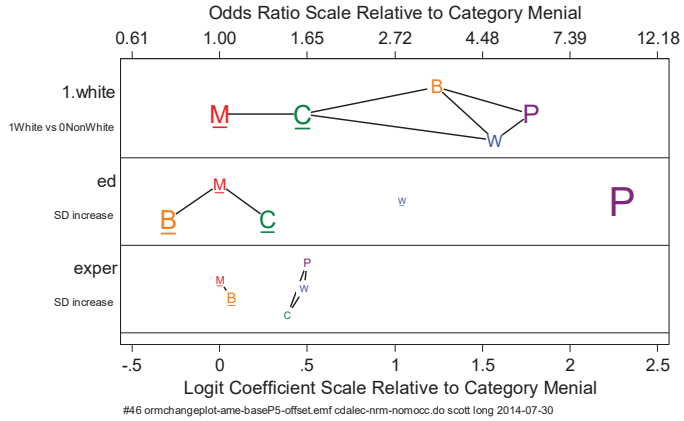
OR and change in probability

1. ORs are the same at all values variables; DC varies by the values
2. Here are the *average discrete changes* for our model



3. We add this to the OR plot
4. Size of letter in OR plot is proportional to *area of the square around letter*
5. *Underline* means a negative DC

#46 using AME for area of coefficients



mlogitplot for OR plots

```
 1] mlogitplot white ed exper,

 2]    amount(sd sd) /// show OR and DC for SD change

 3]    base(1) /// line up on category 1
 4]    offsetlist(0 3 0 -3 0    2 -2 -2 0 0    1 -1 -3 0 3 )

 5]    linep(.1) /// draw line if p>.1
 6]    linegapfactor(.5) /// blank space around letter
 7]    msizefactor(1.1) /// increase size of marker

 8]    mchange /// size of letter based on DC

 9]    min(-.5) max(2.5) ntics(7) mcol(rainbow)

10]    leftmargin(3) /// more room on left for labels
11]    aspect(.5) // aspect ratio of plot
```


Example: Attitudes toward working mothers (-nrm-ordwarm.do)

1. Attitudes toward working mothers

   Working mothers can have a warm relationship with their children?

2. Responses in **warm** are

   1=Strongly Disagree 2=Disagree 3=Agree 4=Strongly Agree

```
Working mom |
   can have |
       warm |
relations w |
     child? |      Freq.     Percent        Cum.
------------+-----------------------------------
       1_SD |        297       12.95       12.95
        2_D |        723       31.53       44.48
        3_A |        856       37.33       81.81
       4_SA |        417       18.19      100.00
------------+-----------------------------------
      Total |      2,293      100.00
```


3. Regressors

```
. codebook warm yr89 male white age ed prst, compact

Variable    Obs Unique      Mean  Min  Max  Label
-------------------------------------------------------------------------------
warm       2293      4  2.607501    1    4  Working mom can have warm...
yr89       2293      2  .3986044    0    1  Survey year: 1=1989 0=1977
male       2293      2  .4648932    0    1  Gender: 1=male 0=female
white      2293      2  .8765809    0    1  Race: 1=white 0=not white
age        2293     72  44.93546   18   89  Age in years
ed         2293     21  12.21805    0   20  Years of education
prst       2293     58  39.58526   12   82  Occupational prestige
-------------------------------------------------------------------------------
```

4. Treating **warm** as *nominal* and fit the model

```
mlogit warm i.yr89 i.male i.white age i.edcat ///
   prst, base(1) nolog
```

5. As we review the results, consider:

   Are the results consistent with warm being an ordinal variables?

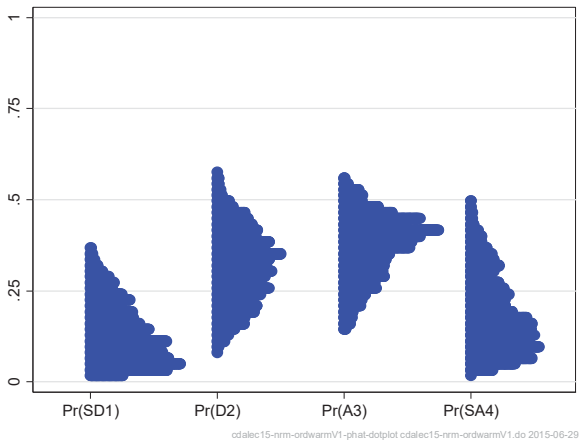
#24 Tests of regressors

```
. mlogtest, wald set(edcat_set=2.edcat 3.edcat 4.edcat)

Wald tests for independent variables (N=2293)

  Ho: All coefficients associated with given variable(s) are 0

                 |      chi2    df   P>chi2
-----------------+------------------------
          1.yr89 |    54.503     3    0.000
          1.male |   100.836     3    0.000
         1.white |     7.638     3    0.054
             age |    86.556     3    0.000
         2.edcat |     1.241     3    0.743
         3.edcat |    10.994     3    0.012
         4.edcat |    15.119     3    0.002
            prst |     6.901     3    0.075
       edcat_set |    26.063     9    0.002

   edcat_set contains: 2.edcat 3.edcat 4.edcat
```


#26 Tests for combining categories

```
. mlogtest, combine

Wald tests for combining alternatives (N=2293)

  Ho: All coefficients except intercepts associated with a given
pair
      of alternatives are 0 (i.e., alternatives can be combined)

                 |      chi2    df   P>chi2
-----------------+------------------------
      1_SD & 2_D |    38.245     8    0.000
      1_SD & 3_A |   134.132     8    0.000
     1_SD & 4_SA |   185.858     8    0.000
       2_D & 3_A |    95.727     8    0.000
      2_D & 4_SA |   172.166     8    0.000
      3_A & 4_SA |    53.660     8    0.000
```


#31 Predictions



#32 Average marginal effects -- we need a plot!

```
. mchange, amount(one sd)

mlogit: Changes in Pr(y) | Number of obs = 2293

Expression: Pr(warm), predict(outcome())
```

	1 SD	2 D	3 A	4 SA
yr89				
1989 vs 1977	-0.095	-0.029	0.079	0.045
p-value	0.000	0.136	0.000	0.006
male				
Male vs Female	0.038	0.120	-0.010	-0.148
p-value	0.006	0.000	0.622	0.000
white				
White vs NonWhite	0.040	0.018	-0.006	-0.052
p-value	0.036	0.529	0.838	0.038
age				
+1	0.002	0.004	-0.003	-0.003
p-value	0.000	0.000	0.000	0.000
+SD	0.037	0.059	-0.054	-0.042
p-value	0.000	0.000	0.000	0.000

<continued>

	1 SD	2 D	3 A	4 SA
edcat				
12 yrs vs 0-11 yrs	-0.012	-0.004	0.028	-0.012
p-value	0.535	0.880	0.305	0.580
13-15 yrs vs 0-11 yrs	-0.056	-0.042	0.053	0.045
p-value	0.013	0.171	0.104	0.084
16-20 yrs vs 0-11 yrs	-0.090	-0.015	0.050	0.055
p-value	0.000	0.694	0.198	0.093
13-15 yrs vs 12 yrs	-0.044	-0.038	0.025	0.057
p-value	0.029	0.167	0.380	0.011
16-20 yrs vs 12 yrs	-0.078	-0.011	0.022	0.066
p-value	0.000	0.745	0.510	0.017
16-20 yrs vs 13-15 yrs	-0.034	0.028	-0.003	0.009
p-value	0.096	0.415	0.934	0.746
prst				
+1	0.000	-0.002	0.001	0.001
p-value	0.818	0.011	0.193	0.220
+SD	0.002	-0.029	0.016	0.012
p-value	0.841	0.009	0.214	0.240

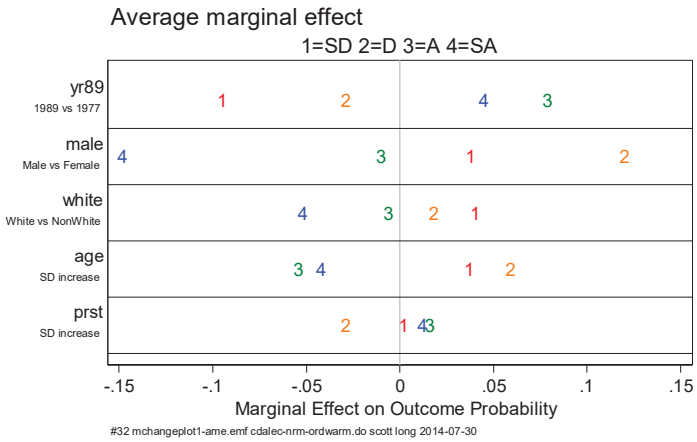
Average predictions

	1_SD	2_D	3_A	4_SA
Pr(y|base)	0.130	0.315	0.373	0.182

A plot gives us a quick summary...

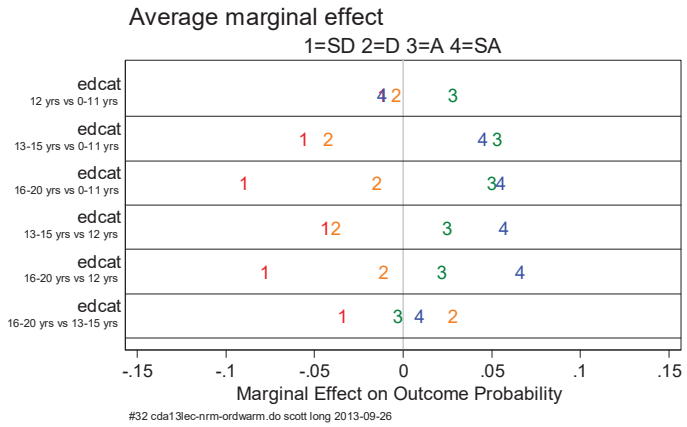
#32 Marginal effects: part 1

Why are the colors ordered red – orange – green – blue used?

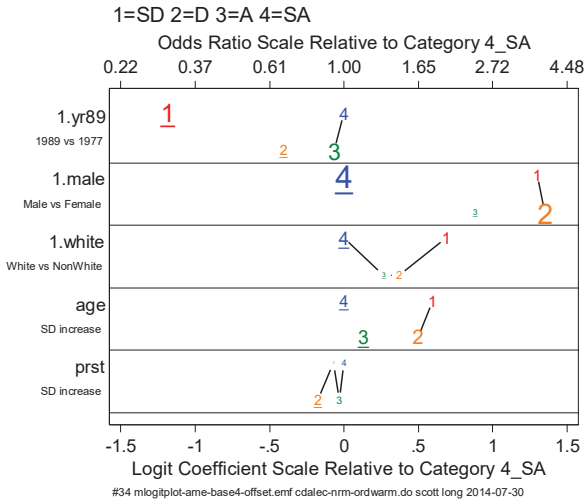


Is this consistent with warm being ordinal?

#32 Marginal effects: part 2

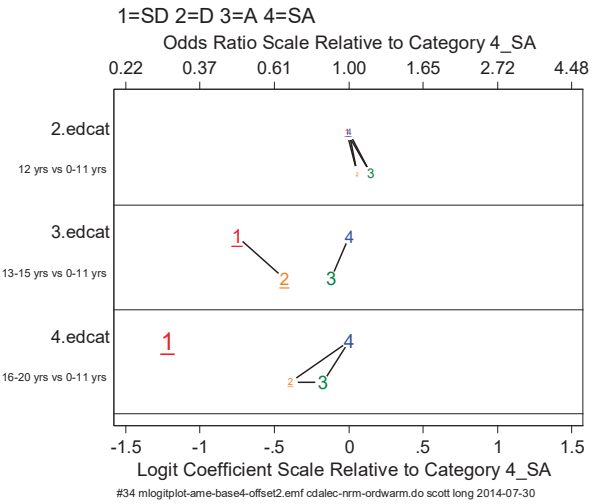


#34 OR plot with ADC: part 1



Is this consistent with warm being ordinal?

#34 OR plot with ADC: part 2



Example: Political orientation (-nrm-partyid.do)

1992 American National Election Study

#11 Outcome: party affiliation

```
. use partyid4, clear
(partyid4.dta | 1992 American National Election Study | 2014-03-12)

. tab party, miss

   Party ID |      Freq.     Percent        Cum.
------------+-----------------------------------
D    StrDem |        266       19.25       19.25
d       Dem |        427       30.90       50.14
i     Indep |        151       10.93       61.07
r       Rep |        369       26.70       87.77
R    StrRep |        169       12.23      100.00
------------+-----------------------------------
      Total |      1,382      100.00
```


#11 Regressors

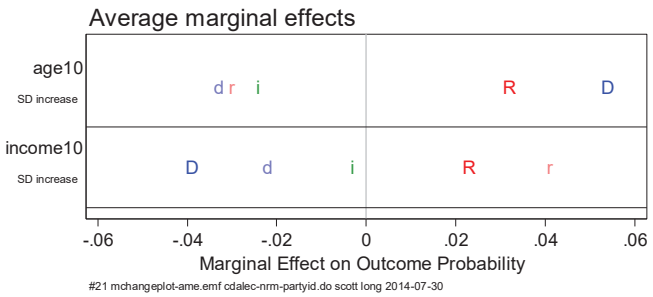
```
. nmlab party age income black female educ

party   Party ID
age     Age
income  Income in $1,000s
black   Black?
female  Female?
educ    Level of education

. sum party age income black female i.educ

  Variable |       Obs        Mean    Std. Dev.       Min        Max
-----------+--------------------------------------------------------
     party |      1382    2.817656    1.342787          1          5
       age |      1382    45.94645    16.78311         18         91
    income |      1382    37.45767    27.78148        1.5     131.25
     black |      1382    .1374819      .34448          0          1
    female |      1382    .4934877    .5001386          0          1
      educ |
  hs only  |      1382    .5803184    .4936854          0          1
  college  |      1382    .2590449    .4382689          0          1
```

"10" versions of variables divide age and income by 10.

#12 Fit MNLM and test regressors

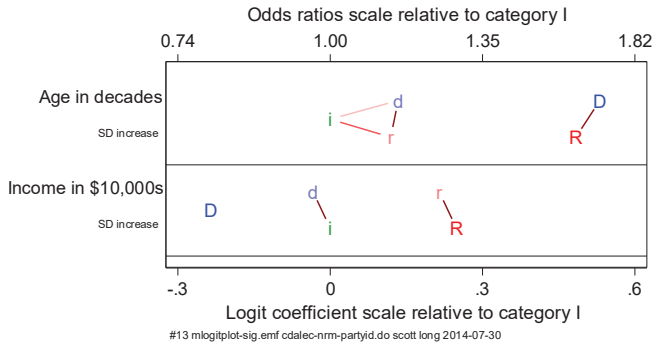
```
. mlogit party age10 income10 i.black i.female i.educ
  <snip>

. mlogtest, lr set(educ_set=1.highschool 1.college)

LR tests for independent variables (N=1382)

  Ho: All coefficients associated with given variable(s) are 0

                |      chi2    df   P>chi2
----------------+-----------------------
          age10 |    45.165     4    0.000
       income10 |    24.361     4    0.000
        1.black |   126.467     4    0.000
       1.female |     9.143     4    0.058
   1.highschool |     5.567     4    0.234
      1.college |    21.582     4    0.000
       educ_set |    26.881     8    0.001

   educ_set contains: 1.highschool 1.college
```

#21 Average marginal effects (AME)



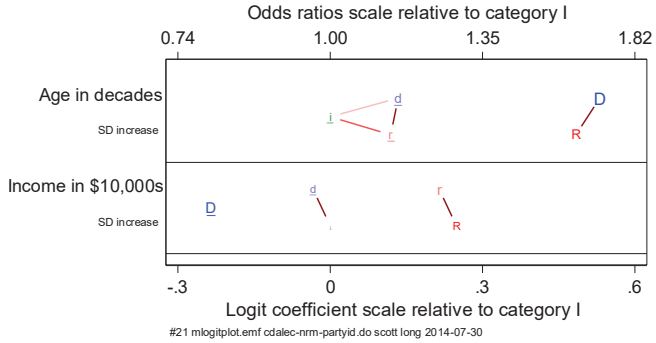
- *Age increases the probability of the extreme affiliations and decreases the probability of other affiliations.*
- *Income increase Republican affiliations, while decreasing Democratic.*

#13 Odds ratios for age and income



Question Could Δ Pr(R) / Δ Age be negative?

#21 Odds ratios for age and income: with marginal effects



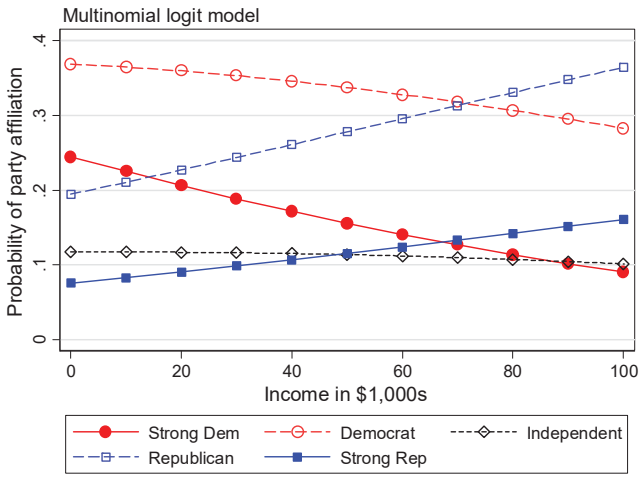
Stata code for mlogitplot

```
1]  mchange age10 income10

2]  mlogitplot age10 income10, base(3) ///
3]    linep(.1 .2 .3) lcolor(red) lshade ///
4]    symbols(D d i r R) linegapfactor(.6) ///
5]    offsetlist(2 2 0 -2 -2    0 2 -2 2 -2 ) ///
6]    titletop(Odds ratios scale relative to category I) ///
7]    titlebot(Logit coefficient scale relative to category I) ///
8]    mcol(`partycol') min(-.3) max(.6) gap(.3) ///
9]    aspect(.4) varlabels mchange
```

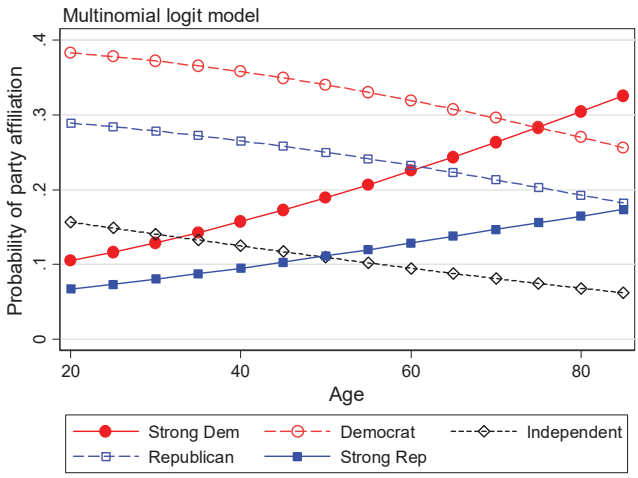

Plots of probabilities (#41--)

1. For continuous regressors, plots can be useful
   - The only way to know if it is useful is to create it!
2. One variable changes while others are held constant
   - Unless variables are linked, like age and age-squared
3. Here I create two plots holding other variables at their global means
   a. Age changes while holding other variables at their means
   b. Income changes holding other variables at their means
4. Commands are discussed after looking at plots

#42 Probabilities by income



#42 Probabilities by age



#41 Code for plots

Generate variables to plot

```
. mgen, atmeans at(age10=(2(.5)8.5)) stub(MNLMage)

Predictions from: margins, atmeans at(age10=(2(.5)8.5)) predict(outcome(5))

Variable      Obs Unique      Mean       Min       Max  Label
-------------------------------------------------------------------------------
MNLMageprr1    14     14  .2046258  .1047786  .3256354  pr(y=SD) from margins
MNLMagell1     14     14  .1675598  .0768825  .2487619  95% lower limit
MNLMageul1     14     14  .2416917  .1326748  .4025089  95% upper limit
MNLMageage10   14     14      5.25         2       8.5  Age in decades
MNLMageCpr1    14     14  .2046258  .1047786  .3256354  pr(y<=SD)
MNLMageprr2    14     14   .329003  .2561781  .3826132  pr(y=D) from margins
MNLMagell2     14     14  .2874055  .1949605  .3331204  95% lower limit
MNLMageul2     14     14  .3706006  .3173957   .435444  95% upper limit
MNLMageCpr2    14     14  .5336288  .4873918  .5818136  pr(y<=D)

<snip>

Specified values of covariates

                    1.         1.         2.         3.
 income10       black     female       educ       educ
----------------------------------------------------------
 3.745767    .1374819    .4934877    .5803184    .2590449
```



```
. mgen, atmeans at(income10=(0(1)10)) stub(MNLMinc)

Predictions from: margins, atmeans at(income10=(0(1)10)) predict(outcome(5))

Variable      Obs Unique      Mean       Min       Max  Label
-------------------------------------------------------------------------------
MNLMincprr1    11     11  .1604032   .090648  .2445128  pr(y=SD) from margins
MNLMincll1     11     11  .1253144  .0479288  .1908274  95% lower limit
MNLMincul1     11     11  .1954919  .1333672  .2981982  95% upper limit
MNLMincin~10   11     11         5         0        10  Income in $10,000s
MNLMincCpr1    11     11  .1604032   .090648  .2445128  pr(y<=SD)
MNLMincprr2    11     11  .3325221  .2827661  .3683488  pr(y=D) from margins
MNLMincll2     11     11  .2882002  .2133308  .3252223  95% lower limit
MNLMincul2     11     11  .3768441  .3522013   .420665  95% upper limit
MNLMincCpr2    11     11  .4929253  .3734141  .6128616  pr(y<=D)

<snip>

Specified values of covariates

                    1.         1.         2.         3.
    age10       black     female       educ       educ
----------------------------------------------------------
 4.594645    .1374819    .4934877    .5803184    .2590449
```

Customize variables and options for plots

```
. gen PLTincome10 = MNLMincincome10
. label var PLTincome10 "Income in $10,000s"
. gen PLTincome = MNLMincincome10*10
. label var PLTincome "Income in $1,000s"
. gen PLTage10 = MNLMageage10
. label var PLTage10 "Age in decades"
. gen PLTage = MNLMageage10*10
. label var PLTage "Age"
. local yaxis_p ///
> "ytitle(Probability of party affiliation)"
. local yaxis_p ///
> "`yaxis_p' ylab(0(.1).4, grid) ylin(0 .4, lcol(gs14))"
. local yaxis_c ///
> "ytitle(Cumulative probability) ylab(0(.2)1, grid) ylin(0 1, lcol(gs14))"
. local titleopt "position(11) size(medium)"

. * line options for probabilities
. local line5_opts "msym(O Oh dh sh s)"
. local line5_opts "`line5_opts' lwid(medium medium medium medium medium)"
. local line5_opts "`line5_opts' lpat(solid dash shortdash dash solid)"
. local line5_opts "`line5_opts' mcol(red red*.8 black blue*.8 blue)"
. local line5_opts "`line5_opts' lcol(red red*.8 black blue*.8 blue)"

. label var MNLMagepr1 "Strong Dem"
. label var MNLMagepr2 "Democrat"
. label var MNLMagepr3 "Independent"
. label var MNLMagepr4 "Republican"
. label var MNLMagepr5 "Strong Rep"
```



```
. label var MNLMincpr1 "Strong Dem"
. label var MNLMincpr2 "Democrat"
. label var MNLMincpr3 "Independent"
. label var MNLMincpr4 "Republican"
. label var MNLMincpr5 "Strong Rep"
```

Create plots

```
. * probability by age
. graph twoway (connected MNLMagepr1 MNLMagepr2 MNLMagepr3 ///
>     MNLMagepr4 MNLMagepr5 PLTage, `line5_opts'), ///
>     title("`title'", `titleopt') `yaxis_p' `xaxis_age' ///
>     legend(rows(2)) // caption("`tag'",size(vsmall))

. * probability by income
. graph twoway (connected MNLMincpr1 MNLMincpr2 MNLMincpr3 ///
>     MNLMincpr4 MNLMincpr5 PLTincome, `line5_opts'), ///
>     title("`title'", `titleopt') `yaxis_p' `xaxis_inc' ///
>     legend(rows(2)) // caption("`tag'",size(vsmall))
```


Independence of irrelevant alternatives (IIA)

1. Fundamental assumption of MNLM and CLM is IIA
   - Independence of Irrelevant Alternatives
2. In these models the odds of two outcomes do *not* depend on other outcomes
   - The odds of A vs B does not depend on what other options there are

Why IIA fails? McFadden's buses

1. A person has two choices

$$\Pr(\text{car}) = 1/2 \quad \text{and} \quad \Pr(\text{red bus}) = 1/2$$

2. The odds of taking the car versus a red bus

$$\frac{\Pr(\text{car})}{\Pr(\text{red bus})} = \frac{1/2}{1/2} = 1$$

3. A new bus company opens with identical service & blue buses
4. IIA *requires*

$$\Pr(\text{car}) = 1/3; \quad \Pr(\text{red bus}) = 1/3; \quad \Pr(\text{blue bus}) = 1/3$$

5. This is necessary so that the odds remain constant

$$\frac{\Pr(\text{car})}{\Pr(\text{red bus})} = 1 = \frac{1/3}{1/3}$$

6. Substantively, we would expect a violation of IIA

$$\Pr(\text{car}) = 1/2; \quad \Pr(\text{red bus}) = 1/4; \quad \Pr(\text{blue bus}) = 1/4$$

IIA, mlogit and logit

```
mlogit party age, base(5) nolog vsquish

------------------------------------------------------------------------------
       party |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
StrDem       |
         age |   .0104532   .0056788     1.84   0.066    -.0006771    .0215834
       _cons |   -.072338   .3003037    -0.24   0.810    -.6609224    .5162464
:::
-------------+----------------------------------------------------------------
StrRep       |  (base outcome)
------------------------------------------------------------------------------

. gen party15 = 1 if party==1
. replace party15 = 0 if party==5

. logit party15 age, nolog

------------------------------------------------------------------------------
     party15 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0102276    .005633     1.82   0.069    -.0008129    .0212681
       _cons |  -.0609598   .2981109    -0.20   0.838    -.6452465    .5233269
------------------------------------------------------------------------------
```


Summary of IIA

1. IIA requires that if a new choice becomes available, probabilities for prior choices adjust *precisely* to retain the original odds among choices
2. **McFadden** suggested that IIA implies that MNLM and CLM should only be used when *outcome categories can plausibly be assumed to be distinct and weighed independently in the eyes of each decision maker*
3. **Amemiya** suggested that the MNLM works well when the *alternatives are dissimilar*
4. Simulations by Cheng and Long (SMR) and other studies found that formal tests do not work well
5. Care in specifying the model to involve *distinct outcomes that are not substitutes for one another* seems to be reasonable, albeit ambiguous, advice
6. But, reviewers sometimes demand to see an IIA test...

Formal tests of IIA

1. Hausman and McFadden proposed a Hausman-type test of IIA
   a. This compares two estimates of the same parameters
      i. One estimate is *consistent and efficient* if the H_0 is true
      ii. The second estimate is *consistent but inefficient*
   b. Cheng and Long (2006) find the Hausman-McFadden test has very poor statistical properties
2. McFadden, Tye, and Train (1977) and Small and Hsiao (1985) proposed a LR type test. The Small-Hsiao test works better, but not always
3. We found no test that works well in all cases

#50 "Testing" IIA with `mlogtest` *(-nrm-partyid.do)*

Hausman test

```
. mlogtest, hausman

Hausman tests of IIA assumption (N=1382)

  Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives

                |      chi2    df   P>chi2
----------------+-----------------------
          StrDem |     4.622    20    1.000
             Dem |     0.919    21    1.000
           Indep |    -2.244    19        .
             Rep |     3.030    21    1.000
          StRep |    -0.580    21        .

  Note: A significant test is evidence against Ho.
```


Small-Hsiao: Seed 124386

```
. mlogtest, smhsiao

Small-Hsiao tests of IIA assumption (N=1382)

  Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives

                | lnL(full)  lnL(omit)       chi2    df   P>chi2
----------------+-------------------------------------------------
          StrDem |  -696.753   -690.654     12.198    21    0.934
             Dem |  -565.571   -557.488     16.166    21    0.760
           Indep |  -764.563   -758.290     12.547    21    0.924
             Rep |  -621.562   -615.492     12.140    21    0.936
          StRep |  -761.598   -752.804     17.587    21    0.675

  Note: A significant test is evidence against Ho.
```

Seed 254331

```
. mlogtest, smhsiao

Small-Hsiao tests of IIA assumption (N=1382)

  Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives

                | lnL(full)  lnL(omit)       chi2    df   P>chi2
----------------+-----------------------------------------------
         StrDem |  -727.367   -692.048     70.639    21    0.000
            Dem |  -610.636   -573.268     74.736    21    0.000
          Indep |  -783.456   -747.654     71.604    21    0.000
            Rep |  -650.962   -615.434     71.057    21    0.000
         StrRep |  -751.887   -740.193     23.388    21    0.324

  Note: A significant test is evidence against Ho.
```


Review of nominal LHS

1. The MNLM is a set of BLMs for all pairs of outcomes
2. MNLM can be *overwhelmingly complex* if you try to absorb all of the coefficients individually
3. Plots of coefficients make it *trivially easy* to uncover patterns
   - Use the plots to find patterns, but you might not want to use them in papers or presentations
4. IIA is a restrictive assumption that does not have an adequate test. If outcomes are reasonably distinct, MNLM works well. No good alternative is readily available

Part 11: Ordinal outcomes

Read and run

Long & Freese	Chapter
cdalec*.do	cdalec17-orm-ordwarm-.do; cdalec17-orm-partyid-.do

Overview

1. What does ordinal mean? What is an ordinal regression model?
2. Derive *the* ordinal regression model (ORM) as
   a. A latent variable model
   b. A nonlinear probability model
3. Apply methods of interpretation from the BRM and MNLM
4. How do you decide if an odinal model is appropriates

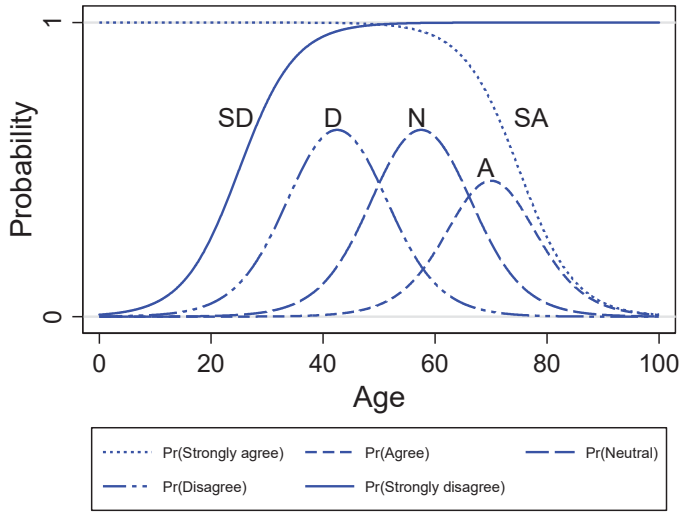
What does ordinal mean?

1. An *ordinal variable* has
   a. Categories *ordered on a single dimension*
   b. Unknown distances between categories

2. An *ordinal model* is defined by the nature of the *relationship* between regressors and outcomes (Anderson 1980)
3. Consider the statement

   A working mother can establish just as warm and secure of a relationship with her child as a mother who does not work.

   with answers Strongly agree, Agree, Neutral, Disagree, Strongly disagree
4. Anderson argues that the following pattern must be found for a model to be ordinal...



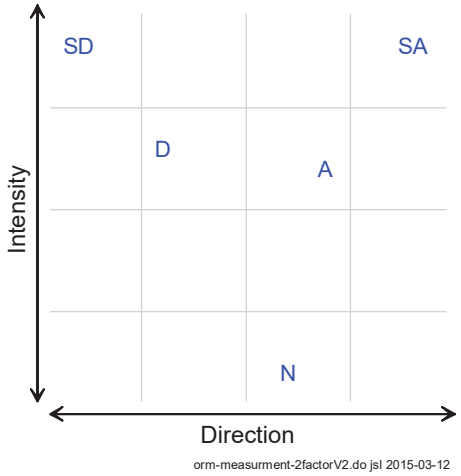
orm-anderson-ordinalityV2.do 2015-06-10

Is the outcome ordered?

1. Because values *can be ordered* does not mean it is *substantively meaningful to order categories*
2. A variable can be *partially ordered* such as SA, A, D, SD, Don't know.
3. A can be ordered on *multiple dimensions*
   a. Miller and Volker ordered occupational groupings both by status and by income of the occupations leading to different conclusions
   b. Likert scales could reflect two dimensions, not one

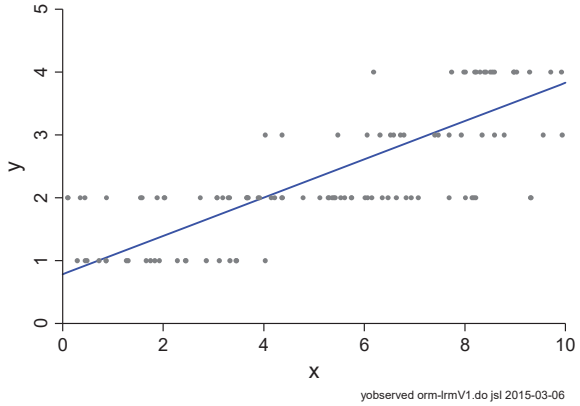
For example...

Multiple ordering of a Likert scale



Ordinal is not interval

1. Researchers often treat ordinal variables (e.g., summated scale) as interval.
2. Results can be misleading in terms of magnitude and significance.



A latent variable model for ordinal outcomes

Structural model relating x's to y*'s

Same as BRM

$$y^* = \mathbf{x}\boldsymbol{\beta} + \varepsilon$$

Measurement model linking y* to y

Observed y obtained from y* by dividing it into segments by thresholds τ_q

$$y_i = q \quad \text{if } \tau_{q-1} \le y_i^* < \tau_q \quad \text{for } q = 1 \text{ to } J$$

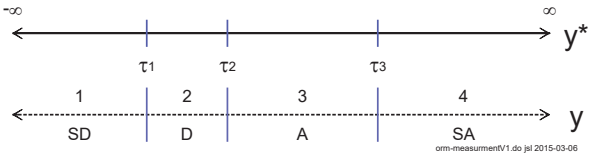
For example...

Attitudes toward working mothers

1. Evaluate the statement

   A working mother can establish just as warm and secure of a relationship with her child as a mother who does not work.

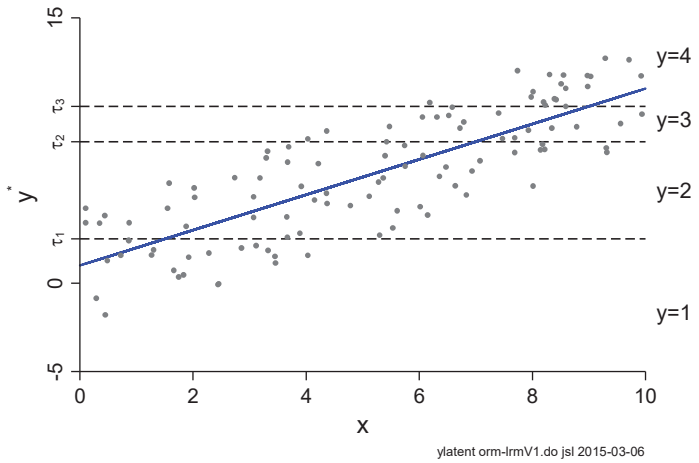
2. Graphically,



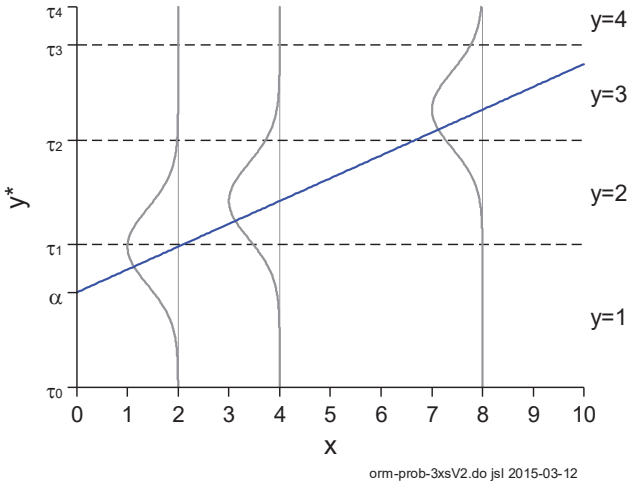
3. Mathematically,

$$y_i = \begin{cases} 1 \Rightarrow \text{SD-Strongly Disagree} & \text{if } \tau_0 = -\infty \le y_i^* < \tau_1 \\ 2 \Rightarrow \text{D-Disagree} & \text{if } \tau_1 \le y_i^* < \tau_2 \\ 3 \Rightarrow \text{A-Agree} & \text{if } \tau_2 \le y_i^* < \tau_3 \\ 4 \Rightarrow \text{SA-Strongly Agree} & \text{if } \tau_3 \le y_i^* < \tau_4 = \infty \end{cases}$$

If we observed y*



Since y* is latent we focus on Pr(y=m|**x**)

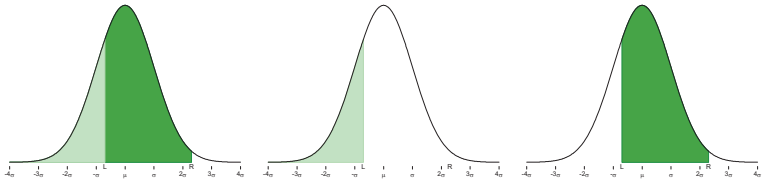


Pr(*y*=m | x) is the area between τ_{m-1} and τ_m

- Ordered probit assumes: $\varepsilon \sim N(0,1)$
- Ordered logit assumes: $\varepsilon \sim \lambda\left(0, \pi^2/3\right)$

Tool: Computing areas between thresholds

Overview

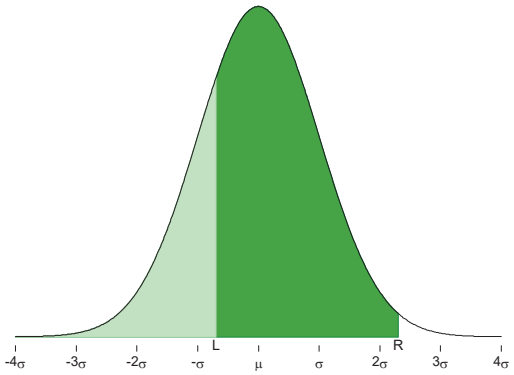


Step 1. Area ≤ R **Step 2**. Area ≤ L **Step 3**. Area between L & R

Details follow...

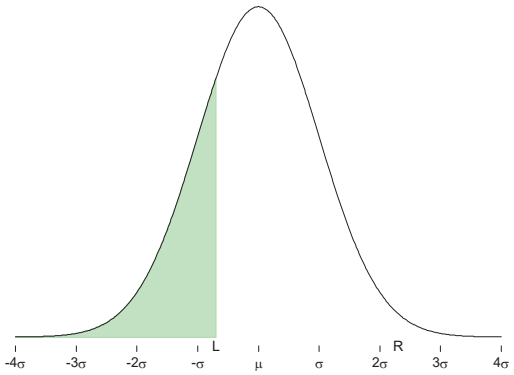
Step 1. Compute area less than R.

CDF(R) is everything to the left of R.



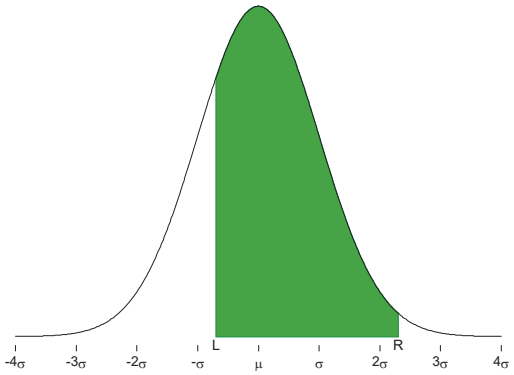
Step 2. Compute area less than L.

CDF(L) is everything to the left of L.



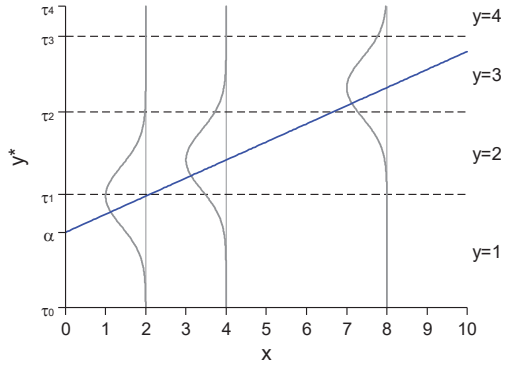
Step 3. Compute area between L and R.

CDF(R) - CDF(L).



The probabilities of observed values

1. Ordered probit assumes: $\varepsilon \sim N(0,1)$
2. Ordered logit assumes: $\varepsilon \sim \lambda\left(0,\pi^2/3\right)$
3. Pr(*y*=m | x) is the area between τ_{m-1} and τ_m:



4. For a given outcome *m*

$$\Pr(y=m|\mathbf{x}) = F(\tau_m - \mathbf{x}\boldsymbol{\beta}) - F(\tau_{m-1} - \mathbf{x}\boldsymbol{\beta})$$

where CDF(-∞) = 0 and CDF(∞) = 1

5. With four outcomes

$$\begin{aligned}
\Pr(y_i=1|x_i) &= \Phi(\tau_1 - \alpha - \beta x_i) - \Phi(\tau_0 - \alpha - \beta x_i) \\
&= \Phi(\tau_1 - \alpha - \beta x_i) - \Phi(-\infty - \alpha - \beta x_i) \\
&= \Phi(\tau_1 - \alpha - \beta x_i) - 0 \\
\Pr(y_i=2|x_i) &= \Phi(\tau_2 - \alpha - \beta x_i) - \Phi(\tau_1 - \alpha - \beta x_i) \\
\Pr(y_i=3|x_i) &= \Phi(\tau_3 - \alpha - \beta x_i) - \Phi(\tau_2 - \alpha - \beta x_i) \\
\Pr(y_i=4|x_i) &= \Phi(\tau_4 - \alpha - \beta x_i) - \Phi(\tau_3 - \alpha - \beta x_i) \\
&= \Phi(\infty - \alpha - \beta x_i) - \Phi(\tau_3 - \alpha - \beta x_i) \\
&= 1 - \Phi(\tau_3 - \alpha - \beta x_i)
\end{aligned}$$

Identification

1. ORM is not identified without additional assumptions
2. Since *y** is latent, the variance of *y** cannot be estimated
   a. For logit we assume: $\text{Var}(\varepsilon|\mathbf{x})=\pi^2/3$
   b. For probit we assume: $\text{Var}(\varepsilon|\mathbf{x})=1$
3. We cannot estimate all of the thresholds and the intercept
4. Consider the thresholds τ_m and the structural model

$$y^* = \alpha + \beta x + \varepsilon$$

5. Let α and the τ_ms be the *true parameters*
6. Pick any δ and create *imposter parameters*

$$\alpha^* = \alpha - \delta \qquad \text{and} \qquad \tau_q^* = \tau_q - \delta$$

7. Probabilities are the same for all values of δ

9. Adding 0 = δ - δ

$$\begin{aligned}
\Pr(y=q\,|\,x) &= F(\tau_q - \alpha - \beta x) - F(\tau_{q-1} - \alpha - \beta x) \\
&= F(\tau_q - \alpha - \beta x + [\delta - \delta]) - F(\tau_{q-1} - \alpha - \beta x + [\delta - \delta]) \\
&= F([\,\tau_q - \delta\,] - [\,\alpha - \delta\,] - \beta x) - F([\,\tau_{q-1} - \delta\,] - [\,\alpha - \delta\,] - \beta x) \\
&= F(\tau_q^* - \alpha^* - \beta x) - F(\tau_{q-1}^* - \alpha^* - \beta x)
\end{aligned}$$

10. Two sets of identifying assumptions are often used
    Alternative 1: $\tau_1 = 0$ (so that δ must equal τ_1)
    Alternative 2: $\alpha = 0$ (so that δ must equal α)
11. These alternative lead to different *parameterizations*
12. The parameterization
    a. Does not affect the βs
    b. Does not affect the significance tests of the βs
    c. Does not affect the probabilities of observed outcomes

13. Different programs use different parameterizations:

	OLM1: beta0=0		OLM2: tau1=0	
	b	z	b	z
yr89	0.524	6.557	0.524	6.557
male	-0.733	-9.343	-0.733	-9.343
white	-0.391	-3.304	-0.391	-3.304
age	-0.022	-8.778	-0.022	-8.778
ed	0.067	4.205	0.067	4.205
prst	0.006	1.844	0.006	1.844
tau1	-2.465	-10.319	.	.
tau2	-0.631	-2.704	1.834	29.098
tau3	1.262	5.392	1.893	32.537
beta0	0.000	0.000	2.465	10.319
2lnL	-2844.912	0.000	-2844.912	0.000
LRchi2	301.716	0.000	301.716	0.000

```
. display -.631 - -2.465
1.834
```

ML estimation

1. The probability of observed outcome *q* for case *i*

$$p_i = \Pr(y_i = q \mid \mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\tau}) = F(\tau_q - \mathbf{x}_i\boldsymbol{\beta}) - F(\tau_{q-1} - \mathbf{x}_i\boldsymbol{\beta})$$

2. If the observations are independent, then

$$L(\boldsymbol{\beta}, \boldsymbol{\tau} \mid \mathbf{y}, \mathbf{X}) = \prod_{i=1}^{N} p_i$$

Software issues

1. You must know which parameterization is used
2. Different methods of maximization produce *slightly* different test statistics
3. ORM takes longer to converge than some models
4. Small N's in a category can lead to failure to converge
   - You can merge adjacent categories and only lose efficiency

Example: Attitudes toward working mothers (-orm-warm.do)

A working mother can establish just as warm and secure a relationship with her child as a mother who does not work.

#12 Descriptive statistics

```
Working mom |
   can have |
       warm |
relations w |
     child? |      Freq.     Percent        Cum.
------------+-----------------------------------
        1SD |        297       12.95       12.95
         2D |        723       31.53       44.48
         3A |        856       37.33       81.81
        4SA |        417       18.19      100.00
------------+-----------------------------------
      Total |      2,293      100.00
```



```
. nmlab warm yr89 male white age ed prst

warm   Working mom can have warm relations w child?
yr89   Survey year: 1=1989 0=1977
male   Gender: 1=male 0=female
white  Race: 1=white 0=not white
age    Age in years
ed     Years of education
prst   Occupational prestige

. sum warm yr89 male white age ed prst

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        warm |      2293    2.607501    .9282156          1          4
        yr89 |      2293    .3986044    .4897178          0          1
        male |      2293    .4648932    .4988748          0          1
       white |      2293    .8765809    .3289894          0          1
         age |      2293    44.93546    16.77903         18         89
          ed |      2293    12.21805    3.160827          0         20
        prst |      2293    39.58526    14.49226         12         82
```

#13 Ordinal logit and probit

```
. ologit warm i.yr89 i.male i.white age ed prst, nolog
Ordered logistic regression                     Number of obs   =       2293
                                                LR chi2(6)      =     301.72
                                                Prob > chi2     =     0.0000
Log likelihood = -2844.9123                     Pseudo R2       =     0.0504
------------------------------------------------------------------------------
        warm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        yr89 |
       1989  |   .5239025   .0798989     6.56   0.000     .3673036    .6805014
        male |
       Male  |  -.7332997   .0784827    -9.34   0.000     -.887123   -.5794765
       white |
       White |  -.3911595   .1183808    -3.30   0.001    -.6231816   -.1591373
         age |  -.0216655   .0024683    -8.78   0.000    -.0265032   -.0168278
          ed |   .0671728    .015975     4.20   0.000     .0358624    .0984831
        prst |   .0060727   .0032929     1.84   0.065    -.0003813    .0125267
-------------+----------------------------------------------------------------
       /cut1 |  -2.465362   .2389128                     -2.933622   -1.997102
       /cut2 |   -.630904   .2333156                     -1.088194   -.1736138
       /cut3 |   1.261854    .234018                      .8031871    1.720521
------------------------------------------------------------------------------
. estimates store olm
. oprobit warm yr89 male white age ed prst, nolog
  <snip>
. estimates store opm
```


#13 Comparing OLM and OPM: ratios on next page

	olm		opm	
	b	z	b	z
warm				
yr89	0.524	6.557	0.319	6.805
male	-0.733	-9.343	-0.417	-9.156
white	-0.391	-3.304	-0.227	-3.260
age	-0.022	-8.778	-0.012	-8.471
ed	0.067	4.205	0.039	4.153
prst	0.006	1.844	0.003	1.705
cut1				
_cons	-2.465	-10.319	-1.429	-10.294
cut2				
_cons	-0.631	-2.704	-0.361	-2.633
cut3				
_cons	1.262	5.392	0.768	5.605
aux				
N	2293.000	.	2293.000	.
LRchi2	301.716	.	294.319	.
BIC	5759.463	.	5766.861	.

Looking at the ratios...

	ratio olm to opm	
	b	z
warm		
yr89	1.643	0.964
male	1.758	1.020
white	1.727	1.014
age	1.773	1.036
ed	1.735	1.012
prst	1.850	1.081
cut1		
_cons	1.726	1.002
cut2		
_cons	1.750	1.027
cut3		
_cons	1.643	0.962
aux		
N	1.000	.
LRchi2	1.025	.
BIC	0.999	.

Interpretation with marginal change in y*

1. LRM is often used with an ordinal outcome which can be *misleading*.
2. If you are considering the LRM for ordinal outcomes, the ORM is a better alternative

$$y^* = \mathbf{x}\boldsymbol{\beta} + \varepsilon = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + \varepsilon$$

3. The scale of y* cannot be estimated, so we use *fully standardized* or *y*-standardized* coefficients
4. We estimate variance of $Var(\hat{y}^*)$ using the assumed variance of the error

$$\begin{aligned} Var\left(\hat{y}^*\right) &= Var\left(\hat{\beta}x + e\right) \\ &= Var\left(\hat{\beta}x\right) + Var\left(e\right) + 2Cov\left(\hat{\beta}x, e\right) \\ &= \hat{\beta}^2 Var\left(x\right) + Var\left(e\right) + 0 \quad \text{where } Var\left(e\right) \text{ is assumed.} \end{aligned}$$

5. Generalizing

$$\hat{\sigma}^2_{y^*} = \hat{\boldsymbol{\beta}}' \mathrm{Var}\left(\mathbf{x}\right)\hat{\boldsymbol{\beta}} + \mathrm{Var}\left(\varepsilon\right)$$

6. The *y* standardized coefficient*

$$\beta_k^{Sy^*} = \frac{\beta_k}{\sigma_{y^*}}$$

   - For a unit increase in x_k, y^* is expected to increase by $\beta_k^{Sy^*}$ standard deviations holding other variables constant.

7. The *fully standardized coefficient* is:

$$\beta_k^S = \frac{\sigma_k \beta_k}{\sigma_{y^*}} = \sigma_k \beta_k^{Sy^*}$$

   - For a standard deviation increase in x_k, y^* is expected to increase by β_k^S standard deviations, holding other variables constant.

#14 y standardized coefficients*

```
. ologit warm i.yr89 i.male i.white age ed prst
. listcoef, help std
ologit (N=2293): Unstandardized and standardized estimates
  Observed SD:  0.9282
    Latent SD:  1.9411
          |          b        z    P>|z|    bStdX    bStdY   bStdXY     SDofX
----------+------------------------------------------------------------------
     yr89 |     0.5239    6.557    0.000    0.257    0.270    0.132     0.490
     male |    -0.7333   -9.343    0.000   -0.366   -0.378   -0.188     0.499
    white |    -0.3912   -3.304    0.001   -0.129   -0.202   -0.066     0.329
      age |    -0.0217   -8.778    0.000   -0.364   -0.011   -0.187    16.779
       ed |     0.0672    4.205    0.000    0.212    0.035    0.109     3.161
     prst |     0.0061    1.844    0.065    0.088    0.003    0.045    14.492
------------------------------------------------------------------------------
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
   bStdX = x-standardized coefficient
   bStdY = y-standardized coefficient
  bStdXY = fully standardized coefficient
   SDofX = standard deviation of X. listcoef, help std
```

Interpretations follow...

1. In 1989 support was .27 standard deviations higher than in 1977, holding other variables constant.

```
             b         z     P>|z|     bStdX     bStdY    bStdXY     SDofX
1.yr89 |   0.52390     6.557   0.000    0.2566    0.2699    0.1322    0.4897
```

2. Each additional year of age decreases support by .01 standard deviations, holding other variables constant. Alternatively, each additional ten years of age decreases support by .11 standard deviations (=10x-.011), holding other variables constant.

```
             b         z     P>|z|     bStdX     bStdY    bStdXY     SDofX
age |  -0.02167    -8.778   0.000   -0.3635   -0.0112   -0.1873   16.7790
```

3. Each standard deviation increase in education increases support by .11 standard deviations, holding other variables constant.

```
             b         z     P>|z|     bStdX     bStdY    bStdXY     SDofX
ed |   0.06717     4.205   0.000    0.2123    0.0346    0.1094    3.1608
```

4. We are holding other variables *constant*, but *not constant at specific values*.

#14 Comparing OLM and LRM

```
. regress warm i.yr89 i.male i.white age ed prst
<snip>
. listcoef
<snip>
```

Extracted output

```
         | b                  | z                  | bstdy              | bstd
         |    lrm      olm    |    lrm      olm    |    lrm      olm    |    lrm      olm
---------+--------------------+--------------------+--------------------+--------------------
  1.yr89 |  0.262    0.524    |  6.944    6.557    |  0.283    0.270    |  0.138    0.132
  1.male | -0.336   -0.733    | -9.171   -9.343    | -0.362   -0.378    | -0.180   -0.188
 1.white | -0.177   -0.391    | -3.166   -3.304    | -0.191   -0.202    | -0.063   -0.066
     age | -0.010   -0.022    | -8.699   -8.778    | -0.011   -0.011    | -0.183   -0.187
      ed |  0.031    0.067    |  4.143    4.205    |  0.034    0.035    |  0.106    0.109
    prst |  0.003    0.006    |  1.734    1.844    |  0.003    0.003    |  0.042    0.045
```

Difference between LRM - OLM

```
             |       b         t     bStdY    bStdXY
-------------+-----------------------------------------
      1.yr89 |   -0.261     0.387     0.013     0.006
      1.male |    0.398     0.173     0.016     0.008
     1.white |    0.214     0.139     0.011     0.004
         age |    0.012     0.078     0.000     0.005
          ed |   -0.036    -0.062    -0.001    -0.003
        prst |   -0.003    -0.111    -0.000    -0.003
```

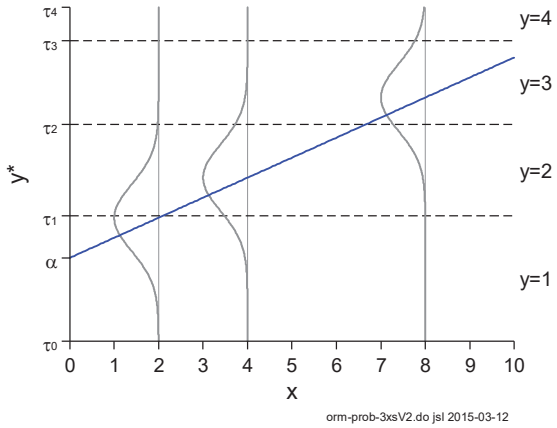

#15 Standardized coefficients for OLM and OPM

```
           | olm                          | opm
           |       b    bstdy     bstd    |       b    bstdy      bstd
-----------+------------------------------+------------------------------
    1.yr89 |   0.524    0.270    0.132    |    0.319    0.296     0.145
    1.male |  -0.733   -0.378   -0.188    |   -0.417   -0.388    -0.193
  1.white  |  -0.391   -0.202   -0.066    |   -0.227   -0.210    -0.069
       age |  -0.022   -0.011   -0.187    |   -0.012   -0.011    -0.191
        ed |   0.067    0.035    0.109    |    0.039    0.036     0.114
      prst |   0.006    0.003    0.045    |    0.003    0.003     0.044
```

1. The unstandardized OLM coefficients are larger than those for OPM
   - This is due to the larger assumed variance of the errors
2. Standardized coefficients are similar
3. Why are they similar but not *exactly* the same?

Predicted probabilities

$$\Pr(y=q\,|\,\mathbf{x}) = F\left(\hat{\tau}_q - \mathbf{x}\hat{\boldsymbol{\beta}}\right) - F\left(\hat{\tau}_{q-1} - \mathbf{x}\hat{\boldsymbol{\beta}}\right)$$



Ways to examine predictions

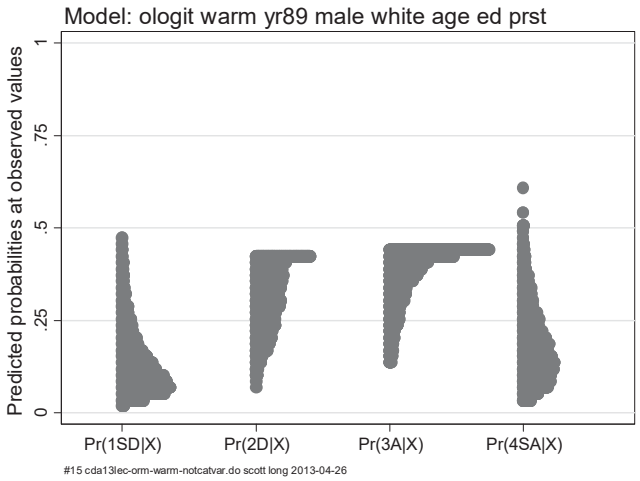
1. Predictions at observed values
2. Discrete change
3. Ideal types
4. Tables
5. Plots

Predictions at observed values

#21 Compute predicted probabilities

```
. predict OLMpr1sd OLMpr2d OLMpr3a OLMpr4sa
(option pr assumed; predicted probabilities)
. label var OLMpr1sd "Pr(1SD|X)"
. label var OLMpr2d  "Pr(2D|X)"
. label var OLMpr3a  "Pr(3A|X)"
. label var OLMpr4sa "Pr(4SA|X)"

. sum OLMpr1sd OLMpr2d OLMpr3a OLMpr4sa

  Variable |       Obs        Mean    Std. Dev.       Min        Max
-----------+--------------------------------------------------------
  OLMpr1sd |      2293    .1291898    .0827858   .0078648   .4583639
  OLMpr2d  |      2293    .3152269    .0702155   .0811076   .4066651
  OLMpr3a  |      2293    .3740882    .0585058   .1494467   .4274917
  OLMpr4sa |      2293    .1814951    .0976008    .018211   .5864362
```

#23 Dotplot of predictions



Tables of predicted probabilities

1. When you have important categorical regressors, tables can be effective
2. An important consideration is where to hold other variables
   a. *Global means* are simpler but can be misleading
   b. *Local means* are harder but sometimes more realistic
3. A *sensitivity analysis* is needed to determine how important this decision is

#30 Gender and year

1. We expect
   a. Men are more negative toward working women
   b. Attitudes are more positive towards working women in 1989
2. Estimates show year and gender are important (see #11):

warm	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.yr89	.5239025	.0798989	6.56	0.000	.3673036	.6805014
1.male	-.7332997	.0784827	-9.34	0.000	-.887123	-.5794765
1.white	-.3911595	.1183808	-3.30	0.001	-.6231816	-.1591373
age	-.0216655	.0024683	-8.78	0.000	-.0265032	-.0168278
ed	.0671728	.015975	4.20	0.000	.0358624	.0984831
prst	.0060727	.0032929	1.84	0.065	-.0003813	.0125267

#32 Probabilities by year and gender using global means

A. 1977	SD	D	A	SA
Men	**0.19**	0.40	0.32	0.10
Women	**0.10**	0.31	0.41	0.18
Men-Women	**0.09**	*0.09*	*-0.09*	*-0.08*
B. 1989	SD	D	A	SA
Men	0.12	0.34	0.39	0.15
Women	0.06	0.23	0.44	0.27
Men-Women	0.06	0.11	-0.05	-0.12
C. 1977 to 1989	SD	D	A	SA
Men	-0.07	-0.06	0.07	0.05
Women	-0.04	-0.08	0.03	0.09

Note: Other variables are held at their means.

Stata code is given below...

#31 mtable with global means

```
. mtable, at(yr89=(0 1) male=(0 1)) atmeans clear

Expression: Pr(warm), predict(outcome())

           |     yr89      male      1_SD       2_D       3_A      4_SA
 ----------+------------------------------------------------------------
         1 |        0         0     0.099     0.308     0.413     0.180
         2 |        0         1     0.186     0.403     0.316     0.095
         3 |        1         0     0.061     0.228     0.441     0.270
         4 |        1         1     0.119     0.339     0.390     0.151

Specified values of covariates

           |       1.
           |    white       age        ed      prst
 ----------+-------------------------------------
   Current |     .877      44.9      12.2      39.6
```

1. **`at(yr89=(0 1) male=(0 1))`** computes predictions for (**`yr89`**, **`male`**) combinations
2. **`atmeans`** holds other variable at global means

3. Since **`i.male`** was a regressor, **`dydx(male)`** computes DC for **`male`** at the two values of **`yr89`**

```
. mtable, dydx(male) at(yr89=(0 1)) atvars(1.yr89 1.male) atmeans

Expression: Marginal effect of Pr(warm), predict(outcome())

           |       1.        1.
           |     yr89      male      1 SD       2 D       3 A      4 SA
 ----------+------------------------------------------------------------
         1 |        0      .465     0.087     0.094    -0.097    -0.085
         2 |        1      .465     0.058     0.111    -0.050    -0.119

Specified values of covariates

           |       1.        1.
           |     male     white       age        ed      prst
 ----------+-----------------------------------------------
   Current |     .465      .877      44.9      12.2      39.6
```

4. **mchange** computes the same values

- Until I am sure I know what I'm doing, I compute things two ways

```
. mchange male, at(yr89=0) atmeans brief

ologit: Changes in Pr(y) | Number of obs = 2293

Expression: Pr(warm), predict(outcome())

               |      1 SD        2 D        3 A       4 SA
---------------+-----------------------------------------------
male           |
 Male vs Female |     0.087      0.094     -0.097     -0.085
        p-value |     0.000      0.000      0.000      0.000

. mchange male, at(yr89=1) atmeans brief

ologit: Changes in Pr(y) | Number of obs = 2293

Expression: Pr(warm), predict(outcome())

               |      1 SD        2 D        3 A       4 SA
---------------+-----------------------------------------------
male           |
 Male vs Female |     0.058      0.111     -0.050     -0.119
        p-value |     0.000      0.000      0.000      0.000
```


* #32 building a table with global means

Creating tables is tedious, but **mtable** can help

```
* predictions for 1977
mtable, at(yr89=0 male=1)     atmeans rowname(Men)       clear roweqnm(1977)
mtable, at(yr89=0 male=0)     atmeans rowname(Women)     below roweqnm(1977)
mtable, dydx(male) at(yr89=0) atmeans rowname(Men_Women) below roweqnm(1977)

           |     1 SD       2 D       3 A      4 SA
 ----------+-------------------------------------
 1977      |
       Men |    0.186     0.403     0.316     0.095
     Women |    0.099     0.308     0.413     0.180
 Men_Women |    0.087     0.094    -0.097    -0.085

<snip>

* predictions for 1989
mtable, at(yr89=1 male=1)     atmeans rowname(Men)       below roweqnm(1989)
mtable, at(yr89=1 male=0)     atmeans rowname(Women)     below roweqnm(1989)
mtable, dydx(male) at(yr89=1) atmeans rowname(Men_Women) below roweqnm(1989)

* DC for year by gender
mtable, dydx(yr89) at(male=1) atmeans rowname(77to89) below roweqnm(Men)
mtable, dydx(yr89) at(male=0) atmeans rowname(77to89) below roweqnm(Women)
```

The table...

* #32 building a table with global means

```
Expression: Marginal effect of Pr(warm), predict(outcome())

            |     1 SD       2 D       3 A      4 SA
 -----------+-------------------------------------
 1977       |
       Men  |    0.186     0.403     0.316     0.095
     Women  |    0.099     0.308     0.413     0.180
  Men Women |    0.087     0.094    -0.097    -0.085
 1989       |
       Men  |    0.119     0.339     0.390     0.151
     Women  |    0.061     0.228     0.441     0.270
  Men Women |    0.058     0.111    -0.050    -0.119
 Men        |
     77to89 |   -0.067    -0.063     0.074     0.056
 Women      |
     77to89 |   -0.038    -0.080     0.028     0.090

Specified values of covariates
<snip>
```

#33 predictions with local means

1. Should you assume the same values for **age**, **ed**, and **prst** for men in 1989 and women in 1977?

```
. sort yr89 male
. by yr89 male: sum white age ed prst

-> yr89 = 1977, male = Female
    Variable |       Obs        Mean    Std. Dev.       Min        Max
       white |       718    .8760446    .3297604          0          1
         age |       718    45.19638    16.59508         19         88
          ed |       718    11.73816    2.813291          3         19
        prst |       718    37.38579    13.53379         12         78
-> yr89 = 1977, male = Male
    Variable |       Obs        Mean    Std. Dev.       Min        Max
       white |       661    .8910741    .3117821          0          1
         age |       661    44.38729    16.49907         19         89
          ed |       661    11.86233     3.53949          0         20
        prst |       661    39.26475    14.58292         12         82
-> yr89 = 1989, male = Female
    Variable |       Obs        Mean    Std. Dev.       Min        Max
       white |       509    .8447937    .3624574          0          1
         age |       509     46.2888    17.17135         18         89
          ed |       509     12.6444     2.70048          3         20
        prst |       509    41.05108    14.81345         12         78
-> yr89 = 1989, male = Male
    Variable |       Obs        Mean    Std. Dev.       Min        Max
       white |       405    .8938272    .3084397          0          1
         age |       405    43.66667    16.98412         19         89
          ed |       405    13.11358    3.368747          0         20
        prst |       405    42.16543      14.999         12         82
```


2. With **mtable**

   a. if I *select cases by* ***yr89*** *and* ***male***,

   b. then **atmeans** holds other variables at means for the selected cases

```
mtable if yr89==0 & male==1, atmeans rowname(Men)   ///
    clear roweqnm(1977) nobs
mtable if yr89==0 & male==0, atmeans rowname(Women) ///
    below roweqnm(1977) nobs
mtable if yr89==1 & male==1, atmeans rowname(Men)   ///
    below roweqnm(1989) nobs
mtable if yr89==1 & male==0, atmeans rowname(Women) ///
    below roweqnm(1989) nobs
```

3. Alternatively, with less elegant output:

```
mtable, atmeans over(yr89 male)
```


#33 Probabilities by year and gender with local means

A. 1977	SD	D	A	SA
Men	**0.19**	0.40	0.31	0.09
Women	**0.10**	0.32	0.41	0.17
Men-Women	***0.09***	*0.09*	*-0.10*	*-0.08*

B. 1989	SD	D	A	SA
Men	0.11	0.33	0.40	0.16
Women	0.06	0.23	0.44	0.27
Men-Women	0.05	0.10	-0.04	-0.11

C. 1977 to 1989	SD	D	A	SA
Men	-0.08	-0.07	0.09	0.07
Women	-0.04	-0.09	0.03	0.10

Note: Other variable are held at means for given year and gender.

Difference between global and local predictions

	1SD	2D	3A	4SA
1977				
Men	-0.00	-0.00	0.00	0.00
Women	-0.00	-0.01	0.00	0.01
1989				
Men	0.01	0.01	-0.01	-0.01
Women	0.00	0.00	-0.00	-0.00
1977				
Men_Women	0.00	0.01	-0.00	-0.01
1989				
Men_Women	0.01	0.01	-0.01	-0.01

1. The differences are small enough that I am confident that the conclusions on the effects of year and gender are not affect by using global means
2. *Which set of probabilities would you use?*

Marginal effects

MEM and MER

1. Discrete change at x* is defined as:

$$\frac{\Delta \Pr\left(y = q \mid \mathbf{x}^*\right)}{\Delta x_k} = \Pr\left(y = q \mid \mathbf{x}^*, \text{End } x_k\right) - \Pr\left(y = q \mid \mathbf{x}^*, \text{Start } x_k\right)$$

2. The change is interpreted as

   When x_k changes from the start value to the end value, the predicted probability of outcome q changes by $\Delta \Pr\left(y = q \mid \mathbf{x}^*\right) / \Delta x_k$, holding other variables at **x***.

3. Since the model is nonlinear, the discrete change depends on
   a. The level of all variables that are not changing
   b. The value at which x_k starts
   c. The amount of change in x_k

4. Mean absolute change summarizes the discrete change for a variable

$$\bar{\Delta} = \frac{1}{J} \sum_{j=1}^{J} \left| \frac{\Delta \Pr\left(y = j \mid \mathbf{x}^*\right)}{\Delta x_k} \right|$$

   (This is not computed by **`mtable`**.)

5. Marginal change can also be computed

$$\frac{\partial \Pr\left(y = q \mid \mathbf{x}^*\right)}{\partial x_k}$$

Average marginal effect (AME)

1. Average discrete change

$$\text{ADC} = \text{mean}\frac{\Delta \Pr(y = j \mid \mathbf{x})}{\Delta x_k} = \frac{1}{N}\sum_{i=1}^{N}\frac{\Delta \Pr(y = j \mid \mathbf{x}_i)}{\Delta x_{ik}}$$

2. Interpretation

   The average change in the probability of outcome *q* is (value of *AME*) when x_k changes from the start value to the end value.

3. For example

   On average being male decreases the probability of strongly agreeing that working mothers can be good mothers by .10.

4. Test that the change is 0 can be computed

#44 AME (some p-values cannot be computed in Stata 12)

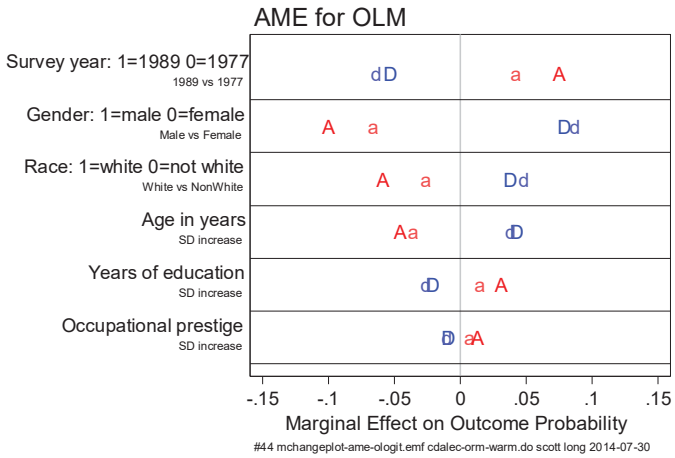
Lots of numbers that are graphed below

```
. mchange

ologit: Changes in Pr(y) | Number of obs = 2293

Expression: Pr(warm), predict(outcome())

                   |      1 SD        2 D        3 A        4 SA
-------------------+----------------------------------------------
yr89               |
      1989 vs 1977 |    -0.053     -0.064      0.042      0.075
           p-value |     0.000      0.000      0.000      0.000
male               |
    Male vs Female |     0.079      0.087     -0.066     -0.100
           p-value |     0.000      0.000      0.000      0.000
white              |
 White vs NonWhite |     0.038      0.048     -0.026     -0.059
           p-value |     0.000      0.001      0.000      0.002
age                |
                +1 |     0.002      0.003     -0.002     -0.003
           p-value |     0.000      0.000      0.000      0.000
               +SD |     0.043      0.038     -0.036     -0.046
           p-value |     0.000      0.000      0.000      0.000
          Marginal |     0.002      0.003     -0.002     -0.003
           p-value |     0.000      0.000      0.000      0.000
```

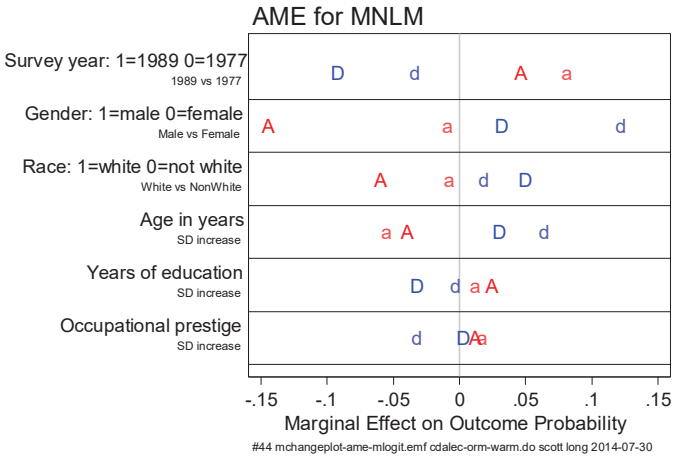


```
ed                 |
                +1 |    -0.007     -0.008      0.005      0.010
           p-value |     0.000      0.000      0.000      0.000
               +SD |    -0.021     -0.026      0.015      0.031
           p-value |     0.000      0.000      0.000      0.000
          Marginal |    -0.007     -0.008      0.006      0.009
           p-value |     0.000      0.000      0.000      0.000
prst               |
                +1 |    -0.001     -0.001      0.001      0.001
           p-value |     0.066      0.065      0.066      0.065
               +SD |    -0.009     -0.010      0.007      0.013
           p-value |     0.058      0.069      0.052      0.071
          Marginal |    -0.001     -0.001      0.001      0.001
           p-value |     0.066      0.065      0.067      0.065

Average predictions

             |      1_SD        2_D        3_A        4_SA
-------------+----------------------------------------------
  Pr(y|base) |     0.129      0.315      0.374      0.182
```

AME for OLM



AME for MNLM: do you see important differences from OLM?



#45 Second differences

1. Gender has significant effects in 1977 and 1989

```
. * DC(male) in 1977
. mtable, dydx(male) at(yr89=0) rowname(DCmale_77) clear
:::
. * DC(male) in 1989
. mtable, dydx(male) at(yr89=1) rowname(DCmale_89) below

           |     1 SD       2 D       3 A      4 SA
-----------+-----------------------------------------
 DCmale 77 |    0.089     0.081    -0.083    -0.088
 DCmale 89 |    0.062     0.099    -0.041    -0.119
```

2. Is the effect of gender significantly larger in 1977?
3. We can test his by computing the *second differences* using
   - **`margins, post`** and **`mlincom`**

#45 margins for predictions with post

1. **mtable** cannot post predictions for multiple outcomes, so we use **margins**
   - Simpler code is possible in Stata 14, but I assume you have 13 or earlier

#44 margins by year and gender for outcome 1

```
. margins, at(yr89=(0 1) male=(0 1)) predict(outcome(1)) post
Predictive margins                                 Number of obs   =       2293

Expression   : Pr(warm==1), predict(outcome(1))
1._at        : yr89            =           0
               male            =           0
2._at        : yr89            =           0
               male            =           1
3._at        : yr89            =           1
               male            =           0
4._at        : yr89            =           1
               male            =           1

             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .1085645   .0077007    14.10   0.000     .0934714    .1236576
          2  |   .1980111   .0116351    17.02   0.000     .1752067    .2208155
          3  |   .0680602   .0058309    11.67   0.000     .0566319    .0794886
          4  |   .1298218   .0099298    13.07   0.000     .1103597    .1492839
------------------------------------------------------------------------------
```


2. We use **lincom** to test the second difference:

```
. lincom (_b[1bn._at]-_b[2._at]) - (_b[3._at]-_b[4._at])

 ( 1)  1bn._at - 2._at - 3._at + 4._at = 0

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   -.027685   .0049492    -5.59   0.000    -.0373854   -.0179847
------------------------------------------------------------------------------
```

3. **mlincom** is more elegant:

```
mlincom (1-2)-(3-4)
             |   lincom    pvalue
-------------+------------------
    Outcome1 |   -0.028     0.000
```

4. Combining results we find

	1 SD	2 D	3 A	4 SA
DCmale 77	0.089	0.081	-0.083	-0.088
DCmale 89	0.062	0.099	-0.041	-0.119
Difference	0.027	-0.018	-0.042	-0.031

The effects of gender changed significantly between 1977 to 1989.

*#45 Loop over outcomes

```
. mlincom, clear // remove any prior results
. estimate restore olmfv

. foreach o in 1 2 3 4 {
 2.  quietly {
 3.    margins, at(yr89=(0 1) male=(0 1)) predict(outcome(`o')) post
 4.    mlincom (1-2)-(3-4), add rowname(Outcome`o') stats(est p)
 5.    estimate restore olmfv
 6.  }
 7. }

. mlincom

             |   lincom    pvalue
-------------+------------------
    Outcome1 |   -0.028     0.000
    Outcome2 |    0.018     0.000
    Outcome3 |    0.042     0.000
    Outcome4 |   -0.032     0.000
```

The effects of gender changed significantly between 1977 to 1989.

Plotting probabilities

1. Plots are useful for examining effects of continuous variables
2. It is significantly more complicated than for the BRM since
   a. With **two** outcomes, plot **one** probability
   b. With **three** outcomes, plot **three** probabilities
3. The formula for the probabilities and cumulative probabilities are:

$$\Pr(y = q \mid \mathbf{x}) = F(\tau_q - \mathbf{x}\boldsymbol{\beta}) - F(\tau_{q-1} - \mathbf{x}\boldsymbol{\beta})$$

$$\Pr(y \le q \mid \mathbf{x}) = \sum_{j=1}^{q} \Pr(y = j \mid \mathbf{x}) = F(\tau_q - \mathbf{x}\boldsymbol{\beta})$$

4. These are computed holding other variables constant
   - Exception: linked variables like x and x^2 change together

Computing probabilities to plot

1. For each outcome q compute
   a. Probability: $\Pr(y=q \mid z, \mathbf{X}^*)$ as z changes
   b. Cumulative probability: $\Pr(y \le q \mid z, \mathbf{X}^*)$ as z changes
2. Create two plots:
   - probability
   - cumulative probabilities
   - Pick the plot that is most effective
3. We will use the cumulative probability plot to explain the parallel regression constraint that is implicit in ordinal models.
4. Then, we will use plots to compare ordinal and nominal models.

#51 mgen for the ORM

For women in 1989, do attitudes change with age?

#51 Computing the predictions

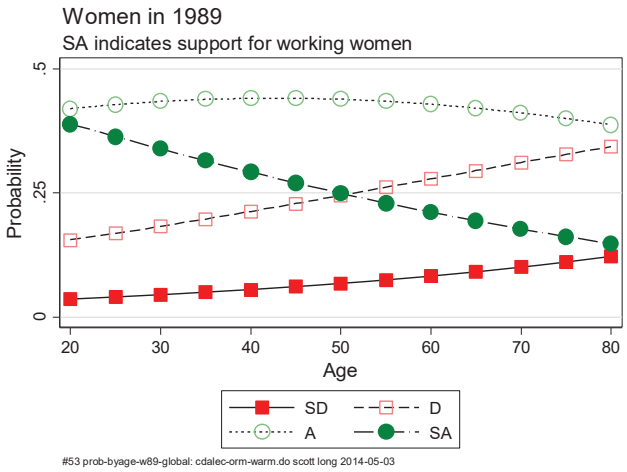
```
. mgen, at(age=(20(5)80) male=0 yr89=1) atmeans stub(W89)
Predictions from: margins, at(age=(20(5)80) male=0 yr89=1) atmeans predict(outco
> me())

Variable   Obs Unique      Mean       Min       Max  Label
------------------------------------------------------------------------------
W89pr1      13     13  .0720984  .0364676   .121933  pr(y=1_SD) from margins
W89ll1      13     13  .0586271  .0281642   .097223  95% lower limit
W89ul1      13     13  .0855696   .044771   .146643  95% upper limit
W89age      13     13        50        20        80  Age in years
W89Cpr1     13     13  .0720984  .0364676   .121933  pr(y<=1_SD)
W89pr2      13     13     .2465  .1551205  .3431755  pr(y=2_D) from margins
W89ll2      13     13  .2198686  .1308338  .3085207  95% lower limit
W89ul2      13     13  .2731315  .1794073  .3778303  95% upper limit
W89Cpr2     13     13  .3185984  .1915881  .4651085  pr(y<=2_D)
<snip>
------------------------------------------------------------------------------

Specified values of covariates

                          1.
     yr89       male      white         ed       prst
-------------------------------------------------------
        1          0   .8765809   12.21805   39.58526
```

#52-54 Plotting predictions

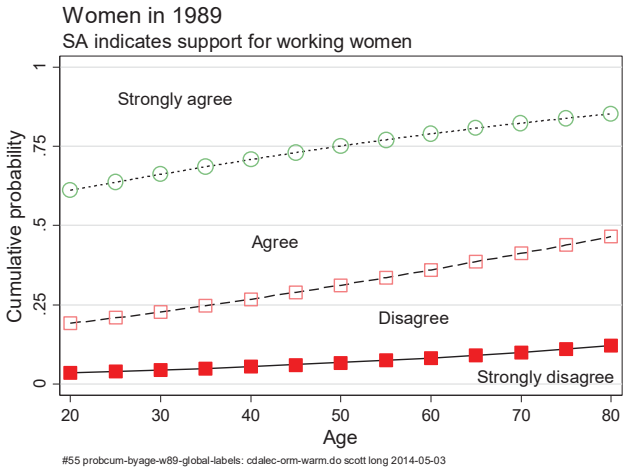
```
. label var W89pr1 "SD"
. label var W89pr2 "D"
. label var W89pr3 "A"
. label var W89pr4 "SA"
. label var W89Cpr1 "SD"
. label var W89Cpr2 "SD or D"
. label var W89Cpr3 "SD, D or A"
. local warmsym "mcol(red red*.5 green*.5 green) "
. local warmsym "`warmsym' msym(s sh Oh O) msiz(3.5 3.5 3 3)"

. graph twoway connected W89pr1 W89pr2 W89pr3 W89pr4 W89age, ///
>     `warmsym' title(Women in 1989, pos(11))  ///
>      subtitle("SA indicates support for working women", pos(11)) ///
>      xtitle("Age") xlabel(20(10)80) ///
>      ylabel(0(.25).50, grid gmin gmax) ///
>      xline(44.93) ytitle("Probability")

. graph twoway connected W89Cpr1 W89Cpr2 W89Cpr3 W89age, ///
>     `warmsym' title(Women in 1989, pos(11)) mcol(`warmcol') ///
>     subtitle("SA indicates support for working women", pos(11)) ///
>     xtitle("Age") xlabel(20(10)80) ylabel(0(.25)1, grid gmin gmax) ///
>     xline(44.93) ytitle("Cumulative probability")
```

5. After viewing these plots, I build the CDF plot in steps

#53 Predicted probabilities

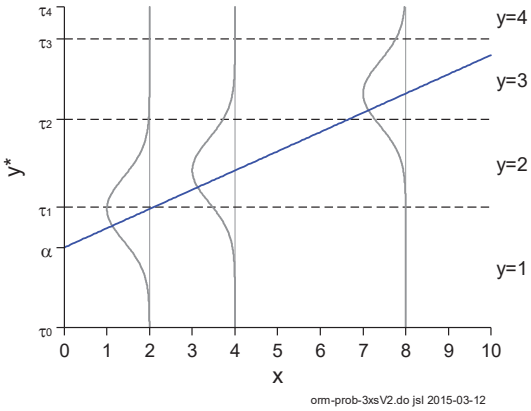


#55 Cumulative probabilities with labels



Odds ratios for the OLM

1. What does this picture tell you about ORs for cumulative probabilities?



2. Tedious algebra shows the *cumulative probability* equals

$$\Pr(y \le q \mid \mathbf{x}) = \sum_{j=1}^{q} \Pr(y = j \mid \mathbf{x}) = \Lambda(\tau_q - \mathbf{x}\boldsymbol{\beta}) \quad \text{for } q = 1, J-1$$

3. This is a BLM with intercept $\alpha_q = (\tau_q - \beta_0)$ and slopes $\boldsymbol{\beta}^* = -\boldsymbol{\beta}$

$$\Pr(y \le q \mid \mathbf{x}) = \Lambda(\tau_q - \mathbf{x}\boldsymbol{\beta}) = \Lambda(\alpha_q + \mathbf{x}\boldsymbol{\beta}^*)$$

4. The *odds* of *y≤q* versus *y>q* given **x** is

$$\Omega_q(\mathbf{x}) = \frac{\Pr(y \le q \mid \mathbf{x})}{\Pr(y > q \mid \mathbf{x})} = \exp(\tau_q - \mathbf{x}\boldsymbol{\beta}) = \exp(\alpha_q + \mathbf{x}\boldsymbol{\beta}^*)$$

5. The *odds ratio* for a change in x_k

$$\frac{\Omega_q(\mathbf{x}, x_k + 1)}{\Omega_q(\mathbf{x}, x_k)} = \exp(\beta_k^*)$$

6. Interpretation

   For a unit increase in x_k the odds of lower outcomes compared to higher outcomes change by the factor $\exp(\beta_k^*)$, holding other variables constant.

#62 Odds ratios for supporting working mothers

NOTE: Odds of higher compared to lower outcome

$$\Omega\left(\frac{P(SD)}{P(D,A,SA)}\right) = \Omega\left(\frac{P(SD,D)}{P(A,SA)}\right) = \Omega\left(\frac{P(SD,D,A)}{P(SA)}\right).$$

```
. listcoef, help
ologit (N=2293): Factor change in odds

  Odds of: >m vs <=m

            |          b         z     P>|z|       e^b   e^bStdX     SDofX
------------+------------------------------------------------------------
       yr89 |
       1989 |     0.5239     6.557     0.000     1.689     1.292     0.490
       male |
       Male |    -0.7333    -9.343     0.000     0.480     0.694     0.499
      white |
      White |    -0.3912    -3.304     0.001     0.676     0.879     0.329
        age |    -0.0217    -8.778     0.000     0.979     0.695    16.779
         ed |     0.0672     4.205     0.000     1.069     1.237     3.161
       prst |     0.0061     1.844     0.065     1.006     1.092    14.492
-------------------------------------------------------------------------
      b = raw coefficient
    e^b = exp(b) = factor change in odds for unit increase in X
e^bStdX = exp(b*SD of X) = change in odds for SD increase in X
  SDofX = standard deviation of X
```

Interpretations on the next page...

1. From 1977 to 1989, the odds of being more positive toward working women increased by a factor of 1.7, holding other variables **constant**.

```
             |          b        z    P>|z|       e^b   e^bStdX      SDofX
-------------+------------------------------------------------------------
      yr89 |
      1989 |     0.5239    6.557    0.000     1.689     1.292      0.490
```

2. Being male decreases the odds of having more favorable attitudes toward working women by a factor of .48, holding other variables constant.

```
             |          b        z    P>|z|       e^b   e^bStdX      SDofX
-------------+------------------------------------------------------------
      male |
      Male |    -0.7333   -9.343    0.000     0.480     0.694      0.499
```

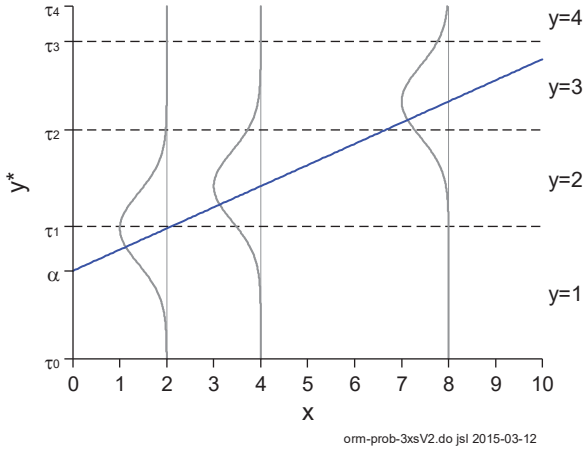
3. As standard deviation increase in age, 17 years, decreases the odds of supporting working mothers by a factor of .70, holding other variables constant.

```
      warm |       b          z      P>|z|     e^b    e^bStdX       SDofX
-----------+------------------------------------------------------------
       age |  -0.02167   -8.778     0.000    0.9786    0.6952     16.7790
```

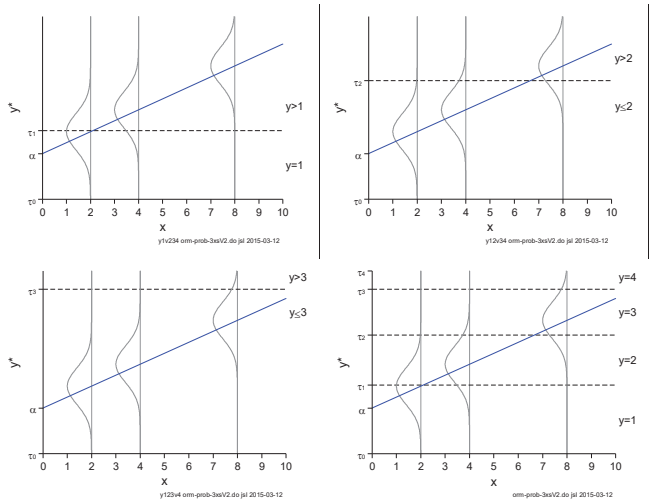
4. We do *not* need to say where variables are held constant.

Parallel regressions (for 4 outcomes)

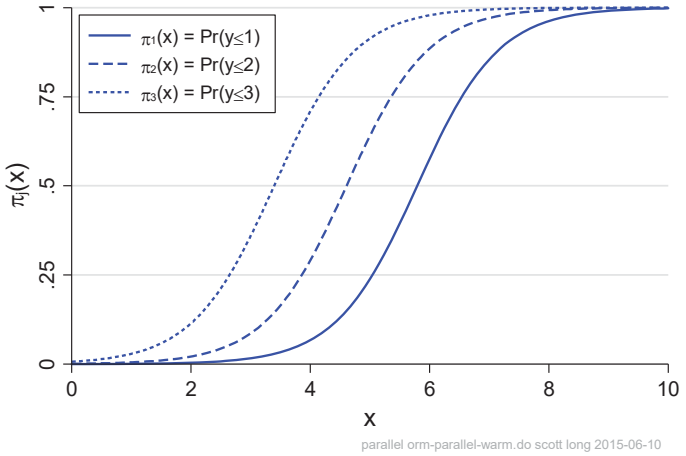
1. The ORM is a set of binary logits on outcomes if $y \leq j$ where each binary logit has: (a) the same slopes; (b) different intercepts



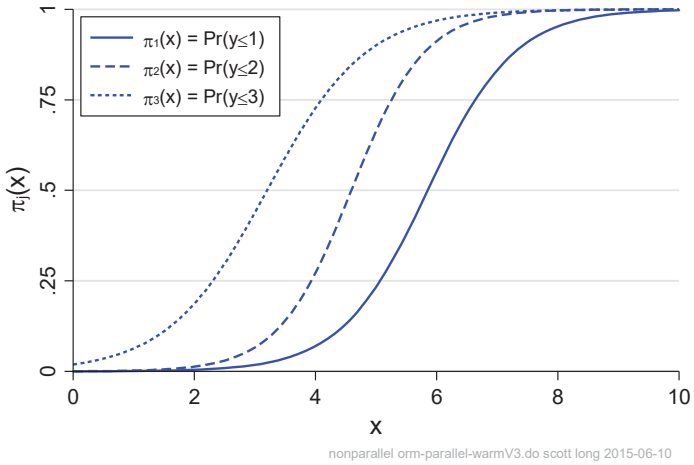
Dichotomizing anyplace and the slope is unchanged



The parallel regression assumption imples



Without the parallel regression assumption



Assessment of parallel regressions

1. You can look at how the curves/coefficients vary when parallel regressions is imposed with the OLM compared to a set of BRMs.
2. You can use the Brant test or other formal tests of parallel regressions.
   - My experience is that these are not very good at differentiating between cases where the parallel regression assumption makes a *substantive* difference in the results.
3. I find it more useful to compare the substantive results from the OLM to those from the MNLM.

#66 Comparing ologit and mlogit predictions

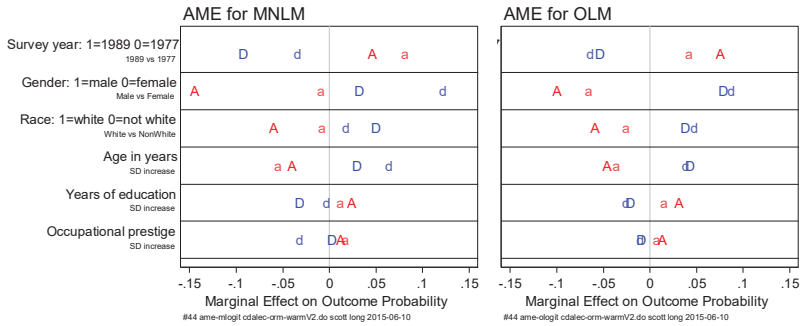
1. Before accepting or rejecting the OLM based on a test of parallel regressions, you should examine whether predictions from the OLM differ from those in a model that *does not impose ordinality*.
2. Here I compare OLM to MNLM. You could also compare predictions to those from the generalized ordered logit model.

#66 estimating models and computing predictions

```
. ologit warm i.yr89 i.male i.white age ed prst
  <snip>
. predict OLMpr1 OLMpr2 OLMpr3 OLMpr4
  <snip>
. mlogit warm i.yr89 i.male i.white age ed prst
  <snip>
. predict NRMpr1 NRMpr2 NRMpr3 NRMpr4
  <snip>
. corr OLMpr1 NRMpr1
  <snip>

    NRMpr1 & OLMpr1: 0.9013
    NRMpr2 & OLMpr2: 0.9239
    NRMpr3 & OLMpr3: 0.8593
    NRMpr4 & OLMpr4: 0.9469
```


#44 AME for MNLM and OLM: ordinality?

Do you see major differences in the conclusions from the two models?



Modeling political party (-orm-partyid.do)

1. The parallel regression assumption is a reason to consider alternatives to ORM
2. More fundamentally: *is an ordinal regression model appropriate?*
3. Consider the voting example from the American National Election Study

#11 The variables

```
. tab party, miss

   Party ID |      Freq.     Percent        Cum.
------------+-----------------------------------
     StrDem |        266       19.25       19.25
        Dem |        427       30.90       50.14
      Indep |        151       10.93       61.07
        Rep |        369       26.70       87.77
     StrRep |        169       12.23      100.00
------------+-----------------------------------
      Total |      1,382      100.00

. sum party age income black female i.educ

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       party |      1382    2.817656    1.342787          1          5
         age |      1382    45.94645    16.78311         18         91
      income |      1382    37.45767    27.78148        1.5     131.25
       black |      1382    .1374819       .34448          0          1
      female |      1382    .4934877    .5001386          0          1
        educ |
    hs only  |      1382    .5803184    .4936854          0          1
    college  |      1382    .2590449    .4382689          0          1
```


#22 OLM

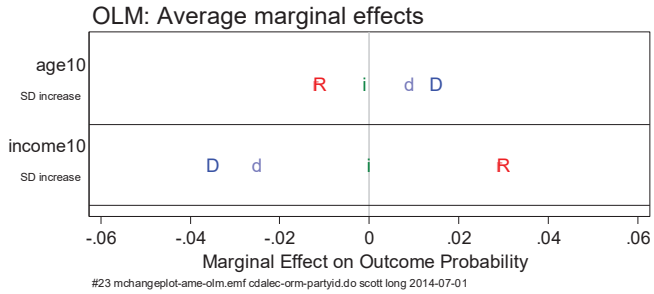
```
. ologit party age10 income10 i.black i.female i.educ
  <snip>
. listcoef, help

ologit (N=1382): Factor Change in Odds

  Odds of: >m vs <=m

------------------------------------------------------------------------------
             |          b        z    P>|z|       e^b   e^bStdX     SDofX
-------------+----------------------------------------------------------------
       age10 |    -0.0636   -2.037    0.042     0.938     0.899     1.678
    income10 |     0.0961    4.792    0.000     1.101     1.306     2.778
       black |
         yes |    -1.4759   -9.824    0.000     0.229     0.601     0.344
      female |
         yes |    -0.1571   -1.584    0.113     0.855     0.924     0.500
        educ |
    hs only  |     0.2942    1.943    0.052     1.342     1.156     0.494
    college  |     0.6420    3.543    0.000     1.900     1.325     0.438
------------------------------------------------------------------------------
       b = raw coefficient
     e^b = exp(b) = factor change in odds for unit increase in X
 e^bStdX = exp(b*SD of X) = change in odds for SD increase in X
   SDofX = standard deviation of X
```

#22 Testing parallel regression assumption

```
. brant

Brant Test of Parallel Regression Assumption

             |       chi2     p>chi2      df
-------------+------------------------------
         All |      89.84      0.000       18
-------------+------------------------------
       age10 |      42.87      0.000        3
    income10 |       2.11      0.550        3
     1.black |      12.82      0.005        3
    1.female |       6.54      0.088        3
      2.educ |       2.92      0.404        3
      3.educ |      12.24      0.007        3
   Variable |      chi2   p>chi2    df

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

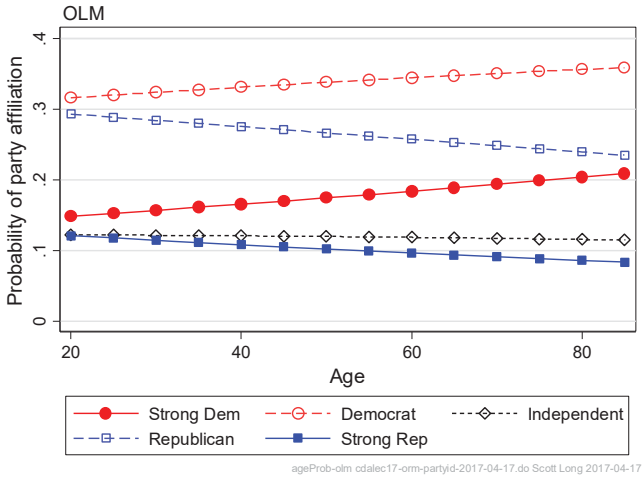
- We explore effects of age and income in ORM and MNLM

#23 OLM: Average discrete change

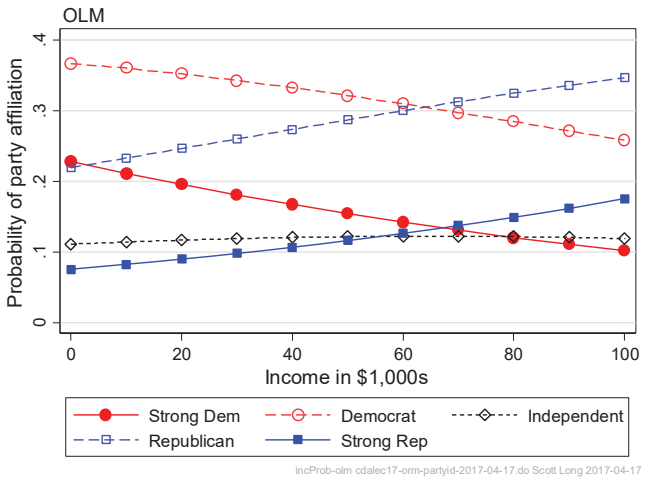


1. Age increase the probability of *both* Democratic affiliations and decreases the probability of *both* Republican affiliations, while income decrease the probability of *both* Democratic affiliations and increases the probability of *both* Republican affiliations.
2. Graphs of predicted probabilities show these relationships.

OLM: Predicted probabilities by age



#24 OLM: Predicted probabilities by income



#12 MNLM model

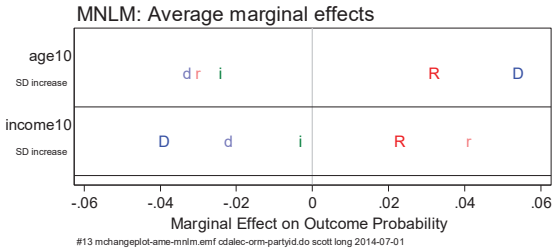
1. This model was fit in the chapter on MNLM
2. Here is a summary of results

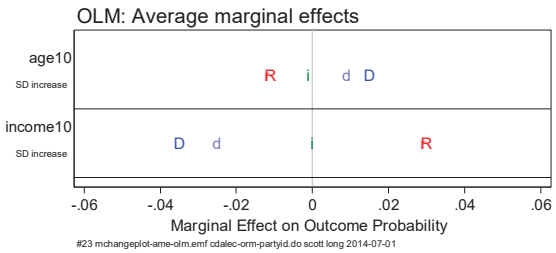
Wald tests for independent variables (N=1382)

Ho: All coefficients associated with given variable(s) are 0

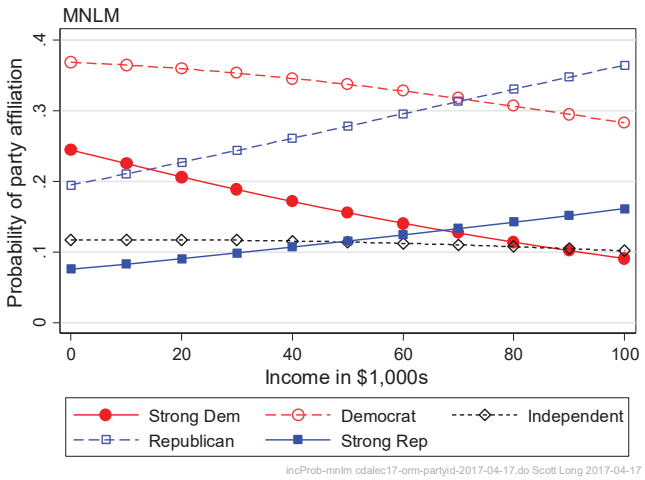
	chi2	df	P>chi2	
age10	43.815	4	0.000	p=.042 for OLM
income10	22.985	4	0.000	p=.000 for OLM
age10	43.815	4	0.000	
income10	22.985	4	0.000	
1.black	83.978	4	0.000	
1.female	9.087	4	0.059	
2.educ	5.569	4	0.234	
3.educ	20.613	4	0.000	

#13 MNLM: ADC of age and income: What's going on?

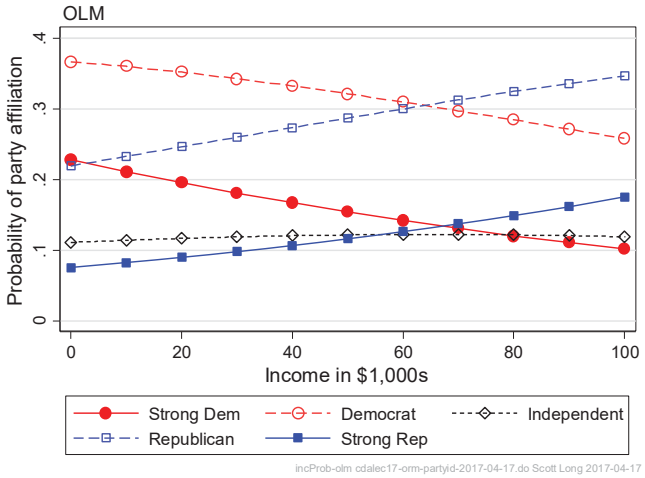




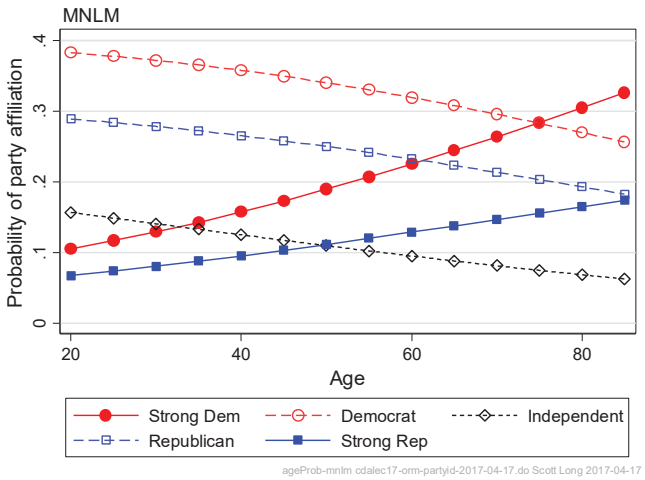
#14 MNLM Predicted probabilities by income



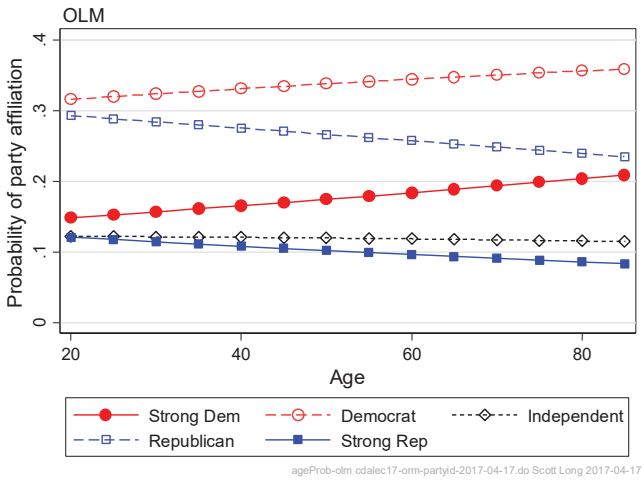
#24: OLM Predicted probabilities by income



#14 MNLM Predicted probabilities by age

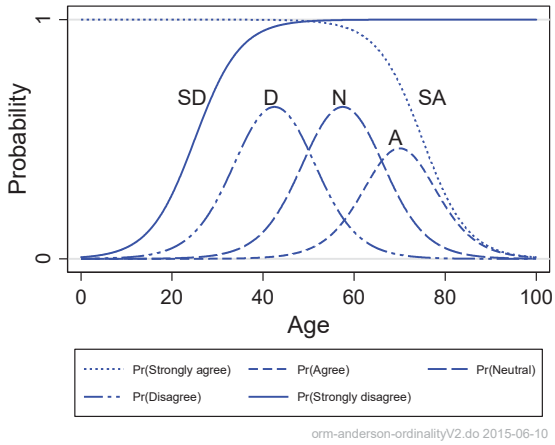


#24 OLM: Predicted probabilities by age

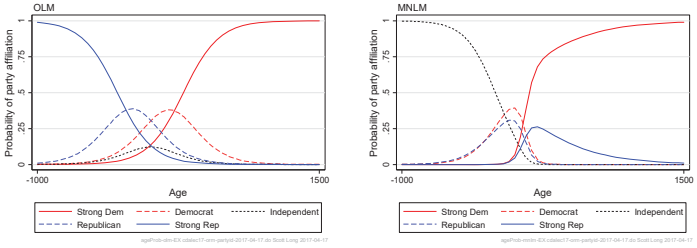


Ordinal or nominal?

1. I find the ordinal regression models (i.e., those meeting Anderson's definition) to be overly restrictive for many substantive applications



2. The ordinality assumptions is relaxed in the MNLM and the generalized ordered logit model (GOLM) which often provide similar predictions
   - o I prefer MNLM because I find odds of two categories to be more intuitive than odds of lower categories vs greater categories
   - o I also find OR plots for MNLM to be useful
3. Any time you are use an ordinal regression model, compare the result to those from a model that is not ordinal.
4. Returning to our example of political party, compare what happens in the OLM and MNLM if we extend the range of age:



* Alternative models for ordinal outcomes

My web page has a recent paper that examines ordinal models

The stereotype regression model

1. The SORM relaxes the parallel regression assumption
2. As the model is made more complex, it becomes the MNLM
3. SORM is estimated with Stata's **`slogit`**
4. See Long and Freese for details

The adjacent logit model

1. This model puts constraints on the MNLM to create an ordinal model
2. Substantively, it is rarely a reasonable model (in my experience)
3. The model is specified as follows where β is the same for all values of q

$$\ln\left[\frac{\Pr(y=q\,|\,\mathbf{x})}{\Pr(y=q+1\,|\,\mathbf{x})}\right]=\tau_q-\mathbf{x}\boldsymbol{\beta}$$

The continuation ratio model

1. The CRM is for outcomes in which the categories represent the progression of events or stages in some process through which an individual can advance
2. Model is estimated by Stata's **`ocratio`**

$$\frac{\Pr(y=m\,|\,\mathbf{x})}{\Pr(y>m\,|\,\mathbf{x})}=\exp(\tau_m-\mathbf{x}\boldsymbol{\beta})$$

GOLM relaxes the assumption of equal β's

1. Define

$$\Omega_{y\le q}(\mathbf{x})=\frac{\Pr(y\le q\,|\,\mathbf{x})}{\Pr(y>q\,|\,\mathbf{x})}$$

2. OLM

$$\ln\Omega_{y\le q}(\mathbf{x})=\tau_q-\mathbf{x}\boldsymbol{\beta}$$

3. GOLM removes the restriction of equal βs and is not an ordinal model

$$\ln\Omega_{y\le q}(\mathbf{x})=\tau_q-\mathbf{x}\boldsymbol{\beta}_q \quad \text{for } q=1,J-1$$

Overview of ordinal LHS

1. If you are using LRM for ordinal outcomes, consider y* standardized coefficients from the ORM. If you *must* use the LRM, at least verify that the conclusions are consistent with those from the ORM.
2. Before using ordinal models, consider whether your variable is ordinal as Stevens defined it:
   - Categories are ranked on a *single dimension*
3. *Always do a sensitivity analysis* before accepting the results of ORM. Compare results to those from a nominal model (MNLM or GOLM)
4. Even if you don't find the ORM useful for your work, this model is the foundation for the IRT and Rasch models for ordinal indicators

Part 12: Count outcomes

Read and run

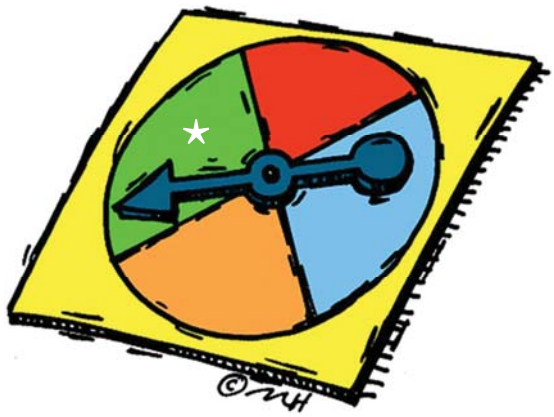
Long & Freese Chapter 9

cdalec*.do cdalec17-crm-couart-.do

Roadmap

1. What random process generates counts?
2. In what ways can people/countries/… differ in the rate at which something occurs?

How many times does the spinner land on green?



Explaining count outcomes

Chance alone

1. **Poisson PDF** — *Chance alone explains variation.*

Chance and heterogeneity

1. **Poisson regression model** — *Add observed heterogeneity.*
2. **Negative binomial regression** — *Add unobserved continuous heterogeneity.*
3. **Mixture count models** — *Add unobserved discrete heterogeneity.*

Overview

1. Start with a Poisson process for modeling counts
   - The bigger the green region, the bigger the rate
2. Let the be determined by *observed* characteristics
   - Characteristics affect the size of the green region
3. This is rarely sufficient so allow *unobserved* heterogeneity

The Poisson Process

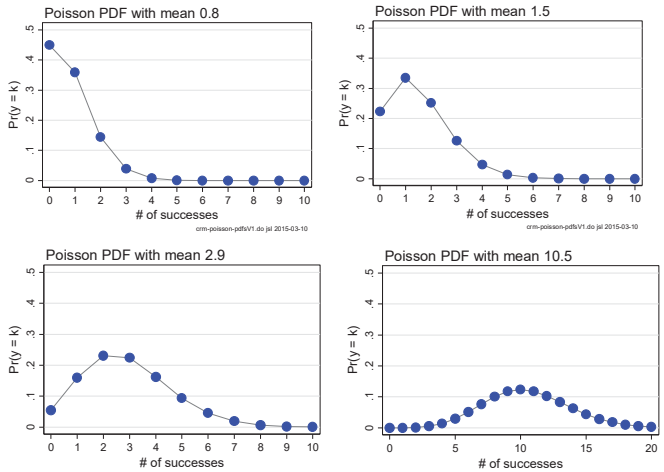
Siméon-Denis Poisson (1781 – 1840)

A Poisson process

1. The *Poisson distribution* is derived from a stochastic process
2. Start the counter at 0
3. Spin
4. If a green *success*, increase the counter; if not green, no change in counter
5. Spin again with next outcome *independent* of prior outcomes
6. Continuing this process leads to a Poisson PDF for the number of successes
7. Repeat for every person in your sample.

Examples of Poisson PDFs with different rates

The size of the green region determines the *rate* at which success occurs.



Formula for Poisson PDF

1. *y* is a random variable that counts the # of successes
2. *e*=2.71828... and *y*!=*y**(*y*-1)*(*y*-2)*...*3*2*1
3. A *Poisson distribution* with the *rate* μ>0 is

$$\Pr(y \mid \mu) = \frac{e^{-\mu}\mu^{y}}{y!} \qquad \text{for } y = 0,1,2,\ldots$$

4. Examples

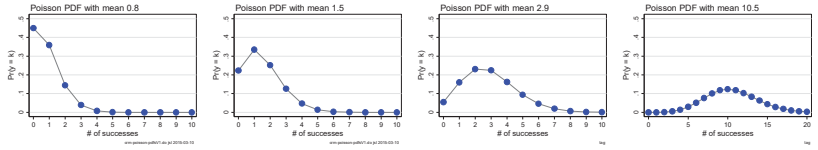
$$\Pr(y = 0 \mid \mu) = \frac{e^{-\mu}\mu^{0}}{0!} = e^{-\mu}$$

$$\Pr(y = 1 \mid \mu) = \frac{e^{-\mu}\mu^{1}}{1!} = e^{-\mu}\mu$$

$$\Pr(y = 3 \mid \mu) = \frac{e^{-\mu}\mu^{3}}{3!} = \frac{e^{-\mu}\mu^{3}}{6}$$

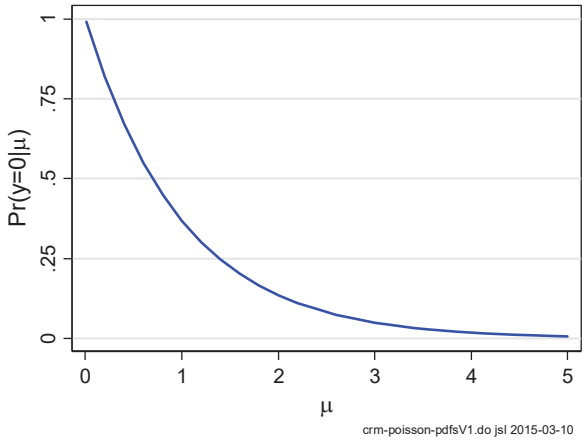
$$\Pr(y = 20 \mid \mu) = \frac{e^{-\mu}\mu^{20}}{20!} = \frac{e^{-\mu}\mu^{20}}{2,432,902,008,176,640,000}$$

The effect of μ on the Poisson distribution



μ=.8 μ=1.5 μ=2.9 μ=10.5

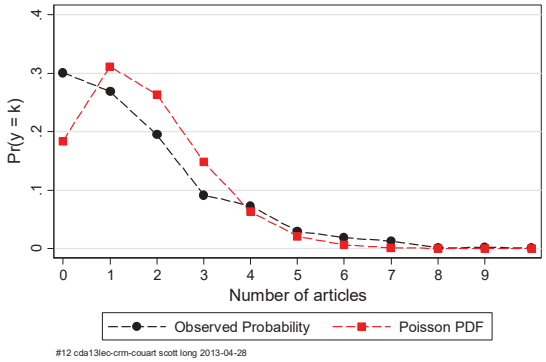
1. *μ* is the rate: *μ= E(y)*
2. As μ increases, mass moves right
3. Variance equals rate: *E(y)=Var(y)=μ*
4. As μ increases, the distribution approaches normal
5. As μ increases, the probability of 0's decreases rapidly as shown here...



$$\Pr(y = 0 \mid \mu = .8) = .45 \qquad \Pr(y = 0 \mid \mu = 1.5) = .22$$

$$\Pr(y = 0 \mid \mu = 2.9) = .05 \qquad \Pr(y = 0 \mid \mu = 10.5) = .00002$$

Fitting Poisson PDF to # of articles



1. Mean articles is 1.7 with variance 3.7
2. Compared to Poisson PDF, observed data has: 1) more 0s; 2) fewer cases in middle; and 3) more cases in the upper tail
3. Data are not consistent with a Poisson process

#11 & 12 Plotting a Poisson PDF (-crm-couart.do)

```
. poisson art, nolog

Poisson regression                                Number of obs   =        915
                                                  LR chi2(0)      =       0.00
                                                  Prob > chi2     =          .
Log likelihood = -1742.5735                       Pseudo R2       =     0.0000

------------------------------------------------------------------------------
         art |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .5264408   .0254082    20.72   0.000     .4766416      .57624
------------------------------------------------------------------------------

. * create variables with mean predictions
. mgen, pr(0/10) meanpred stub(pdf)
  <snip>

. label var pdfpreq "Poisson PDF" // label for plot
. label var pdfobeq "Observed Probability" // label for plot
```

Generated variables are listed on next page...

Results from mgen

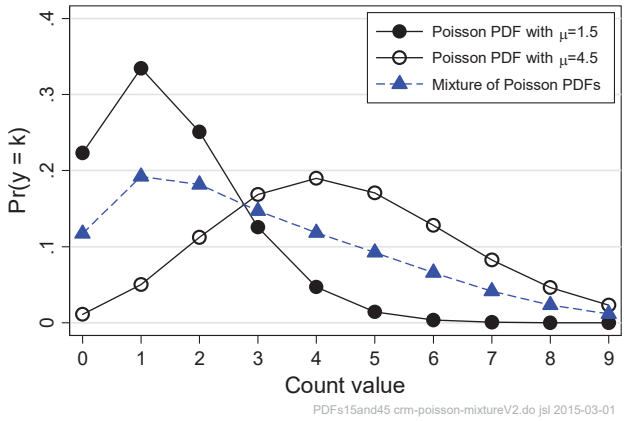
```
. list pdfval pdfobeq pdfpreq in 1/12, clean

      pdfval    pdfobeq    pdfpreq
  1.       0   .3005464   .1839859
  2.       1   .2688525    .311469
  3.       2   .1945355   .2636424
  4.       3   .0918033    .148773
  5.       4    .073224   .0629643
  6.       5   .0295082   .0213184
  7.       6   .0185792    .006015
  8.       7   .0131148   .0014547
  9.       8   .0010929   .0003078
 10.       9   .0021858   .0000579
 11.      10   .0010929   9.80e-06
 12.       .          .          .
```

To create the graph

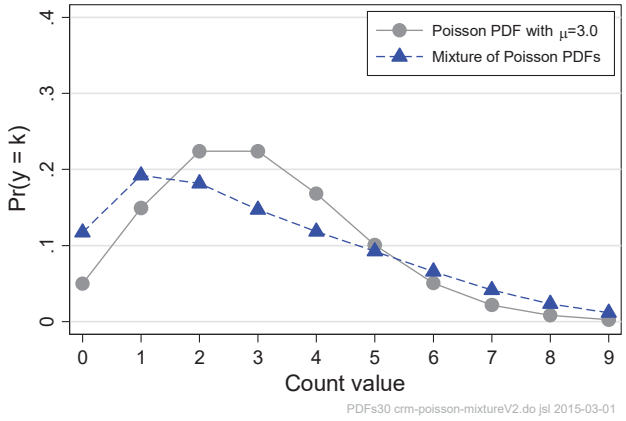
```
twoway connected pdfobeq pdfpreq pdfval, ///
    msym(O s) msiz(2 2.4) mcol(black red) lcol(black red) lpat(dash dash) ///
    ytitle("Pr(y = k)") xtitle("Number of articles") ///
    ylab(0(.1).4, grid gmax gmin) xlab(0(1)9, nogrid)
```

The *BIG* idea of heterogeneity

A 50/50 mix of two Poisson distributions

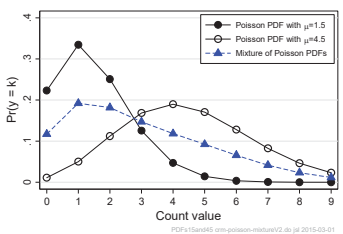


Mixture of Poissons vs Poisson at combined mean

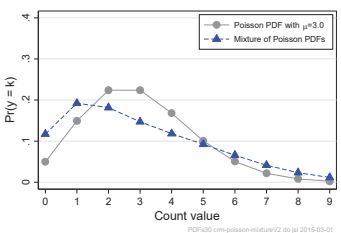


Summary of heterogeneity

Panel A: Mixture versus Poisson at mixed mean



Panel B: Observed data compared to predictions from Poisson PDF



1. Mixture of Poisson distributions does *not* have a Poisson distribution
2. Mixture has excess low and high counts, just like our observed data
3. Failure to account for heterogeneity leads to *overdispersion*

The Poisson regression model (PRM)

1. The PRM adds *observed heterogeneity*

$$\mu_i = \mathrm{E}(y_i \mid \mathbf{x}_i) = \exp(\mathbf{x}\boldsymbol{\beta})$$
$$= \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})$$
$$\ln \mu_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}$$

   where μ_i is the mean of Poisson distribution for observation *i*

2. Taking the exponential of **xβ** forces μ to be positive
3. For observation i,

$$\Pr(y_i = m \mid \mathbf{x}_i) = \frac{e^{-\mu_i}\mu_i^m}{m!}$$

   which differs for each *i*

4. PRM is didactically useful, but *I do not recommend the PRM* for reasons considered below.

PRM with a single binary regressor

1. Let

$$\mu = \exp(-.12 - .13 Female)$$

2. Rates differ by gender

$$\mu_{\text{Female}} = \exp(-.12 - .13) = \exp(-.25) = .779$$
$$\mu_{\text{Male}} = \exp(-.12) = .887$$

3. Since y is distributed Poisson: $\Pr(y_i \mid \text{Female}_i) = e^{-\mu_i}\mu_i^{y_i} / y_i!$

	Women	Men
Pr(y=0)	.46	.41
Pr(y=1)	.36	.37
Pr(y=2)	.14	.16
Pr(y=3)	.04	.05
Pr(y=4)	.007	.011

PRM with a single continuous regressor

1. Let

$$\mu = \exp(-.25 + .13x)$$

2. Since y is Poisson

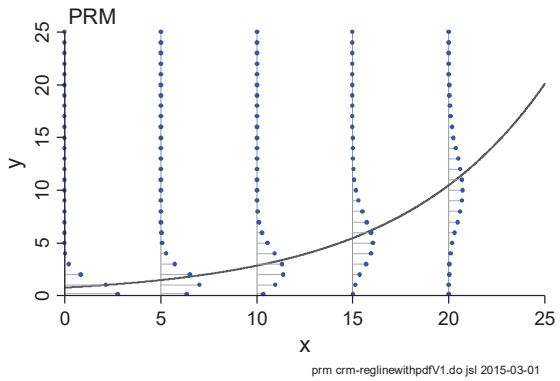
$$\Pr(y_i \mid \mathbf{x}_i) = e^{-\mu_i}\mu_i^{y_i} / y_i!$$

3. For example, at x=0, μ=exp(-.25)=.78

$$\Pr(y = 0 \mid \mu = .78) = .46$$
$$\Pr(y = 1 \mid \mu = .78) = .36$$
$$\Pr(y = 2 \mid \mu = .78) = .14$$
$$\Pr(y = 3 \mid \mu = .78) = .04$$

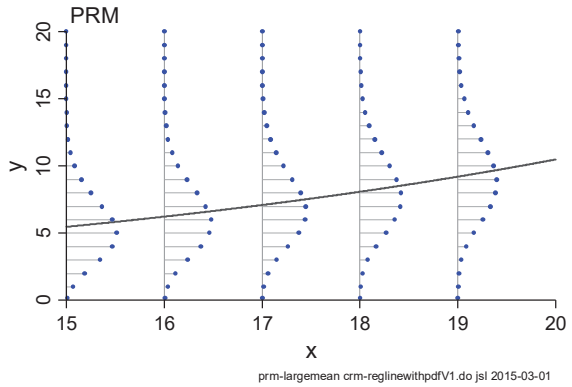
4. Plotting the counts by x...

PRM with a single continuous regressor



1. Distribution around E(y|**x**) is Poisson
2. As μ increases: a) Var(y|**x**) increases; b) proportion of 0s decreases; and 3) distribution becomes normal

PRM with a single continuous regressor at large mean counts



1. For larger E(y|x), LRM i approximates PRM
2. Is LRM acceptable with count outcomes?

LRM with count outcomes

1. Counts are sometimes treated as continuous using the LRM
2. LRM is inefficient, inconsistent, and biased due to nonlinearity and heteroscedasticity
3. LRM of sqrt(y) has some theoretical justification
4. With large mean counts, LRM often works well
5. Count models are preferred, easy to compute, and easy to interpret
   - Is income a count?

ML estimation

1. We maximize the likelihood

$$L(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{X}) = \prod_{i=1}^{N} \Pr(y_i \mid \mu_i)$$

2. Convergence is usually fast and problems are rarely encountered

Example of scientific productivity (-crm-couart.do)

Descriptive statistics

```
. use couart4, clear
(couart4.dta | Long data on productivity of biochemists | 2013-07-15)

. nmlab art fem mar kid5 phd ment

art   Articles in last 3 yrs of PhD
fem   Gender: 1=female 0=male
mar   Married: 1=yes 0=no
kid5  Number of children < 6
phd   PhD prestige
ment  Article by mentor in last 3 yrs

. sum art fem mar kid5 phd ment

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         art |       915    1.692896    1.926069          0         19
         fem |       915    .4601093    .4986788          0          1
         mar |       915    .6622951     .473186          0          1
        kid5 |       915     .495082      .76488          0          3
         phd |       915    3.103109    .9842491       .755       4.62
        ment |       915    8.767213    9.483916          0         77
```


#22 Poisson regression model

```
. poisson art i.fem i.mar kid5 phd ment

Poisson regression                                Number of obs   =        915
                                                  LR chi2(5)      =     183.03
                                                  Prob > chi2     =     0.0000
Log likelihood = -1651.0563                       Pseudo R2       =     0.0525

------------------------------------------------------------------------------
         art |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |
     Female  |  -.2245942   .0546138    -4.11   0.000    -.3316352   -.1175532
             |
     married |
    Married  |   .1552434   .0613747     2.53   0.011     .0349512    .2755356
        kid5 |  -.1848827   .0401272    -4.61   0.000    -.2635305   -.1062349
         phd |   .0128226   .0263972     0.49   0.627     -.038915    .0645601
      mentor |   .0255427   .0020061    12.73   0.000     .0216109    .0294746
       _cons |   .3046168   .1029822     2.96   0.003     .1027755    .5064581
------------------------------------------------------------------------------
```

1. All regressors but **phd** are significant

2. We begin by interpreting effects on rates

Factor change in the rate E(y|x)

1. The expected count focusing on the level of x_3 is

   Start at x_3: $\mathrm{E}(y \mid \mathbf{x}, x_3) = e^{\beta_0}\, e^{\beta_1 x_1} e^{\beta_2 x_2} e^{\beta_3 x_3}$

2. Increasing x_3 by 1

   End at x_3+1: $\mathrm{E}(y \mid \mathbf{x}, x_3 + 1) = e^{\beta_0}\, e^{\beta_1 x_1} e^{\beta_2 x_2} e^{\beta_3 x_3}\, e^{\beta_3 1}$

3. The factor change in the expected count is

$$\frac{\text{End level}}{\text{Start level}} = \frac{\mathrm{E}(y \mid \mathbf{x}, x_3 + 1)}{\mathrm{E}(y \mid \mathbf{x}, x_3)} = \frac{e^{\beta_0}\, e^{\beta_1 x_1} e^{\beta_2 x_2} e^{\beta_3 x_3}\, e^{\beta_3}}{e^{\beta_0}\, e^{\beta_1 x_1} e^{\beta_2 x_2} e^{\beta_3 x_3}} = e^{\beta_3}$$

4. The effect of a change in x_k does not depend on values of the x's

   Factor change: For a unit change in x_k the expected count (rate) changes by a factor of $\exp(\beta_k)$ holding other variables constant.

   Standardized factor change: For a SD change in x_k the expected count changes by a factor of $\exp(s_k\beta_k)$ holding other variables constant.

#22 Factor change in rate

```
. listcoef fem ment, help
poisson (N=915): Factor change in expected count
  Observed SD:  1.9261
             |          b        z    P>|z|       e^b   e^bStdX     SDofX
-------------+------------------------------------------------------------
      female |
     Female  |    -0.2246   -4.112    0.000     0.799     0.894     0.499
      mentor |     0.0255   12.733    0.000     1.026     1.274     9.484
-------------------------------------------------------------------------
     e^b = exp(b) = factor change in expected count for unit increase in X
 e^bStdX = exp(b*SD of X) = change in expected count for SD increase in X

. listcoef fem ment, percent help
poisson (N=915): Percentage change in expected count
  Observed SD:  1.9261
             |          b        z    P>|z|         %     %StdX     SDofX
-------------+------------------------------------------------------------
      female |
     Female  |    -0.2246   -4.112    0.000     -20.1     -10.6     0.499
      mentor |     0.0255   12.733    0.000       2.6      27.4     9.484
-------------------------------------------------------------------------
       % = percent change in expected count for unit increase in X
   %StdX = percent change in expected count for SD increase in X
```

Interpretations on next page...

female

```
             |          b        z    P>|z|       e^b   e^bStdX     SDofX
-------------+------------------------------------------------------------
      female |
     Female  |    -0.2246   -4.112    0.000     0.799     0.894     0.499
```

Being a female scientist decreases the expected number of articles by a factor of .80, holding other variables constant.

mentor

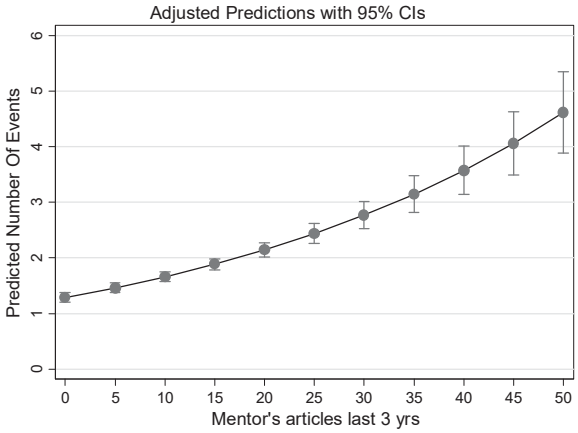
```
             |          b        z    P>|z|         %     %StdX     SDofX
-------------+------------------------------------------------------------
      mentor |     0.0255   12.733    0.000       2.6      27.4     9.484
```

For every additional article by the mentor, a scientist's mean productivity increases by 2.6 percent, holding other variables constant.

For a standard deviation increase in the mentor's productivity, about ten articles, a scientist's mean productivity increases by 27 percent, holding other variables constants.

#22 Plotting the rate

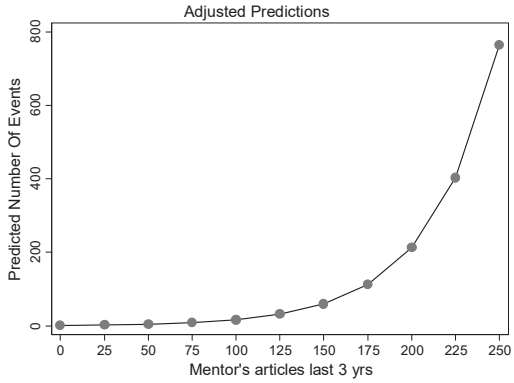
```
. margins, atmeans at(ment=(0(5)50))
. marginsplot, ylabel(0(1)6, grid gmin gmax)
```



Marginal effects on rates are not constant

The larger the rate, the greater the rate of change

$$\frac{\partial \mu_i}{\partial x_k} = \mu_i \beta_k$$



Discrete change in the rate

1. The factor change is constant
2. The DC depends on the values of the regressors
3. Let x_k change from **start value** to **end value**

$$\frac{\Delta \mathrm{E}\left(y \mid \mathbf{x}^*\right)}{\Delta x_k} = \mathrm{E}\left(y \mid \mathbf{x}^*, x_k = \text{end value}\right) - \mathrm{E}\left(y \mid \mathbf{x}^*, x_k = \text{start value}\right)$$

4. Interpretation

   For a change in x_k from **start value** to **end value**, the expected count changes by $\Delta \mathrm{E}\left(y \mid \mathbf{x}^*\right) / \Delta x_k$, holding other variables at the given values.

#23 DCM in the rate

```
. mchange female kid5, amount(one) atmean

poisson: Changes in mu | Number of obs = 915

Expression: Predicted number of art, predict()

                  |    Change    p-value
------------------+----------------------
female            |
    Female vs Male |    -0.359      0.000
kid5              |
               +1 |    -0.272      0.000
<snip>
1: Estimates with margins option atmeans
```

For an average scientist, being female decreases expected productivity by .37 articles.

For an average scientists, an additional young child decreases expected productivity by .29.

#25 DCM in rate at mean with CIs

1. I compute predictions at the mean for men and women

```
. qui mtable, at(fem=0) atmeans ci rowname(Men)
. qui mtable, at(fem=1) atmeans ci rowname(Women) below
```

2. The DC for female is:

```
. mtable, dydx(fem) atmeans ci rowname(Diff)  below
Expression: Predicted number of art, predict()
           |       mu        ll        ul
 ----------+---------------------------
       Men |    1.785     1.664     1.907
     Women |    1.426     1.310     1.542
      Diff |   -0.359    -0.529    -0.190

Specified values of covariates
           |                 1.                                      1.
           |   female    married      kid5       phd    mentor    female
 ----------+-------------------------------------------------------------
     Set 1 |        0       .662      .495       3.1      8.77         .
     Set 2 |        1       .662      .495       3.1      8.77         .
   Current |        .       .662      .495       3.1      8.77       .46
```

The expected number of publications for an average man is 1.78 compared to an average of 1.43 for women, a statistically significant difference of .36.

#23 Aside: ADC in the rate

```
. mchange female kid5, amount(one)

poisson: Changes in mu | Number of obs = 915

Expression: Predicted number of art, predict()

                |    Change    p-value
----------------+----------------------
female          |
 Female vs Male |    -0.375      0.000
kid5            |
             +1 |    -0.286      0.000
```

On average being a female scientist decreases expected productivity by .38 articles.

On average an additional young child decreases expected productivity by .29.

#24 Relationship between exp(β) and DCM

1. The factor or percent change and the DCM are related
2. A change of .36 articles from 1.79 to 1.42 is a **20.1** percent decrease
3. Computed with regression coefficient

```
. listcoef fem, percent

poisson (N=915): Percentage Change in Expected Count

       art |      b         z     P>|z|      %      %StdX      SDofX
-----------+--------------------------------------------------------
     1.fem |  -0.22459   -4.112   0.000    -20.1    -10.6     0.4987
-------------------------------------------------------------------
```

4. Computed with predictions

```
. qui mtable, at(fem=0) atmeans stat(est ll ub) rowname(Men)
. qui mtable, at(fem=1) atmeans stat(est ll ub) rowname(Women) below
. mtable, dydx(fem) atmeans stat(est ll ub) below rowname(Diff)

           |       mu        ll        ul
 ----------+----------------------------
       Men |    1.785     1.664     1.907
     Women |    1.426     1.310     1.542
      Diff |   -0.359    -0.529    -0.190

. di 100*(-0.359/1.785)
-20.112045
```


Probabilities

1. For a given the rate

$$\Pr(y_i = m \mid \mathbf{x}_i) = \frac{e^{-\mu_i}\mu_i^m}{m!}$$

2. Sometimes probabilities provide more substantively useful information than the mean count or rate.
3. Suppose we want to compare the productivity of men and women.
   - We know from above

```
. mtable, dydx(fem) atmeans ci rowname(Diff)  below
Expression: Predicted number of art, predict()
       |       mu        ll        ul
 ----------+----------------------------
       Men |    1.785     1.664     1.907
     Women |    1.426     1.310     1.542
      Diff |   -0.359    -0.529    -0.190
```

   - Can probabilities tell us more?

#27 Comparing probabilities for men and women

```
. qui mtable, at(fem=0) atmeans pr(0/5) stat(est)   clear roweq(Men)
. qui mtable, at(fem=1) atmeans pr(0/5) stat(est)   below roweq(Women)
. mtable,     dydx(fem) atmeans pr(0/5) stat(est p) below roweq(Change)

Expression: Marginal effect of Pr(art), predict(pr()) <= from last mtable

           |        0         1         2         3         4         5
 ----------+-------------------------------------------------------------
 Men       |
         1 |    0.168     0.299     0.267     0.159     0.071     0.025
 Women     |
         1 |    0.240     0.343     0.244     0.116     0.041     0.012
 Change    |
   d Pr(y) |    0.072     0.043    -0.023    -0.043    -0.030    -0.014
         p |    0.000     0.000     0.000     0.000     0.000     0.000
```

1. The largest difference is the greater probability of women having no articles
   - Does this make substantive sense?

Assessing models with average predictions

1. The probability that y=*m* given **x**

$$\Pr(y_i = m \mid \mathbf{x}_i) = \frac{e^{-\mathbf{x}_i\hat{\boldsymbol{\beta}}}\left(\mathbf{x}_i\hat{\boldsymbol{\beta}}\right)^m}{m!}$$

2. *Mean predicted probability* summarizes predictions

$$\overline{\widehat{\Pr}}(y = m) = \frac{1}{N}\sum_{i=1}^{N}\widehat{\Pr}(y_i = m \mid \mathbf{x}_i)$$

3. *Observed probability*

$$\Pr_{Observed}(y = m) = \frac{n(y = m)}{N}$$

4. If model is correct, we expect $\overline{\widehat{\Pr}}(y = m) \approx \Pr_{Observed}(y = m)$

5. This is how **mgen,meanpred** computes this informaton

What is the mean prediction?

#31 Predictions for 1st two observations

```
. list art fem mar kid5 phd ment in 1/2, nolabel clean

       art   fem   mar   kid5    phd   ment
  1.     0     0     1      0   2.52      7
  2.     0     1     0      0   2.05      6

. qui mtable, at(fem=0 mar=1 kid5=0 phd=2.52 ment=7) ///
>     pr(0/5) atmeans rowname(case1) colstub(pr)
. mtable, at(fem=1 mar=0 kid5=0 phd=2.05 ment=6) ///
>     pr(0/5) atmeans below rowname(case2) colstub(pr)
Expression: Pr(art), predict(pr())

        |      pr0       pr1       pr2       pr3       pr4       pr5
--------+------------------------------------------------------------
  case1 |    0.141     0.277     0.271     0.176     0.086     0.034
  case2 |    0.274     0.355     0.230     0.099     0.032     0.008

Specified values of covariates

           |   female   married      kid5       phd    mentor
 ----------+-----------------------------------------------
     Set 1 |        0         1         0      2.52         7
   Current |        1         0         0      2.05         6
```


#32 predict for all observation

```
. predict estpr0, pr(0)
. predict estpr1, pr(1)
. predict estpr2,pr(2)

. list estpr0-estpr2 art fem mar kid5 phd ment in 1/2, nolabel clean

       estpr0     estpr1     estpr2   art   fem   mar   kid5    phd   ment
  1.   .1414034   .2766047   .2705385     0     0     1      0   2.52      7
  2.   .2735238   .3545871   .2298374     0     1     0      0   2.05      6

. * average predictions for all observations

. sum estpr*

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      estpr0 |       915    .2092071    .0794247   .0000659   .4113403
      estpr1 |       915    .3098447    .0634931   .0006345   .3678775
      estpr2 |       915     .242096    .0311473   .0030544   .2706704
```

mgen,meanpred makes these computations automatically

#33 mgen for average predictions

```
. mgen, pr(0/9) meanpred stub(prm)

Variable   Obs Unique       Mean        Min       Max  Label
-------------------------------------------------------------------------------
prmval      10     10        4.5          0         9  Articles in last 3 ...
prmobeq     10     10   .0993443   .0010929  .3005464  Observed proportion
prmoble     10     10   .8328962   .3005464  .9934427  Observed cum. propo...
prmpreq     10     10   .0998819   .0009304  .3098447  Avg predicted Pr(y=#)
prmprle     10     10   .8308733   .2092071  .9988188  Avg predicted cum. P..
prmob_pr    10     10  -.0005376  -.0475604  .0913393  Observed - Avg Pr(y=#)
-------------------------------------------------------------------------------

. list art prmval prmpreq prmobeq in 1/12, nodisplay clean

       art   prmval    prmpreq    prmobeq
  1.     0        0   .2092071   .3005464
  2.     0        1   .3098447   .2688525
  3.     0        2   .242096    .1945355
  4.     0        3   .1346656   .0918033
  5.     0        4   .0611696    .073224
  6.     0        5   .0249554   .0295082
  7.     0        6   .0099346   .0185792
  8.     0        7   .0041384   .0131148
  9.     0        8    .001877   .0010929
 10.     0        9   .0009304   .0021858
 11.     0        .          .          .
```

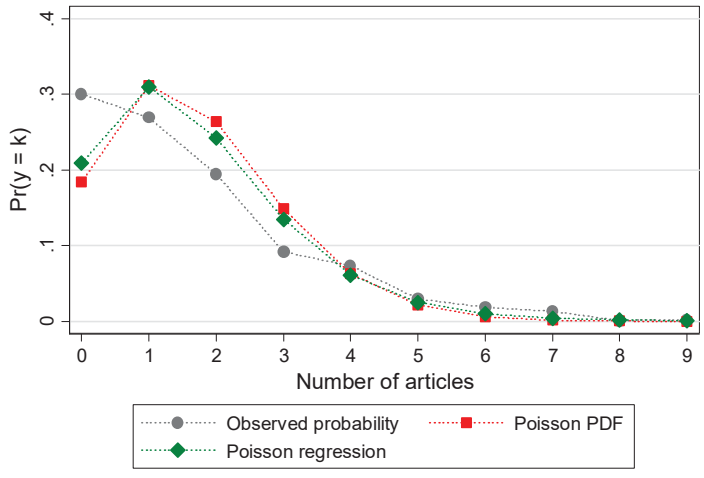

#34 distribution of observed counts

```
. tab art

Articles in |
 last 3 yrs |
     of PhD |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        275       30.05       30.05  <= see prmobeq from mgen
          1 |        246       26.89       56.94  <= see prmobeq from mgen
          2 |        178       19.45       76.39  <= see prmobeq from mgen
          3 |         84        9.18       85.57
          4 |         67        7.32       92.90
          5 |         27        2.95       95.85
          6 |         17        1.86       97.70
<snip>
         12 |          2        0.22       99.78
         16 |          1        0.11       99.89
         19 |          1        0.11      100.00
------------+-----------------------------------
      Total |        915      100.00
```

#35 plotting observed probabilities and mean predictions

```
. twoway connected prmobeq prmpreq prmval, ///
>     msym(O d) msiz(2 2.4) mcol(gs6 green) lcol(gs6 green) lpat(dot dot) ///
>     ytitle("Pr(y = k)") xtitle("Number of articles") ///
>     ylab(0(.1).4, grid gmax gmin) xlab(0(1)9, nogrid)
```


#35 Observed probabilities and average predictions from PRM



#36 cda13lec-crm-couart-prm-pdf-obs: cda13lec-crm-couart scott long 2013-09-27

Negative binomial regression model

1. The PRM rarely fits due to *overdispersion*
2. The model typically under-predicts 0's and over-predicts larger counts
3. If the *PRM mean structure is correct*, but there is over-dispersion
   a. Estimates are consistent, but inefficient
   b. *Z-values are spuriously large*; things appear significant that are not
4. In the PRM

$$Var(y \mid \mathbf{x}) = E(y \mid \mathbf{x}) = \exp(\mathbf{x}\boldsymbol{\beta})$$

5. The NBRM adds a parameter so that

$$Var(y \mid \mathbf{x}) > E(y \mid \mathbf{x})$$

6. The single parameter often makes a huge difference in fit

Unobserved heterogeneity

1. In the NBRM, a new source of error is added

$$\mu = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3) \quad \text{for the PRM}$$
$$\tilde{\mu} = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon) \quad \text{for the NBRM}$$

2. ε is assumed uncorrelated with **x**
3. ε can be due to
   a. Combined effects of excluded variables
   b. Pure randomness
4. $\tilde{\mu}$ from NBRM and μ from PRM are related

$$\begin{aligned}\tilde{\mu} &= \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3) \times \exp(\varepsilon) \\ &= \mu \times \exp(\varepsilon) \\ &= \mu\nu \quad \text{where } \nu \equiv \exp(\varepsilon)\end{aligned}$$

Identification

1. In the LRM, we assume E(ε)=0 to identify the model
2. We need a similar assumption for NBRM
3. Most *conveniently* assume

$$\mathrm{E}(\nu) = 1 = \mathrm{E}[\exp(\varepsilon)]$$

4. With this assumption, PRM and NBRM have *the same mean structure*

$$\mathrm{E}(\tilde{\mu}) = \mathrm{E}(\mu\nu) = \mu\mathrm{E}(\nu) = \mu$$

The negative binomial distribution

Roadmap

1. Show the distribution of y given both x and v.
2. Since v is unobserved, we *average over v* to compute distribution of y given x
3. This is results in the NB distribution

Poisson distribution given x and v

1. v is unobserved, but *assume* we know the value of v for each observation
2. Knowing v, we treat it as a regressor with β_v=1
3. The distribution of y given *both x and v* is Poisson (since v is just another x)

$$\Pr(y \mid \mathbf{x}, \nu) = \frac{e^{-\tilde{\mu}}\,\tilde{\mu}^{y}}{y!} = \frac{e^{-\mu\nu}\,(\mu\nu)^{y}}{y!}$$

4. However, v is unobserved so we cannot compute Pr(y|**x**, v)!
5. Instead, we compute Pr(y|**x**) by *mixing across v*

Example of binary mixing

1. Suppose:
   a. Let v = 1 for those with low motivation
   b. Let v = 2 for those with hi motivation
   c. Let Pr(v = v*) be the probability someone is in group v*
2. The distribution of y|**x** differs for the two groups
   a. Pr(y |**x**, v = 1)
   b. Pr(y |**x**, v = 2)
3. We mix these distributions by how frequently v = 1 and v = 2 occur

   Pr(y |**x**) = [Pr(v = 1) Pr(y |**x**, v = 1)]
   + [Pr(v = 2) Pr(y |**x**, v = 2)]

4. Pr(y|**x**) is a *mixture* of the distribution of y in the two groups

Continuous mixing (the Poisson-gamma mixture model)

1. If the mixing variable is continuous (infinite v groups)

$$\Pr(y \mid \mathbf{x}) = \int_0^{\infty} \left[\Pr(y \mid \mu, \nu) \times g(\nu) \right] d\nu$$

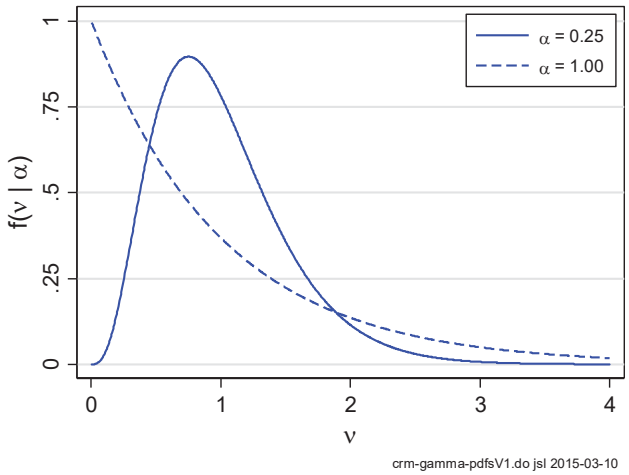
2. Assume v has a *gamma distribution* with parameter δ

$$g(\nu) = \frac{\delta^{\delta}}{\Gamma(\delta)} \nu^{\delta-1} e^{-\nu\delta} \quad \text{for } \delta > 0 \text{ where } \Gamma(\delta) = \int_0^{\infty} t^{\delta-1} e^{-t}\, dt$$

3. With the gamma distribution
   a. E(v) = 1
   b. Var(v) = 1/δ ≡ α.
4. The distribution varies in shape with changes in α

Graph on next page...

Gamma distributions with varying parameters



crm-gamma-pdfsV1.do jsl 2015-03-10

Negative binomial distribution

1. The mixture of Poisson PDFs by gamma is a *negative binomial distribution*

$$\Pr(y \mid \mathbf{x}) = \frac{\Gamma\left(y+\alpha^{-1}\right)}{y!\Gamma\left(\alpha^{-1}\right)}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}}\left(\frac{\mu}{\alpha^{-1}+\mu}\right)^{y}$$

2. It has the same mean structure as the PRM

$$\mathrm{E}(y \mid \mathbf{x}) = \exp(\mathbf{x}\boldsymbol{\beta}) = \mu$$

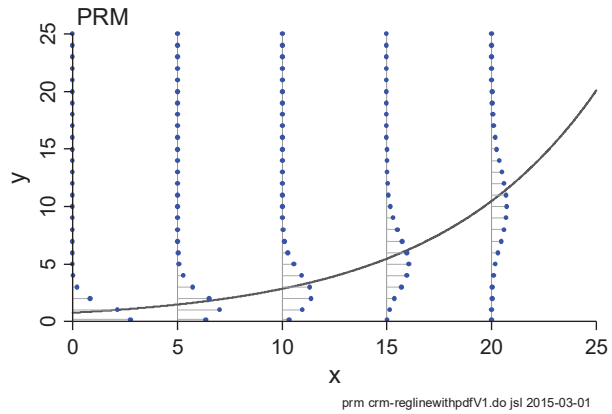
3. Where the variance is larger

$$\mathrm{Var}(y \mid \mathbf{x}) = \mu(1+\alpha\mu) = \exp(\mathbf{x}\boldsymbol{\beta})(1+\alpha\exp(\mathbf{x}\boldsymbol{\beta})) > \mu$$

   - Since μ and α are positive, the *NB has overdispersion*

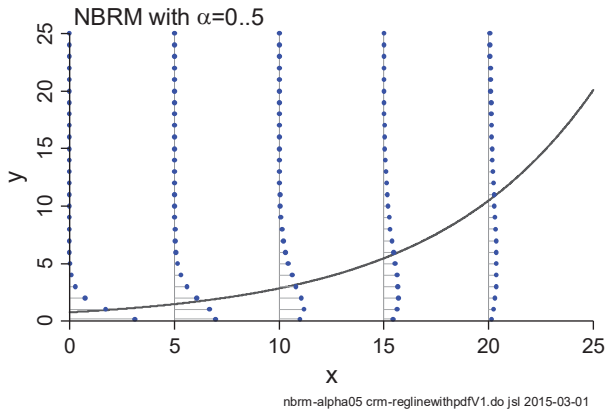
Graphs on next pages...

PRM or NBRM with α = 0.0



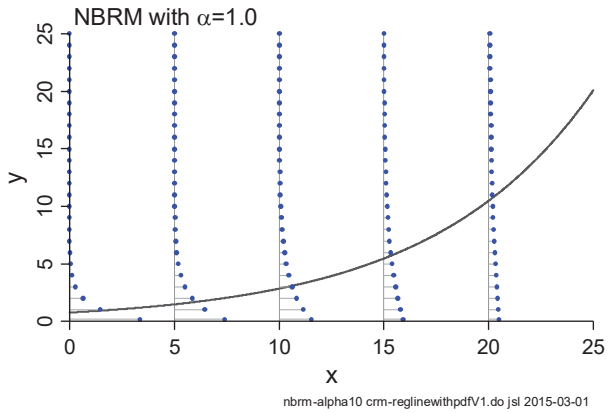
prm crm-reglinewithpdfV1.do jsl 2015-03-01

NBRM with α = .5



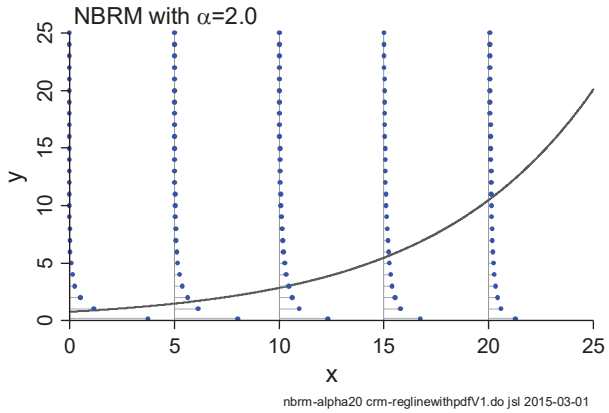
nbrm-alpha05 crm-reglinewithpdfV1.do jsl 2015-03-01

NBRM with α = 1.0



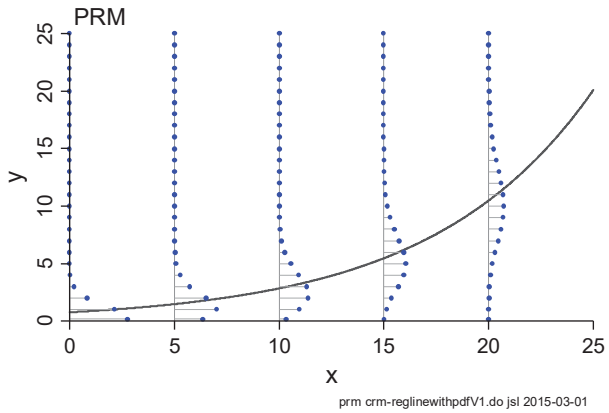
nbrm-alpha10 crm-reglinewithpdfV1.do jsl 2015-03-01

NBRM with α = 2.0



nbrm-alpha20 crm-reglinewithpdfV1.do jsl 2015-03-01

NBRM with α = 0.0



Heterogeneity and contagion

1. Our derivation is based on *unobserved heterogeneity* (v)
2. *Contagion* also leads to the NB distribution
   - Contagion is when people start with the same rate but the rate changes when an event occurs
3. Two, identical scientists start with the same productivity rate μ
   a. Success in publishing increases the rate of future publishing
   b. If scientist 1 publishes, her rate increases as the result of contagion
   c. Scientist 2's rate does not change
   d. Now scientist 1 has an advantage that will accumulate
4. Contagion violates the independence assumption of the PRM
5. With cross-sectional data, heterogeneity and contagion are indistinguishable

ML Estimation

The NBRM model can be estimated by ML

$$L(\boldsymbol{\beta}\,|\,\mathbf{y},\mathbf{X}) = \prod_{i=1}^{N} \Pr(y_i\,|\,\mathbf{x}_i)$$

$$= \prod_{i=1}^{N} \frac{\Gamma\left(y_i+\alpha^{-1}\right)}{y_i!\,\Gamma\left(\alpha^{-1}\right)}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_i}\right)^{\alpha^{-1}}\left(\frac{\mu_i}{\alpha^{-1}+\mu_i}\right)^{y_i}$$

where $\mu = \exp(\mathbf{x}\boldsymbol{\beta})$.

#41 The NBRM for articles

```
. nbreg art i.fem i.mar kid5 phd ment, nolog

Negative binomial regression                    Number of obs   =        915
                                                LR chi2(5)      =      97.96
Dispersion     = mean                           Prob > chi2     =     0.0000
Log likelihood = -1560.9583                     Pseudo R2       =     0.0304

------------------------------------------------------------------------------
         art |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.fem |  -.2164184   .0726724    -2.98   0.003    -.3588537   -.0739832
       1.mar |   .1504895   .0821063     1.83   0.067    -.0104359    .3114148
        kid5 |  -.1764152   .0530598    -3.32   0.001    -.2804105     -.07242
         phd |   .0152712   .0360396     0.42   0.672    -.0553652    .0859075
        ment |   .0290823   .0034701     8.38   0.000     .0222811    .0358836
       _cons |    .256144   .1385604     1.85   0.065    -.0154294    .5277174
-------------+----------------------------------------------------------------
    /lnalpha |  -.8173044   .1199372                     -1.052377   -.5822318
-------------+----------------------------------------------------------------
       alpha |   .4416205   .0529667                      .3491069    .5586502
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) =  180.20 Prob>=chibar2 = 0.000
```

Clicking on the blue...

What is chibar2? The likelihood-ratio (LR) test that is displayed is testing on the boundary of the parameter space. You are probably testing whether an estimated variance component (something that is always greater than zero) is different from zero by using an LR test.

Suppose for now that the two models being compared differ only with respect to the variance component in question, in which case the test statistic will be displayed as "chibar(01)". In such cases, the limiting distribution of the maximum-likelihood estimate of the parameter in question is a normal distribution that is halved, or chopped off at the boundary -- zero here. The distribution of the LR test statistic is therefore not the usual chi-squared with 1 degree of freedom but is instead a 50:50 mixture of a chi-squared with no degrees of freedom (that is, a point mass at zero) and a chi-squared with 1 degree of freedom.

The p-value of the LR test takes this into account and will be set to 1 if it is determined that your estimate is close enough to zero to be, in effect, zero for purposes of significance. Otherwise, the p-value displayed is set to one-half of the probability that a chi-squared with 1 degree of freedom is greater than the calculated LR test statistic.

Sometimes you are testing whether a variance component is zero in addition to testing whether k other parameters (not affected by boundary conditions) are zero. Such situations often arise when comparing mixed-effects models, such as those fit by xtmixed. For such tests, the distribution of the likelihood-ratio test statistic is a 50:50 mixture of chi-squared distributions with k and k+1 degrees of freedom, shown on the output as "chibar(4_5)", for example. As for chibar(01), significance levels are adjusted accordingly.

Finally, if you are testing more than one boundary-affected parameter, the theory is much more complex and usually intractable. When this occurs, Stata will either display significance levels that are conservative and marked as such or will not display an LR test at all.

#42 Comparing PRM and NBRM

```
. estimates table prm nbrm, stats(N bic r2_p) b(%9.3f) t(%6.2f) eform

-------------------------------------
    Variable |    prm        nbrm
-------------+-----------------------
      fem 1  |     0.799       0.805
             |     -4.11       -2.98
      mar 1  |     1.168       1.162
             |      2.53        1.83
        kid5 |     0.831       0.838
             |     -4.61       -3.32
         phd |     1.013       1.015
             |      0.49        0.42
        ment |     1.026       1.030
             |     12.73        8.38
       _cons |     1.356       1.292
             |      2.96        1.85
-------------+-----------------------
lnalpha_cons |                 0.442
             |                -6.81
-------------+-----------------------
Statistics   |
           N |       915         915
         bic |  3343.026    3169.649
        r2_p |     0.053       0.030
-------------------------------------
                          legend: b/t
```

Testing for overdispersion

1. With overdispersion PRM estimates are *inefficient with standard errors that are biased downward.*
2. We test H_0: α=0 since when α=0 the NBRM becomes the PRM
3. To estimate the NBRM, Stata maximizes a function with *ln α*, not with α
   - Testing H_0: ln(α)=0 is equivalent to H_0: α=1 which we do not want
4. A LR test of H_0: α=0 is

$$G^2 = 2\left(\ln L_{\text{NBRM}} - \ln L_{\text{PRM}}\right)$$

5. In this example, there is strong evidence of overdispersion:

```
Likelihood-ratio test of alpha=0: chibar2(01) = 180.20 Prob>=chibar2 = 0.000
```

Since there is significant evidence of overdispersion ($G^2(1)=180.2$, $p<.001$), the NBRM is preferred to the PRM.

Interpretation of the NBRM

1. Interpretation based on rates is identical to the PRM
2. The same methods for predicted probabilities can be used where

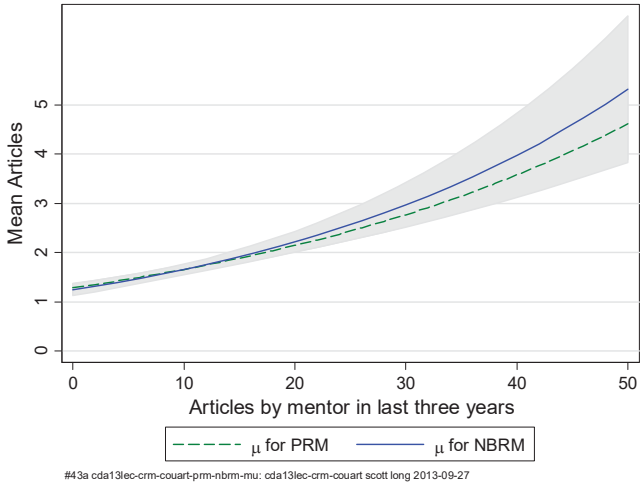
$$\Pr(y \mid \mathbf{x}) = \frac{\Gamma\left(y+\hat{\alpha}^{-1}\right)}{y!\,\Gamma\left(\hat{\alpha}^{-1}\right)}\left(\frac{\hat{\alpha}^{-1}}{\hat{\alpha}^{-1}+\hat{\mu}}\right)^{\hat{\alpha}^{-1}}\left(\frac{\hat{\mu}}{\hat{\alpha}^{-1}+\hat{\mu}}\right)^{y}$$

Comparing PRM and NBRM using mgen

#43 Comparing rates

```
. estimates restore nbrm
. mgen, at(ment=(0(2)50)) atmeans stub(NB)

Predictions from: margins, at(ment=(0(2)50)) atmeans

Variable   Obs Unique      Mean       Min       Max  Label
------------------------------------------------------------------------------
NBmu        26     26   2.82001  1.241455  5.314301  mean art from margins
NBll        26     26  2.333907  1.122784  3.825262  95% lower limit
NBul        26     26  3.306114  1.360127   6.80334  95% upper limit
NBment      26     26        25         0        50  Mentor's arts last 3...
------------------------------------------------------------------------------

Specified values of covariates

        1.         1.
   female    married       kid5        phd
-------------------------------------------
 .4601093   .6622951    .49508    3.103109

. estimates restore prm
. mgen, at(ment=(0(2)50)) atmeans stub(PR)
<snip>
```

Comparing rates for PRM and NBRM



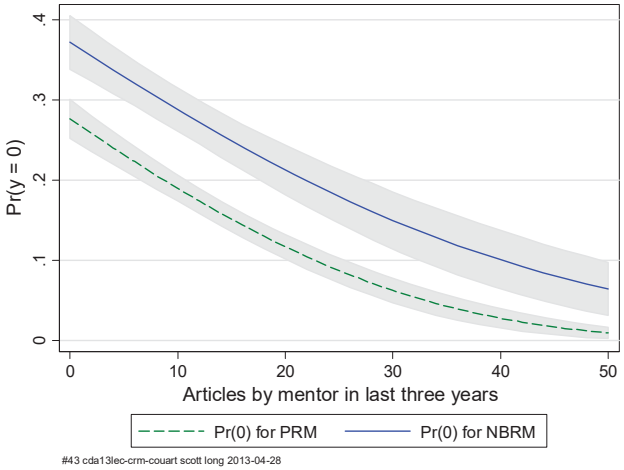
#43a cda13lec-crm-couart-prm-nbrm-mu: cda13lec-crm-couart scott long 2013-09-27

#44 Comparing probabilities of 0

```
. estimates restore nbrm
. mgen, at(ment=(0(2)50)) pr(0/9) atmeans stub(NB)

Predictions from: margins, at(ment=(0(2)50)) atmeans predict(pr(9))

Variable Obs Unique      Mean       Min       Max  Label
------------------------------------------------------------------------------
NBpr0     26     26  .1939652  .0648641   .371642  pr(y=0) from margins
NBll0     26     26  .1615177  .0318453  .3382501  95% lower limit
NBul0     26     26  .2264127  .0978828  .4050339  95% upper limit
NBment    26     26        25         0        50  Mentor's articles last 3...
<snip>
------------------------------------------------------------------------------

Specified values of covariates

        1.         1.
   female    married       kid5        phd
-------------------------------------------
 .4601093   .6622951    .495082   3.103109

estimates restore prm
. mgen, at(ment=(0(2)50)) pr(0/9) atmeans stub(PR)
<snip>
```

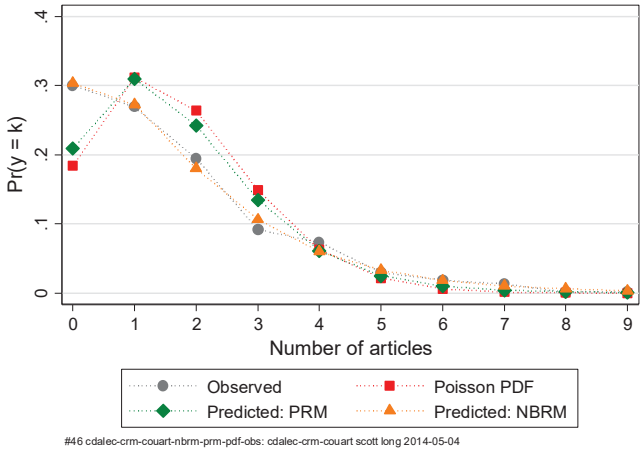

Probabilities of no articles for PRM and NBRM

Which model makes the most substantive sense?



#43 cda13lec-crm-couart scott long 2013-04-28

Average predictions



Zero modified count models

1. NBRM increases 0 over PRM by increasing the conditional variance
2. PRM and NBRM assume every scientist has a positive probability of publishing
3. In zero modified models
   a. Zeros
      - Some people *always have 0 counts*; some *might have 0 counts*
      - This leads to a greater proportion of 0s since they occur two ways
      - We say, "The zeros are inflated"
      - The variance increases since mass is added to the end of the PDF
   b. *Positive counts* are generated by *one process* just like PRM or NBRM

The with zeros model: didactically useful

1. Zeroes occur through two processes

   Group A: people who *always* have 0 counts

   Group S: people who *sometimes* have 0s, but *sometimes not*

2. Probabilities of being in each group

   Pr(Group A)= ψ

   Pr(Group S)=1-ψ

3. An observed 0 could come from either group
4. Probabilities of counts by group

   Group A: $\Pr(y = 0 \mid \mathbf{x}) = 1$

   Group S: $\Pr(y \mid \mathbf{x}) = \dfrac{e^{-\mu}\mu^{y}}{y!}$ where $\mu = \exp(\mathbf{x}\boldsymbol{\beta})$

   $\Pr(y = 0 \mid \mathbf{x}) = \exp(-\mu)$

5. This is discrete, unobserved heterogeneity.

6. *Total probability of 0* mixes the two sources of 0's

$$\begin{aligned}\Pr(y=0\,|\,\mathbf{x}) &= \{\Pr(\text{Group A})\times 1\}+\{\Pr(\text{Group S})\times PRM(0\,|\,\text{Group S})\}\\ &= \{\psi\times 1\}+\{(1-\psi)\times e^{-\mu_i}\}\\ &= \psi+(1-\psi)e^{-\mu_i}\end{aligned}$$

7. The Poisson process applies to those in S

   Pr(Group S)=1-ψ

8. The probability of positive counts is adjusted

$$\Pr(y\,|\,\mathbf{x}) = (1-\psi)\frac{e^{-\mu}\mu^{y}}{y!} \quad \text{for } y>0$$

   - I could use a NB distribution in stead of a Poisson distribution

Zero inflated models model ψ

Next we model whether a person is in A or S

Step 1: Model group membership as a BRM

$$\Pr(\textit{Group A}\,|\,\mathbf{z}_i) = \Pr(\textit{Always 0}\,|\,\mathbf{z}_i) = \psi_i = F(\mathbf{z}_i\boldsymbol{\gamma})$$

where *F*() is logistic or normal.

Step 2: Model counts in Group S as PRM or NBRM

1. Zero inflated Poisson (ZIP) model

$$\Pr(y\,|\,\mathbf{x}\,\&\,\text{Group S}) = \frac{e^{-\mu}\mu^{y}}{y!} \quad \text{where } \mu=\exp(\mathbf{x}\boldsymbol{\beta})$$

2. Zero inflated NB (ZINB) model

$$\Pr(y\,|\,\mathbf{x}\,\&\,\text{Group S}) = \frac{\Gamma(y+\alpha^{-1})}{y!\,\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}}\left(\frac{\mu}{\alpha^{-1}+\mu}\right)^{y}$$

Step 3: Model counts in Group A (always 0)

$$\Pr(y=0\,|\,\mathbf{x}\,\&\,\text{Group A}) = 1$$

Step 4: Mixing Group A and Group S

Predicted probabilities

$$\Pr(y=0\,|\,\mathbf{x}) = (1-\psi)\times\Pr(y=0\,|\,\mathbf{x}\,\&\,\text{Group S})+\psi$$

$$\Pr(y\,|\,\mathbf{x}) = (1-\psi)\times\Pr(y\,|\,\mathbf{x}\,\&\,\text{Group S})$$

Rates

$$\begin{aligned}E(y\,|\,\mathbf{x},\mathbf{z}) &= \{0\times\Pr(\text{Group A})\}+\{\mu\times\Pr(\text{Group S})\}\\ &= \{0\times\psi\}+\{\mu(1-\psi)\}\\ &= \mu(1-\psi)<\mu\end{aligned}$$

Variance for the ZIP

1. If ψ =0, we have the standard PRM/NBRM
2. If ψ>0 the dispersion is greater than for the PRM/NBRM

Two types of zeroes

1. Zeros come from two groups

$$\Pr(y=0\,|\,\mathbf{x}) = (1-\psi)\Pr(y=0\,|\,\mathbf{x}\ \&\ \text{Group S}) + \psi$$

2. Splitting the groups

$$\Pr(\text{always } 0\,|\,\mathbf{x}) = \psi$$

$$\Pr(\text{sometimes } 0\,|\,\mathbf{x}) = (1-\psi)\Pr(y=0\,|\,\mathbf{x}\ \&\ \text{Group S})$$

3. For *ZIP*:

$$\Pr(y=0\,|\,\mathbf{x}) = \psi + (1-\psi)e^{-\mu}$$

$$\Pr(y\,|\,\mathbf{x}) = (1-\psi)\frac{e^{-\mu_i}\mu^y}{y!} \quad \text{for } y>0$$

4. For *ZINB*

$$\Pr(y=0\,|\,\mathbf{x}) = \psi + (1-\psi)\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_i}\right)^{\alpha^{-1}}$$

$$\Pr(y\,|\,\mathbf{x}) = (1-\psi)\frac{\Gamma(y+\alpha^{-1})}{y!\,\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}}\left(\frac{\mu}{\alpha^{-1}+\mu}\right)^{y} \quad \text{for } y>0$$

Example of scientific productivity

#51 Estimation of ZIP

```
. zip art i.fem i.mar kid5 phd ment, inflate(i.fem i.mar kid5 phd ment) nolog
Zero-inflated Poisson regression                Number of obs   =        915
                                                Nonzero obs     =        640
                                                Zero obs        =        275
Inflation model = logit                         LR chi2(5)      =      78.56
Log likelihood  = -1604.773                     Prob > chi2     =     0.0000
------------------------------------------------------------------------------
         art |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
art          |
       1.fem |  -.2091446   .0634047    -3.30   0.001    -.3334155   -.0848737
       1.mar |    .103751    .071111     1.46   0.145     -.035624     .243126
        kid5 |  -.1433196   .0474293    -3.02   0.003    -.2362793   -.0503599
         phd |  -.0061662   .0310086    -0.20   0.842     -.066942    .0546096
        ment |   .0180977   .0022948     7.89   0.000     .0135999    .0225955
       _cons |    .640839   .1213072     5.28   0.000     .4030814    .8785967
-------------+----------------------------------------------------------------
inflate      |
       1.fem |   .1097465   .2800813     0.39   0.695    -.4392028    .6586958
       1.mar |  -.3540107   .3176103    -1.11   0.265    -.9765155    .2684941
        kid5 |   .2171001    .196481     1.10   0.269    -.1679956    .6021958
         phd |   .0012702   .1452639     0.01   0.993    -.2834418    .2859821
        ment |   -.134111   .0452461    -2.96   0.003    -.2227918   -.0454302
       _cons |  -.5770618   .5093853    -1.13   0.257    -1.575439     .421315
------------------------------------------------------------------------------
```

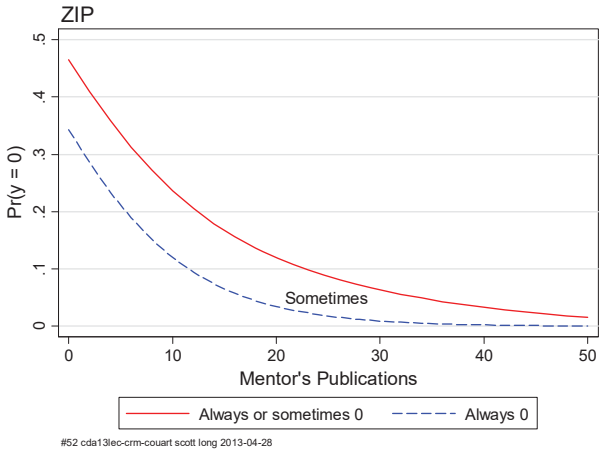
#51 factor change coefficients for ZIP

```
listcoef, help

zip (N=915): Factor change in expected count

  Observed SD:  1.9261

Count equation: Factor change in expected count for those not always 0

------------------------------------------------------------------------------
             |          b         z     P>|z|       e^b   e^bStdX     SDofX
-------------+----------------------------------------------------------------
      female |
      Female |    -0.2091    -3.299     0.001     0.811     0.901     0.499
             |
     married |
     Married |     0.1038     1.459     0.145     1.109     1.050     0.473
        kid5 |    -0.1433    -3.022     0.003     0.866     0.896     0.765
         phd |    -0.0062    -0.199     0.842     0.994     0.994     0.984
      mentor |     0.0181     7.886     0.000     1.018     1.187     9.484
    constant |     0.6408     5.283     0.000         .         .         .
------------------------------------------------------------------------------
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
     e^b = exp(b) = factor change in expected count for unit increase in X
 e^bStdX = exp(b*SD of X) = change in expected count for SD increase in X
   SDofX = standard deviation of X
```

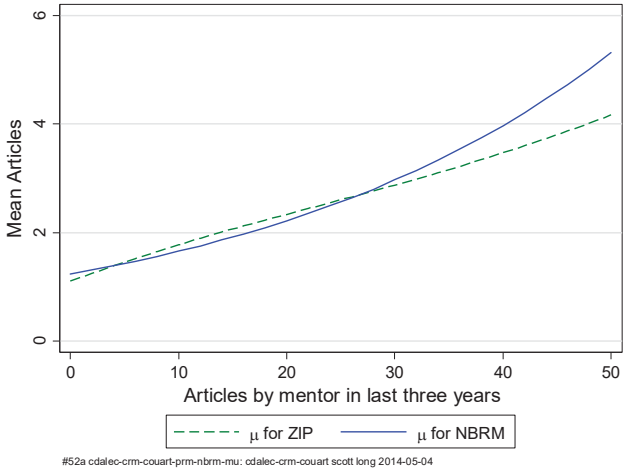


```
Binary equation: factor change in odds of always 0
------------------------------------------------------------------------------
             |          b         z     P>|z|       e^b   e^bStdX     SDofX
-------------+----------------------------------------------------------------
      female |
      Female |     0.1097     0.392     0.695     1.116     1.056     0.499
             |
     married |
     Married |    -0.3540    -1.115     0.265     0.702     0.846     0.473
        kid5 |     0.2171     1.105     0.269     1.242     1.181     0.765
         phd |     0.0013     0.009     0.993     1.001     1.001     0.984
      mentor |    -0.1341    -2.964     0.003     0.874     0.280     9.484
    constant |    -0.5771    -1.133     0.257         .         .         .
------------------------------------------------------------------------------
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
     e^b = exp(b) = factor change in odds for unit increase in X
 e^bStdX = exp(b*SD of X) = change in odds for SD increase in X
   SDofX = standard deviation of X
```


#52 Plotting sometimes 0's and total 0's from ZIP



#52 Comparing mean rates for ZIP and NBRM



Comparisons among count models

Count models we have considered

PRM: Poisson regression

NBRM: Negative binomial regression model

ZIP: Zero inflated Poisson model

ZINB: Zero inflated negative binomial model

countfit

1. This program that was so successful that SAS released **`countreg`** without citing **`countfit`**, although used my dataset as an example!
2. P. Trivedi suggested they change their name to **`count(er)fit`**.

Comparing mean observed probabilities

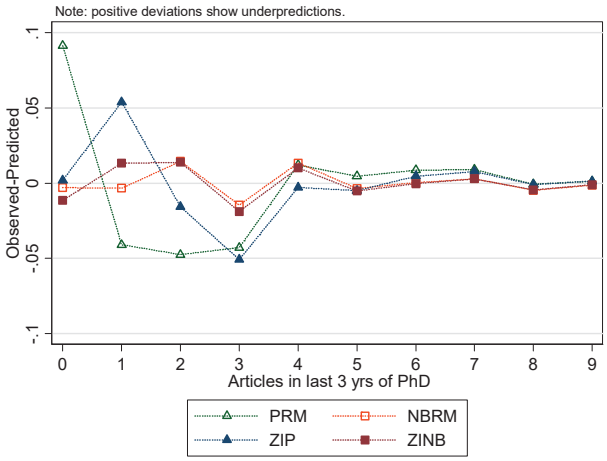
1. The mean predicted probability at observed values is

$$\overline{\Pr}(y=m)=\frac{1}{N}\sum_{i=1}^{N}\Pr(y_i=m\,|\,\mathbf{x}_i)$$

2. We plot the difference between observed and mean probability

$$\left(\text{Observed - Mean predicted}\right)=\bar{y}-\overline{\Pr}(y=m)$$

Lambert plot: summarizing mean predictions



Tests to compare count models

#62 LR tests of nested models

1. *PRM* vs *NBRM*: Test the dispersion parameter α

   We found that α was significant (G^2=180.2) supporting NBRM over PRM

2. *ZIP* vs *ZINB* can be compared the same way

```
. qui zip art fem mar kid5 phd ment, inflate(fem mar kid5 phd ment)
. est store zip
. qui zinb art fem mar kid5 phd ment, inf(fem mar kid5 phd ment)
. est store zinb

. lrtest zip zinb, force

Likelihood-ratio test                                 LR chi2(1)  =    109.56
(Assumption: zip nested in zinb)                      Prob > chi2 =    0.0000
```

There is evidence that the ZINB improves the fit over the ZIP model.

#63 Vuong test of non-nested models

1. PRM and ZIP are not nested; NBRM and ZINB are not nested.
   a. For ZIP to reduce to PRM, ψ must equal 0
   b. You cannot constrain the parameters to make ψ=0.
   c. If **γ=0**, then ψ=F(**z0**)=.5
2. A Vuong test is used instead of a LR test to compare the models

Details on next page...

3. The *Vuong test* compares non-nested models

a. $\Pr_1(y_i \mid \mathbf{x}_i)$ is probability of y_i from the first model

b. $\Pr_2(y_i \mid \mathbf{x}_i)$ is probability for the second model

c. Define

$$m_i = \ln\left[\frac{\Pr_1(y_i \mid \mathbf{x}_i)}{\Pr_2(y_i \mid \mathbf{x}_i)}\right]$$

d. Let $\overline{m} = \sum_{i=1}^{N} \frac{m_i}{N}$ and s_m be the standard deviation of m_i

e. The Vuong statistic is

$$V = \frac{\sqrt{N}\,\overline{m}}{s_m} \overset{a}{\sim} N.$$

- If $V > 1.96$, the first model is favored
- If $V < -1.96$, the second model is favored

For example...


```
. zip art i.fem mar kid5 phd ment, ///
>     inflate(i.fem mar kid5 phd ment) vuong nolog
  <snip>

Vuong test of zip vs. standard Poisson:            z =     4.18  Pr>z = 0.0000

. zinb art i.fem mar kid5 phd ment, ///
>     inflate(i.fem mar kid5 phd ment) vuong nolog
  <snip>

Vuong test of zinb vs. standard negative binomial: z =     2.24  Pr>z = 0.0125
```

`countfit` makes it easier to interpret the results

#63 countfit

```
. countfit art i.fem i.mar kid5 phd ment, ///
>          inf(i.fem i.mar kid5 phd ment)

-------------------------------------------------------------------------------
                     Variable |     PRM        NBRM         ZIP        ZINB
------------------------------+------------------------------------------------
art                           |
                       female |
                       Female |     0.799       0.805       0.811       0.822
                              |     -4.11       -2.98       -3.30       -2.59
                      married |
                      Married |     1.168       1.162       1.109       1.103
                              |      2.53        1.83        1.46        1.16
                # of kids < 6 |     0.831       0.838       0.866       0.859
                              |     -4.61       -3.32       -3.02       -2.80
                 PhD prestige |     1.013       1.015       0.994       0.999
                              |      0.49        0.42       -0.20       -0.02
       Mentor's # of articles |     1.026       1.030       1.018       1.025
                              |     12.73        8.38        7.89        7.10
                     Constant |     1.356       1.292       1.898       1.517
                              |      2.96        1.85        5.28        2.90
------------------------------+------------------------------------------------
lnalpha                       |
                     Constant |                 0.442                   0.377
                              |                -6.81                   -7.21  .
:: plus inflation results ::
```

And so on for all models...

```
Comparison of Mean Observed and Predicted Count

            Maximum       At      Mean
Model     Difference    Value    |Diff|
-----------------------------------------
PRM          0.091         0      0.026
NBRM        -0.015         3      0.006
ZIP          0.054         1      0.015
ZINB        -0.019         3      0.008

PRM: Predicted and actual probabilities

Count   Actual    Predicted    |Diff|   Pearson
-------------------------------------------------
0        0.301       0.209      0.091    36.489
1        0.269       0.310      0.041     4.962
2        0.195       0.242      0.048     8.549
3        0.092       0.135      0.043    12.483
4        0.073       0.061      0.012     2.174
5        0.030       0.025      0.005     0.760
6        0.019       0.010      0.009     6.883
7        0.013       0.004      0.009    17.815
8        0.001       0.002      0.001     0.300
9        0.002       0.001      0.001     1.550
-------------------------------------------------
Sum      0.993       0.999      0.259    91.964
```

And so on for all models...


```
Tests and Fit Statistics

PRM            BIC=  3343.026  AIC=  3314.113  Prefer  Over  Evidence
-----------------------------------------------------------------------------
  vs NBRM      BIC=  3169.649  dif=   173.377  NBRM    PRM   Very strong
               AIC=  3135.917  dif=   178.196  NBRM    PRM
               LRX2=  180.196  prob=    0.000  NBRM    PRM   p=0.000
-----------------------------------------------------------------------------
  vs ZIP       BIC=  3291.373  dif=    51.653  ZIP     PRM   Very strong
               AIC=  3233.546  dif=    80.567  ZIP     PRM
               Vuong=   4.180  prob=    0.000  ZIP     PRM   p=0.000
-----------------------------------------------------------------------------
  vs ZINB      BIC=  3188.628  dif=   154.398  ZINB    PRM   Very strong
               AIC=  3125.982  dif=   188.131  ZINB    PRM
-----------------------------------------------------------------------------
NBRM           BIC=  3169.649  AIC=  3135.917  Prefer  Over  Evidence
-----------------------------------------------------------------------------
  vs ZIP       BIC=  3291.373  dif=  -121.724  NBRM    ZIP   Very strong
               AIC=  3233.546  dif=   -97.629  NBRM    ZIP
-----------------------------------------------------------------------------
  vs ZINB      BIC=  3188.628  dif=   -18.979  NBRM    ZINB  Very strong
               AIC=  3125.982  dif=     9.935  ZINB    NBRM
               Vuong=   2.242  prob=    0.012  ZINB    NBRM  p=0.012
-----------------------------------------------------------------------------
ZIP            BIC=  3291.373  AIC=  3233.546  Prefer  Over  Evidence
-----------------------------------------------------------------------------
  vs ZINB      BIC=  3188.628  dif=   102.745  ZINB    ZIP   Very strong
               AIC=  3125.982  dif=   107.564  ZINB    ZIP
               LRX2=  109.564  prob=    0.000  ZINB    ZIP   p=0.000
```


Commands and model extensions

1. **predict** after count-data models, such as **gnbreg**, **nbreg**, **poisson**, **xtgee**, **xtnbreg**, **xtpoisson**, **zinb**, and **zip** has two new options

   predict *varname*, [**pr(**n**) pr(**a,b**)**]

   a. **pr(**n**)** stores the probability Pr(y = n) in *varname.*

   b. **pr(**a**,** b**)** stores the probability Pr(a<y<b) for *varname*.

2. **margins** and **m*** commands compute predicted rates and probabilities
3. **tnbreg** is for the truncated negative binomial regression for any nonnegative truncation point; **tpoisson** is for truncated Poisson regression
4. Mixed models for counts were added to Stata 13
5. Exposure time can be added to include the amount of time each case is "at risk" of the event occurring.

*Finite Mixture Models

fmm by Partha Deb fits a finite mixture regression model using ML. maximum likelihood estimation. In Stata, **findit fmm**

1. *Unobserved discrete heterogeneity*: Assume two or more types of people in the population but you do not know which group a person is in
2. Suppose we have groups A and B
3. For Group A

$$\Pr(y_i \mid \mathbf{x}_i, \text{Group}_i = A) = \frac{e^{-\mu_{Ai}}\mu_{Ai}^{y_i}}{y_i!} \quad \text{where} \quad \mu_A = \exp(\mathbf{x}\boldsymbol{\beta}_A)$$

4. For Group B

$$\Pr(y_i \mid \mathbf{x}_i, \text{Group}_i = B) = \frac{e^{-\mu_{Bi}}\mu_{Bi}^{y_i}}{y_i!} \quad \text{where} \quad \mu_B = \exp(\mathbf{x}\boldsymbol{\beta}_B)$$

5. The parameters can be interpreted for each groups

6. The observed counts come from both sources

$$\Pr(y_i \mid \mathbf{x}_i) = \left[\Pr(\text{Group}_i = A)\frac{e^{-\mu_{iA}}\mu_{iA}^{y_i}}{y_i!}\right] + \left[\Pr(\text{Group}_i = B)\frac{e^{-\mu_{iB}}\mu_{iB}^{y_i}}{y_i!}\right]$$

7. When would a mixture model make sense? Why do we need latent classes?

Review of count LHS

1. The PRM is rarely appropriate.
2. Start with NBRM since you use NBRM to test if PRM is appropriate.
3. Inflated models deal with "excess" zeros.
4. Statistical and ad hoc tests help select a model, but knowing what makes substantive sense is essential
   - If you don't think, you will use *ZIP* when you need a *1IP* model!
5. Other models deal with hurdles, mixtures, truncation and censoring.

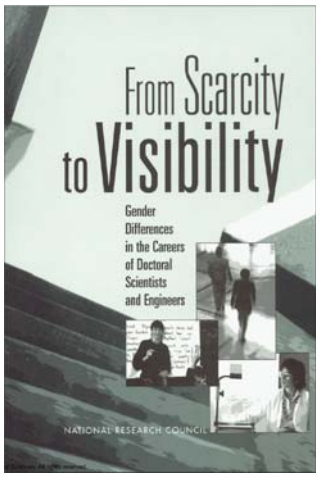
Part 15: Conclusions

1. LRM, BRM, ORM, MNLM, PRM, NBRM, and ZIP/ZINB are building blocks for models in many areas.
   - o Extensions add panels, hierarchies, clustering, survey sampling, and more
   - o The *basic structure* of the models stays the same
2. Since the models are nonlinear, the challenge is to determine what is substantively important and to find the best way to summarize the results
   - o Alternative strategies need to be tried to find the most convincing approach
3. Remember what Neal Henry told me many years ago:

   Don't let the numbers get in the way of the data.

   Think about what you want to know, then focus on answering that question.

** Part 9: Comparing groups



A motivating example

Are the "**effects**" of scientific productivity on the probability of tenures the same for men and women?

Read and run

Long 2009; Long & Mustillo 2017

cdalec*.do cdalec*-brmgroups-tenure-.do

Statistical and substantive issues

1. *Traditional LRM* approach for comparing groups
   a. Estimate model for women.
   b. Estimate same model for men.
   c. Compare coefficients across groups.
2. *Substantive concern*: Academy panel did not find logit coefficients informative.
3. *Statistical problem*: Paul Allison sent me a working paper that said:

   Differences in the estimated coefficients *tell us nothing* about the differences in the underlying impact of [publications] on [tenure for] the two groups.

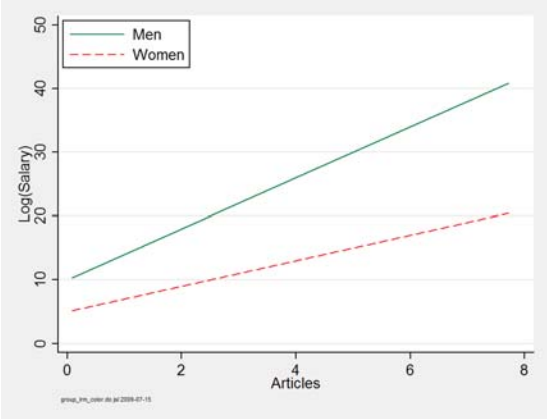
Roadmap

1. *Linear regression model* for group comparisons
2. *Binary regression model*: review of Chapter 3
   a. A latent variable model for y*
   b. A probability model for Pr(y=1|x)
3. *Identification in the BRM*
4. *Group comparisons in BRM*
   a. Allison's test for comparing coefficients.
   b. Tests to compare predictions between groups.
5. I discuss predictions; see Long & Mustillo for comparing effects across groups

Group comparisons in the LRM

Men: $y = \alpha^m + \beta^m_{articles} articles + \beta^m_{prestige} prestige + \varepsilon$

Women: $y = \alpha^w + \beta^w_{articles} articles + \beta^w_{prestige} prestige + \varepsilon$



LRM: testing equality of regression coefficients

1. Do men and women have the same return for articles?

$$H_0^A: \beta^w_{articles} = \beta^m_{articles}$$

2. Standard t-test

$$t = \frac{\hat{\beta}^w_{articles} - \hat{\beta}^m_{articles}}{\sqrt{Var\left(\hat{\beta}^w_{articles}\right) + Var\left(\hat{\beta}^m_{articles}\right)}}$$

3. More generally, tests that all regression coefficients are equal

$$H_0^B: \alpha^w = \alpha^m;\ \beta^w_{articles} = \beta^m_{articles};\ \beta^w_{prestige} = \beta^m_{prestige}$$

4. This hypothesis does *not* imply the models are the same for men and women!

Here's why...

LRM: differences in explained variation (R^2)

1. If

$$\alpha^w = \alpha^m;\ \beta^w_{articles} = \beta^m_{articles};\ \beta^w_{prestige} = \beta^m_{prestige}$$

2. This does *not* imply

$$R^2_w = R^2_m$$

3. Since I expect less explained variation for women, unexplained variation is critical for group comparison in the BRM

BRM review

Logit

$$\Pr(y=1\,|\,\mathbf{x}) = \frac{\exp(\beta_0 + \beta_x x + \beta_z z)}{1+\exp(\beta_0 + \beta_x x + \beta_z z)}$$

Probit

$$\Pr(y=1\,|\,\mathbf{x}) = \int_{-\infty}^{\beta_0+\beta_x x+\beta_z z} \frac{1}{\sqrt{2\pi}}\left(\frac{-t^2}{2}\right) dt$$

Generally

$$\Pr(y=1\,|\,\mathbf{x}) = F(\beta_0 + \beta_x x + \beta_z z)$$

Gender differences in tenure (-brmgroups-tenure.do)

Variable	Mean	StdDev	Minimum	Maximum	Label
tenure	0.12	0.33	0.00	1.00	Is tenured?
female	0.38	0.48	0.00	1.00	Scientist is female?
year	3.86	2.30	1.00	10.00	Years in rank.
yearsq	20.17	22.15	1.00	100.00	Years in rank squared.
select	5.00	1.41	1.00	7.00	Selectivity of bachelor's
articles	7.05	6.58	0.00	73.00	Total number of articles.
prestige	2.65	0.78	0.65	4.80	Prestige of department.
presthi	0.05	0.21	0.00	1.00	Prestige is 4 or higher?

N = 2797 (person-years)

Models consider (## indicates interaction)

M1: tenure on female + controls

M2: tenure on female##articles

M3: tenure on female##(articles presthi)

M4: tenure on female##(articles other controls)

Testing group differences in the BRM

Tests of coefficients

1. In M1, test of dummy variable

$$H_0: \beta_{female} = 0$$

2. In M2+: Equality of regression coefficients across gender

$$H_0: \beta^{w}_{articles} = \beta^{m}_{articles}$$

Tests of predictions

1. Equality of probabilities for men and women

$$H_0: \Pr(y = 1 \mid \mathbf{x})_w = \Pr(y = 1 \mid \mathbf{x})_m$$

2. Equivalently, the discrete change is 0:

$$H_0: \Delta_{m-w}(\mathbf{x}) = \Pr(y = 1 \mid \mathbf{x})_m - \Pr(y = 1 \mid \mathbf{x})_w = 0$$

#2 M1: dummy variable for gender

$$\Pr(\text{tenure} = 1 \mid \mathbf{x}) = \Lambda\left(\begin{array}{l}\beta_0 + \beta_{female}\text{female} + \beta_{year}\text{year} + \beta_{yearsq}\text{yearsq} \\ + \beta_{select}\text{select} + \beta_{articles}\text{articles} + \beta_{presthi}\text{presthi}\end{array}\right)$$

Odds of: Tenure vs NoTenure

tenure	b	z	P>|z|	e^b	e^bStdX	SDofX
female	-0.35260	-2.677	0.007	0.7029	0.8429	0.4849
year	1.69865	10.426	0.000	5.4666	49.9816	2.3028
c.year#c.y.	-0.12295	-8.748	0.000	0.8843	0.0656	22.1512
select	0.12228	2.699	0.007	1.1301	1.1878	1.4075
articles	0.04948	5.986	0.000	1.0507	1.3845	6.5757
presthi	-1.05052	-2.662	0.008	0.3498	0.8009	0.2113

```
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
     e^b = exp(b) = factor change in odds for unit increase in X
 e^bStdX = exp(b*SD of X) = change in odds for SD increase in X
```


Predicted probabilities at given x's

1. Compute predictions at specific values of the variables

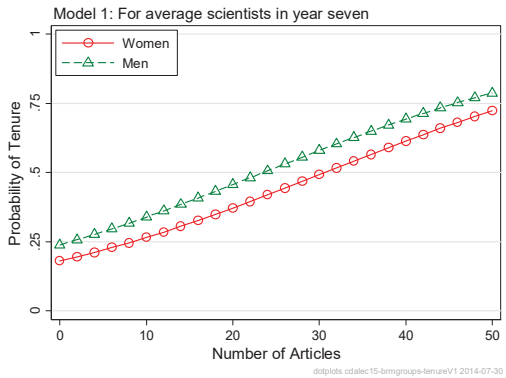
$$\Pr(\text{tenure} = 1 \mid \mathbf{x}) = \Lambda\left(\begin{array}{l}\beta_0 + \beta_{female}\text{female} + \beta_{year}\text{year} + \beta_{yearsq}\text{yearsq} \\ + \beta_{select}\text{select} + \beta_{articles}\text{articles} + \beta_{presthi}\text{presthi}\end{array}\right)$$

2. Probability for women in year 7 with selectivity 4, low prestige, and no articles

$$0.16 = \Lambda\left(\begin{array}{l}\beta_0 + \beta_{female}(1) + \beta_{year}(7) + \beta_{yearsq}(49) \\ + \beta_{select}(4) + \beta_{articles}(0) + \beta_{presthi}(.05)\end{array}\right)$$

3. Extending this idea, plots of probabilities are constructed
   - Probability curves for x_k for men and women have the same slope but intercepts differ by β_{female}

#3 M1: Probabilities for men & women by # of articles



1. Pr(Tenure) is about .06 greater for men than women at all levels of productivity
2. The lack of gender interactions is *unrealistic*

BRM with β's differing by group

1. Allowing different coefficients for men and women:

$$\text{Women:}\quad \Pr(y=1)=\Lambda\left(\alpha^{w}+\beta^{w}_{articles}articles+\beta^{w}_{prestige}prestige\right)$$

$$\text{Men:}\quad \Pr(y=1)=\Lambda\left(\alpha^{m}+\beta^{m}_{articles}articles+\beta^{m}_{prestige}prestige\right)$$

2. Can we use a Chow-type test?

$$H_0\colon \beta^{w}_{articles}=\beta^{m}_{articles}$$

3. Allison (1999) writes:

   Because of an *identification problem*, the usual tests of this hypothesis tell us *nothing* about the underlying impact of articles for men and women.

4. Is he right? Why?

Identification in the BRM

1. Start with a regression with a latent y*

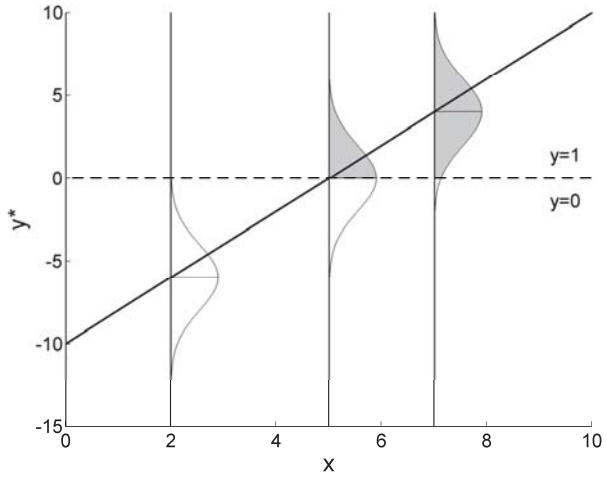
$$y^{*}=\alpha+\beta x+\varepsilon$$

2. For identification, the mean and variance of ε are assumed
   - ε is normal(0,1) for probit or logistic(0, $\pi^2/3$) for logit.
3. y and y* are linked by

$$y=\begin{cases}1 & \text{if } y^{*}>0\\ 0 & \text{if } y^{*}\le 0\end{cases}$$

4. Pr(y=1|**x**) is the shaded region in the following graph

*Linear model for y**



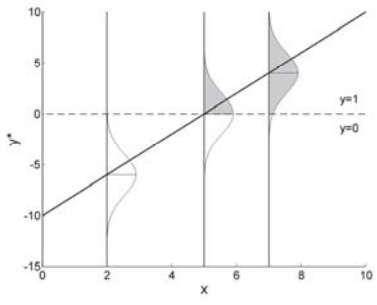
*Computing Pr(y) from y**

1. The probability depends on
   a. The error distribution
   b. The regression coefficients
   c. The value of x
2. To compute the probability:

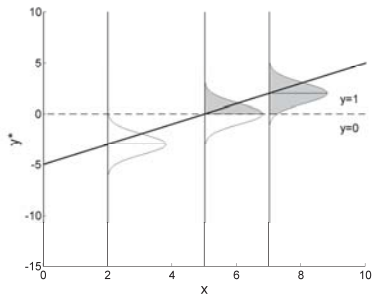
$$\Pr(y=1\mid x)=\Pr(y^{*}>0\mid x)$$
$$=\Pr(\varepsilon<[\alpha+\beta x]\mid x)$$

3. The identification problem is illustrated in this graph

Group W: α=-12, β=2, σ=2



Group M: α=-6, β=1, σ=1



1. The change in y* is twice as large for Women than Men
2. The change in probability is identical for Women and Men
3. Empirically, the effect of x is indistinguishable for the two groups

Identification and group comparisons in the BRM

1. Regress y* on articles

Women: $y^* = \alpha^w + \beta^w_{\text{articles}}\text{articles} + \varepsilon_m$

Men: $y^* = \alpha^m + \beta^m_{\text{articles}}\text{articles} + \varepsilon_w$

2. I want to test

$\beta^w_{\text{articles}} = \beta^m_{\text{articles}}$

3. Substantively, I expect

$\sigma^2_w > \sigma^2_m$

4. For identification, software *assumes*

Logit: $\text{Var}(\varepsilon) = \pi^2/3$ *Probit*: $\text{Var}(\varepsilon) = 1$

6. For probit, software rescales the "true" ε to have variance 1

$$Var\left(\frac{\varepsilon}{\sigma}\right) = Var(\tilde{\varepsilon}) = 1$$

7. The *estimated* model for women is

$$\frac{y^*}{\sigma_w} = \frac{\alpha^w}{\sigma_w} + \frac{\beta^w_{\text{articles}}}{\sigma_w}\text{articles} + \frac{\varepsilon_w}{\sigma_w}$$
$$= \tilde{\alpha}^w + \tilde{\beta}^w_{\text{articles}}\text{articles} + \tilde{\varepsilon}_w, \text{ where } \tilde{\sigma}_w \equiv 1$$

8. For men

$$\frac{y^*}{\sigma_m} = \frac{\alpha^m}{\sigma_m\sigma_m} + \frac{\beta^m_{\text{articles}}}{\sigma_m}\text{articles} + \frac{\varepsilon_m}{\sigma_m}$$
$$= \tilde{\alpha}^m + \tilde{\beta}^m_{\text{articles}}\text{articles} + \tilde{\varepsilon}_m, \text{ where } \tilde{\sigma}_m \equiv 1$$

9. **Substantively**, we want to test

$H_0^{NoTilda}: \beta^w_{\text{articles}} = \beta^m_{\text{articles}}$

10. Standard software tests

$H_0^{Tilda}: \tilde{\beta}^w_{\text{articles}} = \tilde{\beta}^m_{\text{articles}}$

11. The problem is

a. Equal tilde coefficients $\tilde{\beta}^w_{\text{articles}} = \tilde{\beta}^m_{\text{articles}}$

b. Does not imply equal non-tilde $\beta^w_{\text{articles}} = \beta^m_{\text{articles}}$

c. Unless $\sigma^2_m = \sigma^2_w$

and we think the variances vary by gender

12. This is the problem raised by Allison (1999)

* Aside: rescaling errors in logit

1. The model

$$y^* = \alpha + \beta_{\text{articles}}\text{articles} + \varepsilon$$

2. Rescale the errors

$$Var(\tilde{\varepsilon}) = \frac{\pi^2}{3} \text{ rather than 1 for probit}$$

3. The estimated equation

$$\frac{\pi}{\sqrt{3}}\frac{y^*}{\sigma} = \frac{\pi}{\sqrt{3}}\frac{\alpha}{\sigma} + \frac{\pi}{\sqrt{3}}\frac{\beta_{\text{articles}}}{\sigma}\text{articles} + \frac{\pi}{\sqrt{3}}\frac{\varepsilon}{\sigma}$$

Allison's test of the equality of non-tilde coefficients

1. *Assume* for some variable *z*

$$\beta_z^w = \beta_z^m \text{ or equivalently } \frac{\beta_z^w}{\beta_z^m} = 1$$

2. The ratio $\tilde{\beta}_z^w / \tilde{\beta}_z^m$ is the relative size of σ_m and σ_w

$$\frac{\tilde{\beta}_z^w}{\tilde{\beta}_z^m} = \frac{\beta_z^w / \sigma_w}{\beta_z^m / \sigma_m} = \frac{\sigma_m}{\sigma_w}$$

3. This provides leverage to test the underlying coefficients

$$H_0: \beta_x^w = \beta_x^m$$

4. This works *only if* you can justify the assumption $\beta_z^w = \beta_z^m$

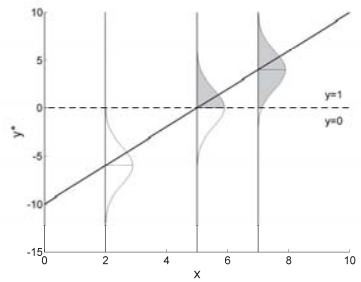
5. To avoid this assumption, I use tests of probabilities

Testing probabilities

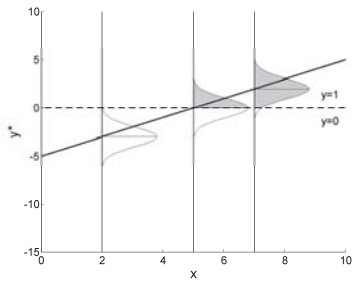
1. I am interested in the probabilities, not the slopes

2. Probabilities are invariant to Var(ε) which allows us to test

$$H_0: \Pr(y=1 \mid \mathbf{x})_w = \Pr(y=1 \mid \mathbf{x})_m$$



Women: α=-12, β=2, σ=2



Men: α=-6, β=1, σ=1

#4 M2: articles and gender

```
logit (N=2797): Factor Change in Odds

  Odds of: Tenure vs NoTenure

WOMEN              b         z     P>|z|     e^b    e^bStdX
-------------------------------------------------------------
    constant   -2.50116  -17.858   0.000   0.0820   0.2974
    articles    0.04714    4.490   0.000   1.0483   1.3150

MEN                b         z     P>|z|     e^b    e^bStdX
-------------------------------------------------------------
    constant   -2.72101  -22.402   0.000   0.0658   0.2673
    articles    0.10239    9.756   0.000   1.1078   1.8054

------------------------------------------------------------------
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
     e^b = exp(b) = factor change in odds for unit increase in X
 e^bStdX = exp(b*SD of X) = change in odds for SD increase in X
```


Comparing groups using probabilities

1. Compute the discrete change for gender (*group difference*):

$$\Delta_{m-w}(\text{articles}) = \Pr(y=1\,|\,\text{articles})_m - \Pr(y=1\,|\,\text{articles})_w$$

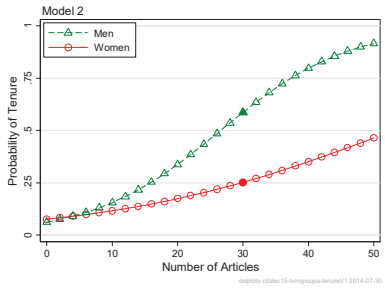
2. With a 95% confidence interval (CI):

$$\left[\Delta_{m-w}(\text{articles})_{\text{LowerBound}}, \Delta_{m-w}(\text{articles})_{\text{UpperBound}}\right]$$

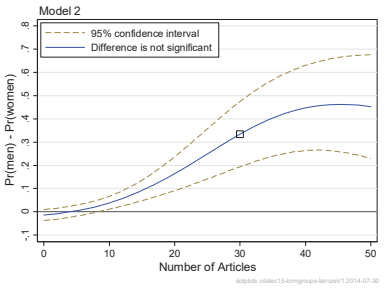
3. With repeated sampling, we expect $\Delta_{m\text{-}w}$ to fall within interval 95% of the time.
4. The CI can be computed using
   a. Delta method using **`margins`** which is fast
   b. Bootstrap which is 1,000 time slower
5. With one RHS variable, we can plot all probabilities and discrete changes....

#6 Probabilities and group difference

A: Probabilities by group

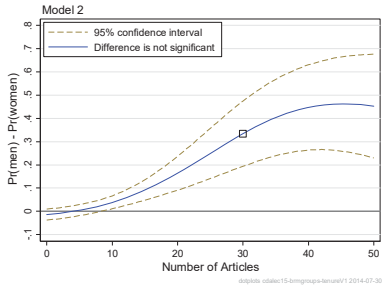


B: Discrete change with CI

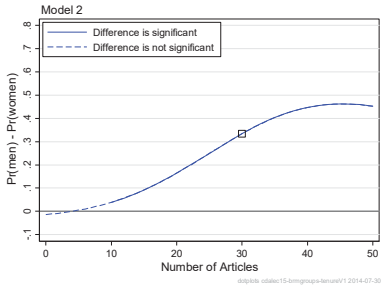


Group difference with CI or a broken line

B: Group difference with CI



C: Group difference with broken line



Adding variables

1. Adding variables complicates things
2. With two independent variables

$$\Pr(y=1\mid x,z)=F(\alpha+\beta_x x+\beta_z z)$$

3. Setting $z=Z^*$ changes the intercept in an equation including only *x*

$$\begin{aligned}\Pr(y=1\mid x,Z^*)&=F(\alpha+\beta_x x+\beta_z Z^*)\\&=F\left(\left[\alpha+\beta_z Z^*\right]+\beta_x x\right)\\&=F(\alpha^*+\beta_x x)\end{aligned}$$

4. The probabilities and group differences depend on the levels of **all** variables

Comparing groups with additional variables

1. For a given z=Z*

Men: $\Pr(y=1\mid x,Z^*)_m=F(\alpha^{*m}+\beta_x^m x)$

Women: $\Pr(y=1\mid x,Z^*)_w=F(\alpha^{*w}+\beta_x^w x)$

2. Group difference depends on the level of all variables

$$\Delta_{m-w}(x,Z^*)=\Pr(y=1\mid x,Z^*)_m-\Pr(y=1\mid x,Z^*)_w$$

#7 M3: articles and prestigious jobs

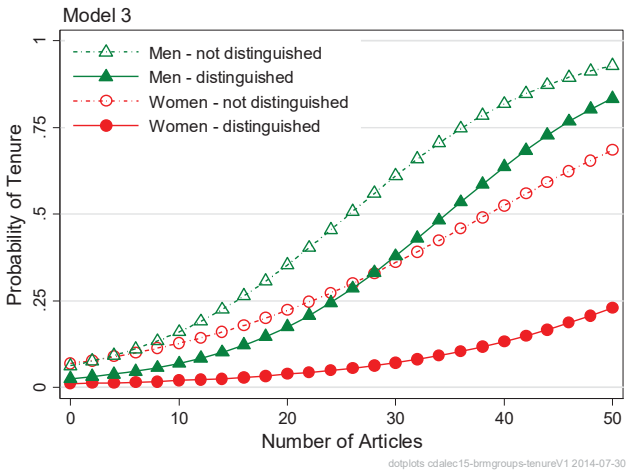
logit (N=2797): Factor Change in Odds

Odds of: Tenure vs NoTenure

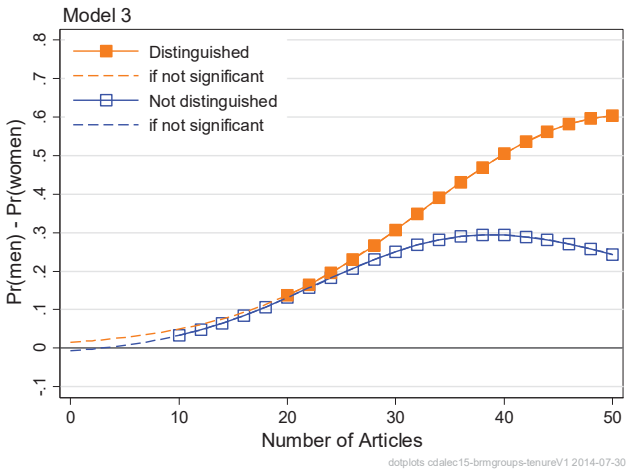
WOMEN	b	z	P>|z|	e^b	e^bStdX
constant	-2.60432	-17.320	0.000	0.0740	0.2829
articles	0.06761	5.358	0.000	1.0699	1.4811
presthi	-1.98396	-2.685	0.007	0.1375	0.7484

MEN	b	z	P>|z|	e^b	e^bStdX
constant	-2.71499	-22.268	0.000	0.0662	0.2681
articles	0.10554	9.890	0.000	1.1113	1.8385
presthi	-0.94529	-2.058	0.040	0.3886	0.8627

M3: Plot of probabilities



M3: Group differences



#10 M4: full model for women

logit (N= 2797): Factor Change in Odds

Odds of: Tenure vs NoTenure

WOMEN	b	z	P>|z|	e^b	e^bStdX
constant	-5.84198	-6.747	0.000	0.0029	0.0589
year	1.40777	5.472	0.000	4.0868	30.1273
yearsq	-0.09559	-4.364	0.000	0.9088	0.1857
select	0.05513	0.769	0.442	1.0567	1.1534
articles	0.03395	2.693	0.007	1.0345	1.2181
prestige	-0.37079	-2.376	0.017	0.6902	0.6013

b = raw coefficient
z = z-score for test of b=0
P>|z| = p-value for z-test
e^b = exp(b) = factor change in odds for unit increase in X
e^bStdX = exp(b*SD of X) = change in odds for SD increase in X

#10 M4: full model for men

logit (N= 2797): Factor Change in Odds

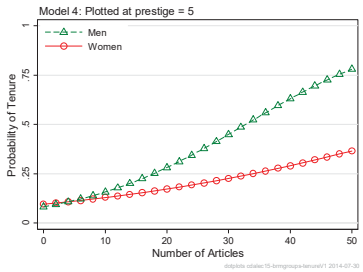
Odds of: Tenure vs NoTenure

MEN	b	z	P>|z|	e^b	e^bStdX
constant	-7.68016	-11.271	0.000	0.0005	0.0241
year	1.90885	8.915	0.000	6.7454	130.9789
yearsq	-0.14322	-7.699	0.000	0.8666	0.0622
select	0.21577	3.513	0.000	1.2408	1.7711
articles	0.07369	6.367	0.000	1.0765	1.5299
prestige	-0.43119	-3.963	0.000	0.6497	0.5418

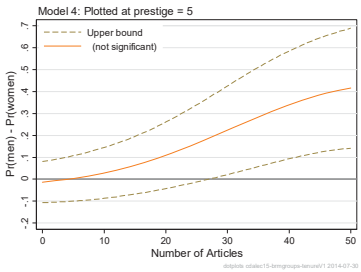
b = raw coefficient
z = z-score for test of b=0
P>|z| = p-value for z-test
e^b = exp(b) = factor change in odds for unit increase in X
e^bStdX = exp(b*SD of X) = change in odds for SD increase in X

M4: Plots with prestige = 5

A. Probabilities by gender



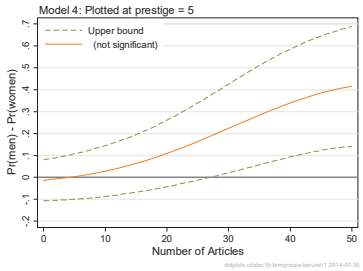
B. Group differences with CI



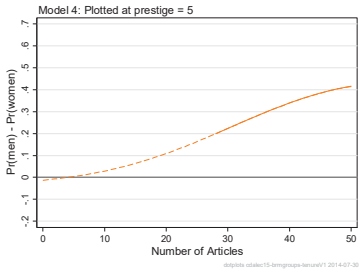
M4: Plots with prestige = 5

Converting CI to a broken line

B. Group difference with CI

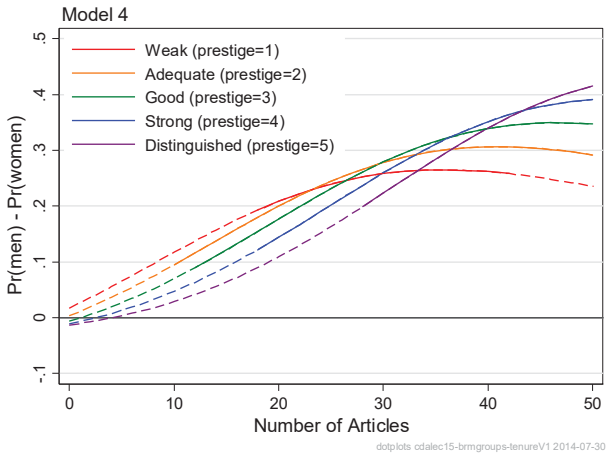


C. Group difference with broken line

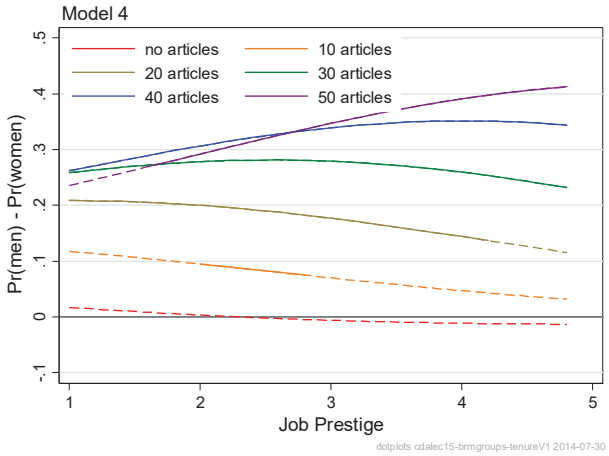


1. Do this for each level of prestige
2. Then combine the group differences in a single graph

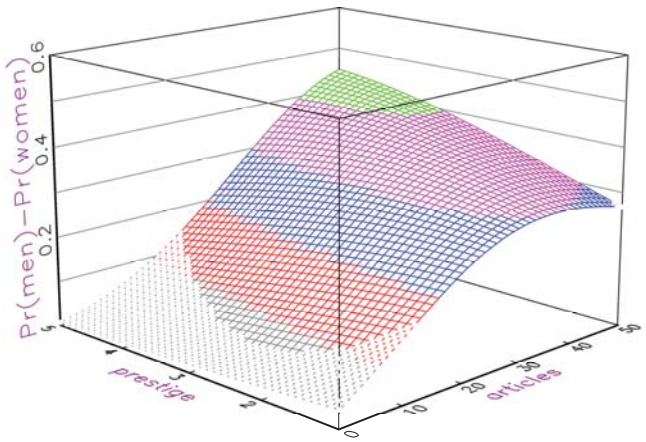
M4: Group differences by articles & prestige



M4: Group differences by prestige & articles



M4: Group differences by prestige and articles



Conclusions

1. **LRM**: Chow-type tests are used to compare coefficients across groups
2. **BRM**: Chow-type tests should not be used due to identification
3. Two approaches for comparing groups
   a. **Slope coefficients**: With added assumptions tests are possible
   b. **Probabilities**: Tests are not affected by the identification issues
4. Both approaches have limitations

Tests of regression coefficients

1. Can the equality assumption be justified?
2. Technical issues regarding tests (see Williams 2009).
3. Are the coefficients what you **want** to test?
4. Does it matter whether gender differences in tenure are due to differences in the effect of articles or differences in unexplained variation?

Tests of predictions

1. The substantive question is
   "Is the effect of articles the same for men and women?"
2. With probabilities, this does not have a simple answer
3. Do you need to adjust for multiple tests? If so, how?
4. Do you have sufficient observations to support your conclusions?
   Are the predictions *on the support*?

Was Allison right?

"Differences in the estimated coefficients *tell us nothing* about the differences in the underlying impact of [publications] on [tenure for] the two groups."

1. Predicted probabilities are not affected by the identification problem
2. Odds ratios can be computed as an average of predicted probabilities
3. Accordingly, you can test if the OR's are equal across groups. But would you want to?

Fundamentally, what is an effect?

1. Can probabilities show the effect of articles?
2. Are regression coefficients required to describe an effect?

Bibliography

Allison, Paul D. 1999. Comparing logit and probit coefficients across groups. *Sociological Methods and Research* 28:186-208.

Chow, G.C. 1960. Tests of equality between sets of coefficients in two linear regressions. *Econometrica* 28:591-605.

Long, J.S. 2009 (2005). Comparing groups using predicted outcomes. (www.indiana.edu/~jslsoc/research_groupdif.htm)

Long, J.S. and S. Mustillo. 2017. Working paper.

Williams, Richard. 2009. Using heterogeneous choice models to compare logit and probit coefficients across groups. *Sociological Methods & Research* 37: 531-559