ICPSR Summer Program

Comparing groups in regression models for binary outcomes

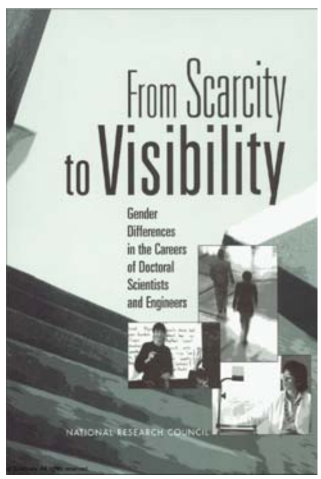
Scott Long

Indiana University

Blalock Lecture

August 2010

An motivating example



Statistical and substantive problems

Are the "effects" of <u>productivity on tenure</u> the same for <u>men and women</u>?

1. A common solution for comparing groups:
   a. Estimate model for women.
   b. Estimate same model for men.
   c. Compare coefficients across groups.
   d. Conclude either they are the same or they are not.
2. Two problems:
   a. NRC panel did not find the coefficients informative.
   b. Paul Allison sent me a working paper that said:

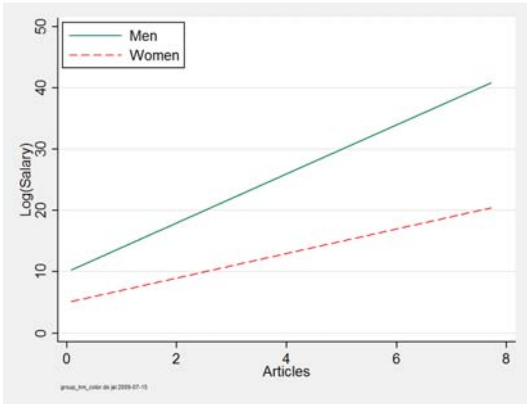
      *Differences in the estimated coefficients tell us **nothing** about the differences in the underlying impact of [publications] on [tenure for] the two groups.*

Objectives for today's talk

1. Review methods for comparing groups in the **LRM**.
2. Review binary logit and probit (**BRM**) in terms of:
   a. A latent variable model for y*.
   b. A probability model for Pr(y=1|**x**).
3. Why is there a problem comparing coefficients across groups in BRM?
4. Two approaches to comparing groups in BRM:
   a. Briefly: Allison's tests for comparing coefficients.
   b. In depth: Using predicted probabilities to compare groups.

Group comparisons in the LRM

$$\text{Men:} \quad y = \alpha^m + \beta^m_{articles} articles + \beta^m_{prestige} prestige + \varepsilon$$
$$\text{Women:} \quad y = \alpha^w + \beta^w_{articles} articles + \beta^w_{prestige} prestige + \varepsilon$$



Testing coefficients in the LRM

1. Do men and women have the same return for articles?

$$H_0^A: \beta^w_{articles} = \beta^m_{articles}$$

2. We compute a 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. We could test that all coefficients are equal:

$$H_0^B: \alpha^w = \alpha^m;\ \beta^w_{articles} = \beta^m_{articles};\ \beta^w_{prestige} = \beta^m_{prestige}$$

4. Critically, H_0^B this does **not** imply:

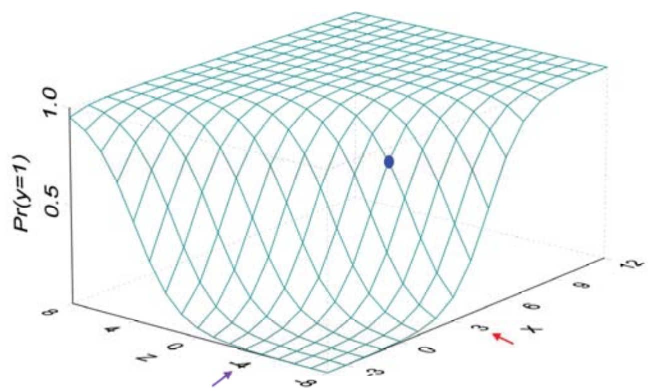
$$H_0^C: R^2_w = R^2_m$$

LRM vs BRM for comparing groups

1. Testing $H_0 : \beta^{w}_{articles} = \beta^{m}_{articles}$ is appropriate in LRM.
2. In the BRM the test confounds:
   a. Group differences in **the effect of x**.
   b. Group differences in **unexplained variation**.
3. Two approaches can be used for the BRM:
   a. Allison's test that assumes effects of other variables are equal across groups.
   b. Tests comparing predicted probabilities.

BRM & the challenge of nonlinearity

$$\Pr\left(y=1\,|\,\mathbf{x}\right)=F\left(\beta_0+\beta_x x+\beta_z z\right)$$



Nonlinear models for binary outcomes

Logit

$$\Pr\left(y=1\,|\,\mathbf{x}\right)=\Lambda\left(\beta_0+\beta_x x+\beta_z z\right)$$
$$=\frac{\exp\left(\beta_0+\beta_x x+\beta_z z\right)}{1+\exp\left(\beta_0+\beta_x x+\beta_z z\right)}$$

Probit

$$\Pr\left(y=1\,|\,\mathbf{x}\right)=\Phi\left(\beta_0+\beta_x x+\beta_z z\right)$$
$$=\int_{-\infty}^{\beta_0+\beta_x x+\beta_z z}\frac{1}{\sqrt{2\pi}}\left(\frac{-t^2}{2}\right)dt$$

Models that allow groups differences

1. **Dummy only**: Use only a dummy variable for group.
2. **Interactions**: Allow group differences in the effects of the predictors such as $\beta^{w}_{articles}$ and $\beta^{m}_{articles}$.
   a. Test the equality of coefficients.
   $$\beta^{w}_{articles} = \beta^{m}_{articles}$$
   b. Compare predictions across groups.
   $$\Pr\left(y=1 \mid \mathbf{x}\right)_{w} = \Pr\left(y=1 \mid \mathbf{x}\right)_{m}$$

Example: gender differences in tenure

Descriptive statistics for data as career years

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

M1: logit model with dummy variable for gender

$$\Pr\left(\text{tenure}=1 \mid \mathbf{x}\right) = \Lambda\left(\begin{array}{l}\beta_0 + \beta_{female}\text{female} + \beta_{year}\text{year} + \beta_{yearsq}\text{yearsq} \\ +\beta_{select}\text{select} + \beta_{articles}\text{articles} + \beta_{presthi}\text{presthi}\end{array}\right)$$

logit (N=2797): Factor Change in Odds

Odds of: Tenure vs NoTenure

tenure	b	z	P>|z|	e^b	e^bStdX
female	-0.35260	-2.677	0.007	0.7029	0.8429
year	1.69865	10.426	0.000	5.4666	49.9816
yearsq	-0.12295	-8.748	0.000	0.8843	0.0656
select	0.12228	2.699	0.007	1.1301	1.1878
articles	0.04948	5.986	0.000	1.0507	1.3845
presthi	-1.05052	-2.662	0.008	0.3498	0.8009
_cons	-7.59000	-15.025	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 ratios for interpretation

1. Odds:

$$\text{Odds}(x,z)=\frac{\Pr(y=1\mid x,z)}{\Pr(y=0\mid x,z)}$$

2. Odds ratios:

$$\frac{\text{Odds}(x+1,z)}{\text{Odds}(x,z)}=\exp(\beta_x)$$

3. Odds ratio for female: $\exp(\beta_{\text{female}})=0.70.$

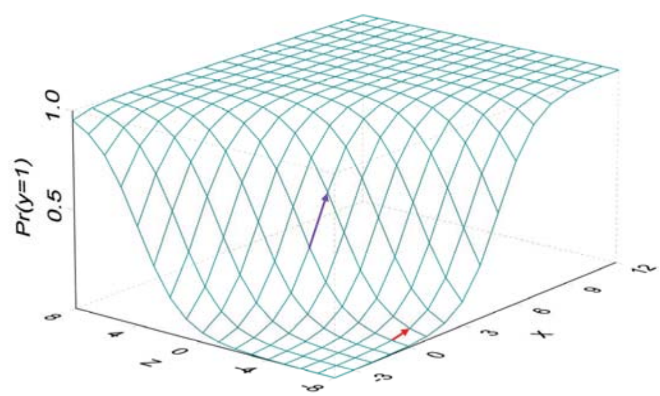
   Being a female scientist decreases the odds of tenure by a factor of .70, holding all other variables constant.

2. Odds ratio for articles: $\exp(\beta_{\text{articles}})=1.05.$

   For each additional article, the odds of tenure increase by a factor of 1.05, holding all other variables constant.

Odds ratios compared to changes in probabilities

Both arrows correspond to the **same odds ratio**.



Predicted probabilities at a given x

1. Compute the predicted probability at specific values of the independent variables:

$$\Pr(\text{tenure}=1\mid \mathbf{x})=\Lambda\begin{pmatrix}\beta_0+\beta_{female}\text{female}+\beta_{year}\text{year}+\beta_{yearsq}\text{yearsq}\\ +\beta_{select}\text{select}+\beta_{articles}\text{articles}+\beta_{presthi}\text{presthi}\end{pmatrix}$$

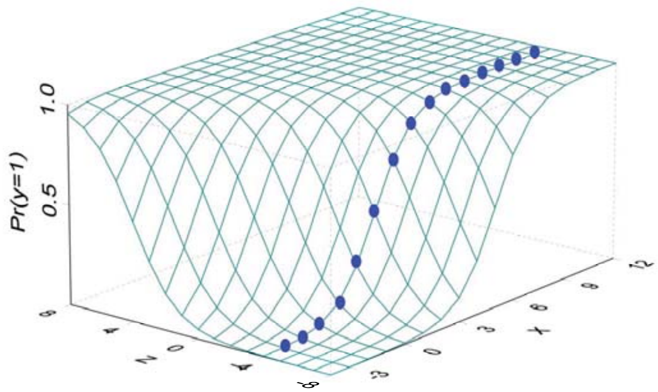
2. For example, the probability of tenure for non-publishing women in year 7, with selectivity 4 and low prestige:

$$0.16=\Lambda\begin{pmatrix}\beta_0+\beta_{female}(1)+\beta_{year}(7)+\beta_{yearsq}(49)\\ +\beta_{select}(4)+\beta_{articles}(0)+\beta_{presthi}(.05)\end{pmatrix}$$

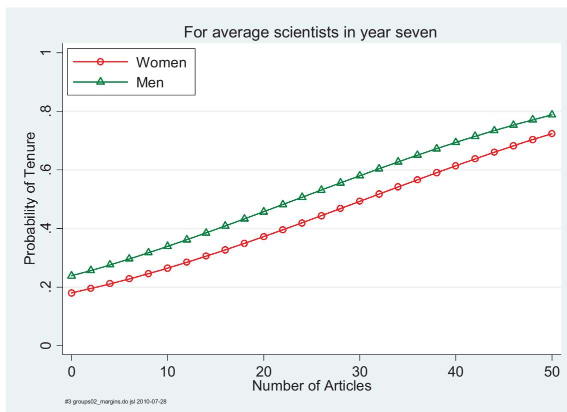
3. Extending this idea, plots of probabilities can be constructed...

For a simple model with two predictors

$$\Pr(y=1\mid \mathbf{x}) = F(\beta_0 + \beta_x x + \beta_z z)$$



M1: Probabilities for men & women by # of articles



1. Probability of tenure is about .06 higher for men at all levels of productivity.
2. Lack of interactions with gender is unrealistic.

Group comparisons in BRM when β's differ

1. In the BRM:

   Women: $\Pr(y=1) = \Lambda(\alpha^w + \beta^w_{articles} articles + \beta^w_{prestige} prestige)$

   Men: $\Pr(y=1) = \Lambda(\alpha^m + \beta^m_{articles} articles + \beta^m_{prestige} prestige)$

2. Can we use a Chow-type test?

   $H_0\colon \beta^w_{articles} = \beta^m_{articles}$

3. Allison (1999) show that because of an **identification problem**, the usual tests of this hypothesis tells us **nothing** about the underlying impact of articles for men and women.

Regression on a latent y* for the BRM

The equations

1. **Structural model** with a latent y*:

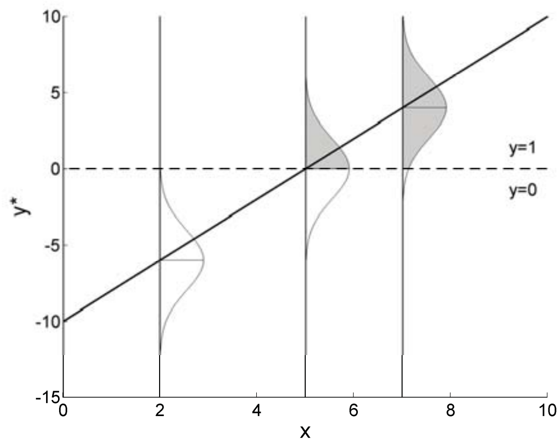
$$y^* = \alpha + \beta x + \varepsilon$$

2. **Error** ε is normal(0,1) for probit; ε is logistic(0, $\pi^2 / 3$) for logit.
3. **Observed y and latent y*** are linked by:

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

4. **Graphically...**

Graphing the latent y*



Computing Pr(y) from y*

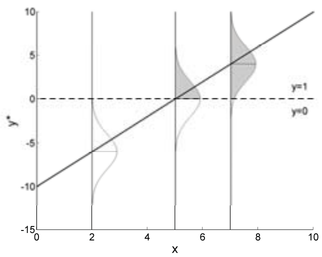
5. The probability depends on the <u>error distribution</u> *and* the <u>coefficients</u>:

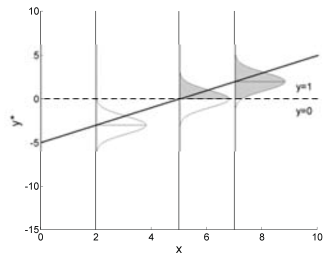
$$\Pr(y=1 \mid x) = \Pr\left(y^* > 0 \mid x\right)$$
$$= \Pr\left(\varepsilon < [\alpha + \beta x] \mid x\right)$$

6. There is an **identification problem** that can be illustrated graphically.

Identification problem in BRM

Group W: α=-12, β=2, σ=2 **Group M: α=-6, β=1, σ=1**





7. The "effect" (i.e., β) of x is **twice as large** for W than M.
8. In terms of $\Pr(y=1)$, these M and W are **empirically indistinguishable**:

   For W: A change of 1 in x when β=2 and σ=2.

   For M: A change of 1 in x when β=1 and σ=1.

Identification and group comparisons

1. Let y* be the latent variable associated with receipt of tenure:

   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. Assume the coefficients for articles are equal:

   $\beta^w_{\text{articles}} = \beta^m_{\text{articles}}$

3. Assume women have **more unexplained variation** (why would they?):

   $\sigma^2_w > \sigma^2_m$

4. When you estimate the model, the **software** makes implicit assumptions:

   Logit: $Var(\varepsilon) = \pi^2/3$ Probit: $Var(\varepsilon) = 1$

5. What is the effect of these implicit assumptions?

6. With **probit**, ε is rescaled so that:

   $$Var\left(\frac{\varepsilon}{\sigma}\right) = Var(\tilde{\varepsilon}) = 1$$

7. For women, the **estimated** model for probit is:

   $$\frac{y^*}{\sigma_w} = \frac{\alpha^w}{\sigma_w} + \frac{\beta^w_{\text{articles}}}{\sigma_w}\text{articles} + \frac{\varepsilon_w}{\sigma_w}$$
   $$= \tilde{\alpha}^w + \tilde{\beta}^w_{\text{articles}}\text{articles} + \tilde{\varepsilon}_w, \text{ where } \tilde{\sigma}_w = 1$$

8. For men, the **estimated** model for probit is:

   $$\frac{y^*}{\sigma_m} = \frac{\alpha^m}{\sigma_m} + \frac{\beta^m_{\text{articles}}}{\sigma_m}\text{articles} + \frac{\varepsilon_m}{\sigma_m}$$
   $$= \tilde{\alpha}^m + \tilde{\beta}^m_{\text{articles}}\text{articles} + \tilde{\varepsilon}_m, \text{ where } \tilde{\sigma}_m = 1$$

9. We want to test:

$$H_0^{NoTilda}: \beta_{\text{articles}}^{w} = \beta_{\text{articles}}^{m}$$

10. But, standard software only allows us to test:

$$H_0^{Tilda}: \tilde{\beta}_{\text{articles}}^{w} = \tilde{\beta}_{\text{articles}}^{m}$$

11. Unless $\sigma_m^2 = \sigma_w^2$,

$$H_0^{NoTilda} \textit{ is not equivalent to } H_0^{Tilda}$$

Aside: rescaling errors in logit is messier

1. The model is:

$$y^* = \alpha + \beta_{\text{articles}}\text{articles} + \varepsilon$$

2. We rescale the errors so that:

$$Var(\tilde{\varepsilon}) = \frac{\pi^2}{3} \text{ rather than 1 for probit}$$

3. This leads to the equation that is estimated:

$$\frac{\pi}{\sqrt{3}}\frac{y^*}{\sigma} = \frac{\pi}{\sqrt{3}}\frac{\alpha}{\sigma} + \frac{\pi}{\sqrt{3}}\frac{\beta_{\text{articles}}}{\sigma}\text{articles} + \frac{\pi}{\sqrt{3}}\frac{\varepsilon}{\sigma}$$

Alternatives for testing group differences in BRM

Two distinct approaches address the identification problem.

1. Allison's test of $H_0: \beta_x^w = \beta_x^m$ disentangles the β's and $Var(\varepsilon)$.
   a. The test requires the **strong assumption**:

   $$\beta_z^w = \beta_z^m \text{ or equivalently } \frac{\beta_z^w}{\beta_z^m} = 1$$

   b. The ratio of $\tilde{\beta}_z$'s estmates group differences in unexplained variation:

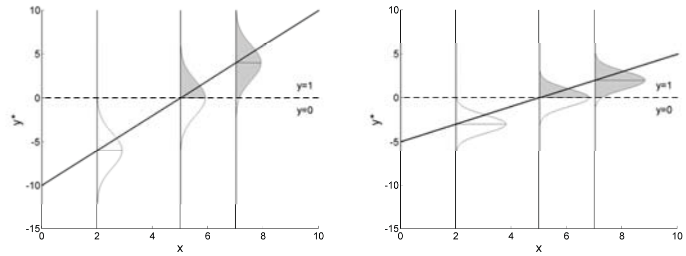
   $$\frac{\tilde{\beta}_z^w}{\tilde{\beta}_z^m} = \frac{\beta_z^w / \sigma_w}{\beta_z^m / \sigma_m} = \frac{\beta_z^w \sigma_w}{\beta_z^m \sigma_m} = 1 \times \frac{\sigma_m}{\sigma_w}$$

   c. This provides leverage to test:

   $$H_0: \beta_x^w = \beta_x^m$$

2. Alternatively, since probabilities are invariant to $Var(\varepsilon)$, I propose testing:

$$H_0\!: \Pr\left(y=1\,|\,\mathbf{x}\right)_w = \Pr\left(y=1\,|\,\mathbf{x}\right)_m$$



Women: α=-12, β=2, σ=2 Men: α=-6, β=1, σ=1

Setting up a model for group differences in β's

1. Let $w=1$ for women, else 0 and $wx = w \times x$;

   let $m=1$ for men, else 0 and $mx = m \times x$.

$$\Pr\left(y=1\,|\,\mathbf{x}\right) = F\left(\alpha^w w + \beta_x^w wx + \alpha^m m + \beta_x^m mx\right)$$

2. Then:

$$\Pr\left(y=1\,|\,\mathbf{x}\right)_w = F\left(\alpha^w + \beta_x^w x\right) \quad \text{if } w=1, m=0$$

$$\Pr\left(y=1\,|\,\mathbf{x}\right)_m = F\left(\alpha^m + \beta_x^m x\right) \quad \text{if } w=0, m=1$$

3. The gender difference in the probability of tenure is:

$$\Delta_{m-w}\left(\mathbf{x}\right) = \Pr\left(y=1\,|\,\mathbf{x}\right)_m - \Pr\left(y=1\,|\,\mathbf{x}\right)_w$$

M2: articles and gender

Start with a simple model with only publications predicting tenure:

```
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 with predicted probabilities and CIs

To compare groups at different levels of articles:

1. Compute **discrete change** (aka: first difference):

$$\Delta_{m-w}(\text{articles}) = \Pr(y=1\,|\,\text{articles})_m - \Pr(y=1\,|\,\text{articles})_w$$

2. **Confidence intervals** for predictions:

$$\left[\Delta_{m-w}(\text{articles})_{\text{LowerBound}}, \Delta_{m-w}(\text{articles})_{\text{UpperBound}}\right]$$

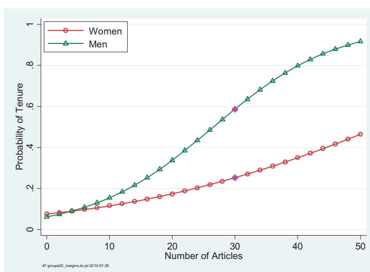
With repeated sampling, we would expect the predicted probability to fall within the CI ninety-five percent of the time.

3. For CI, **delta** or **end-point method** are easy and fast;
bootstrap requires at least 1,000 replications to get reliable results.

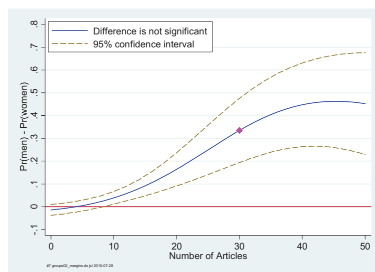
4. With one RHS variable, we can plot all comparisons.

5. Moving from predictions for each group to differences in predictions:

A: Probablities by group

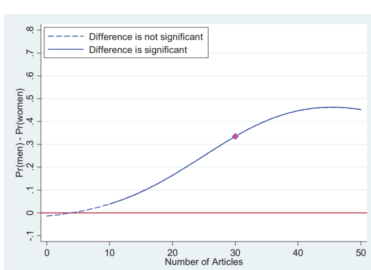


B: Discrete change with CI

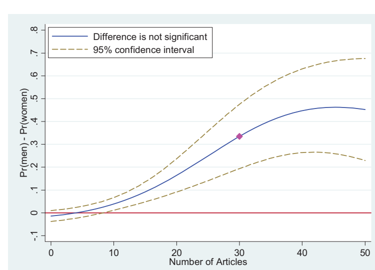


6. Or using dashed lines to indicate differences that are not significant.

C: Discrete change with broken line



B: Discrete change with CI



Adding additional variables

1. Adding variables introduces substantial complications for interpretation:
2. With two independent variables:

$$\Pr(y=1\,|\,x,z)=F(\alpha+\beta_x x+\beta_z z)$$

3. Setting $z=Z^*$ changes the intercept in an equation with only x:

$$\Pr(y=1\,|\,x,Z^*)=F(\alpha+\beta_x x+\beta_z Z^*)$$
$$=F([\alpha+\beta_z Z^*]+\beta_x x)$$
$$=F(\alpha^*+\beta_x x)$$

4. Thus, probabilities and discrete changes depend on the levels of each variable.

Comparing groups with additional variables

Control variables change the intercept

1. For a given $z=Z$:

Men: $\Pr(y=1\,|\,x,Z^*)_m=F(\alpha^{*m}+\beta_x^m x)$

Women: $\Pr(y=1\,|\,x,Z^*)_w=F(\alpha^{*w}+\beta_x^w x)$

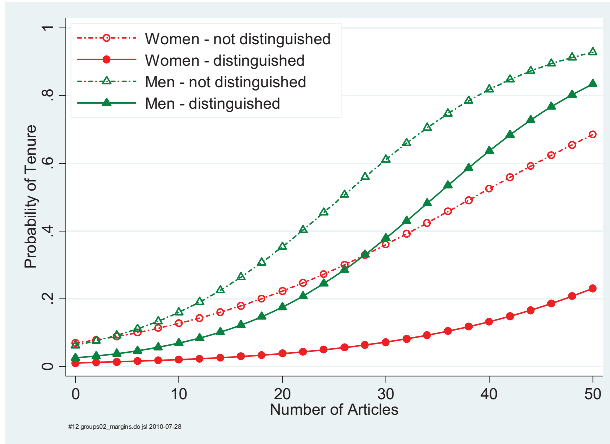
2. Differences in probabilities for a given x depends on the level of other variables:

$$\Delta_{m-w}(x,Z^*)=\Pr(y=1\,|\,x,Z^*)_m-\Pr(y=1\,|\,x,Z^*)_w$$

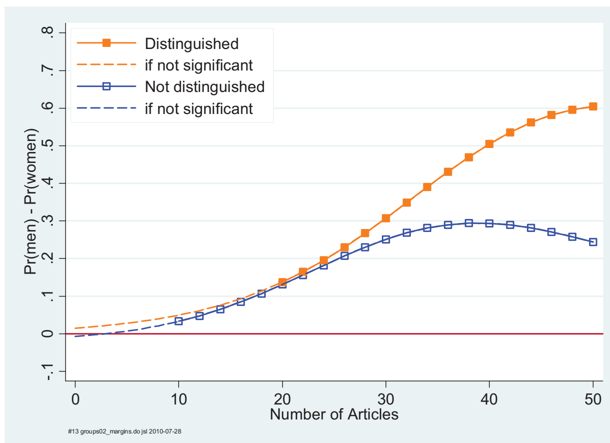
M3: ORs for articles and prestigious jobs

```
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: Plots of all predictions



M3: Alternatively, plots the discrete changes



M4: the full model - OR 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

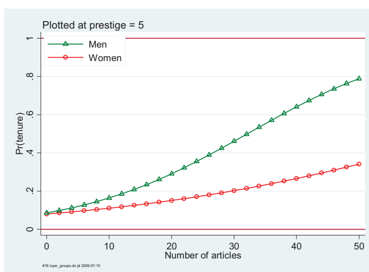
```
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
     e^b = exp(b) = factor change in odds for unit increase in X
 e^bStdX = exp(b*SD of X) = change in odds for SD increase in X
```

M4: the full model - OR for men

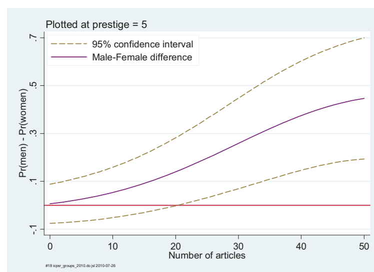
```
logit (N= 2797): Factor Change in Odds

  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: Plotting probabilities and group differences

Probabilities by group

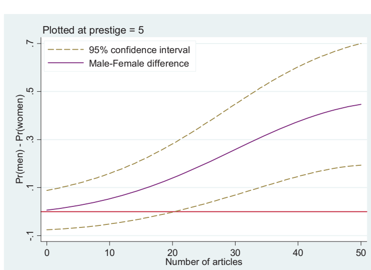


Discrete change with CI

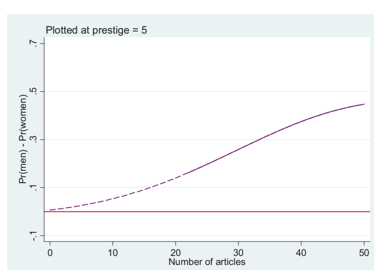


M4: Plotting probabilities and group differences

Discrete change with CI

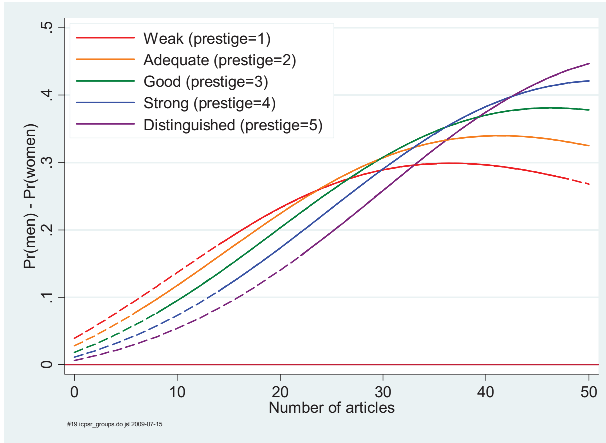


Discrete change with broken line



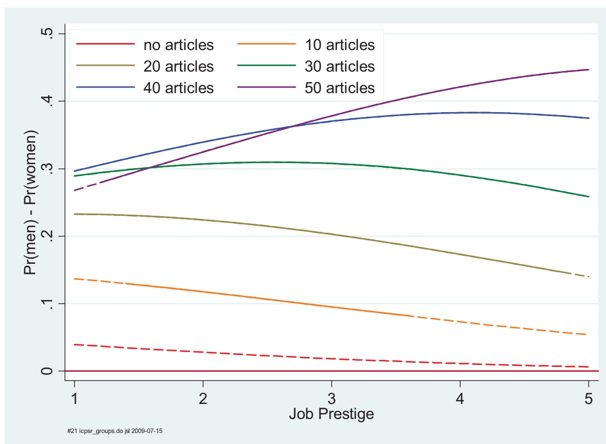
M4: Plotting probabilities and group differences

Holding all other variables constant:

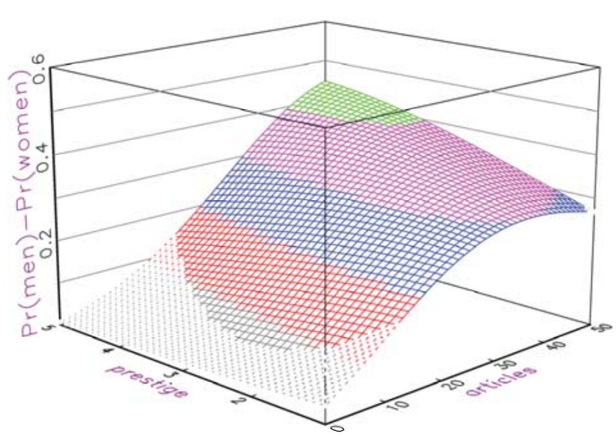


M4: Plotting probabilities and group differences

Holding all other variables constant:



M4: in three dimensions



Summary

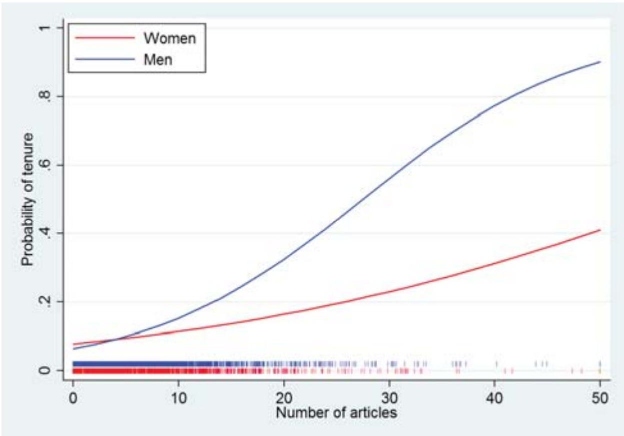
Basic issues

1. In the LRM we can use standard tests to compare coefficients across groups.
2. In the BRM, these tests are problematic due to identification issues.
3. We can make stronger assumptions to test the coefficients.
4. Tests comparing predicted probabilities are unaffected by the identification problem.

Difficulties

1. You must deal with the interpretation of nonlinear models since a single test is not possible. Should you adjust for multiple tests?
2. Do probabilities show the "effect" of (say) articles?
3. Do you have sufficient observations to support your conclusions. Are the preditions "on the support"?

Data support for making generalizations



References

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.

Long, J.S. and Freese, J. 2005. *Regression Models for Categorical and Limited Dependent Variables with Stata*. Second Edition. College Station, TX: Stata Press.

Long, J. Scott, Paul D. Allison, and Robert McGinnis. 1993. Rank Advancement in Academic Careers: Sex Differences and the Effects of Productivity. *American Sociological Review* 58:703-722.

Williams, Richard. 2009. Using Heterogeneous Choice Models to Compare Logit and Probit Coefficients across Groups. *Sociological Methods & Research* 37: 531-559.

Xu, Jun and Long, J. 2005, Confidence intervals for predicted outcomes in regression models for categorical outcomes. *The Stata Journal* 5: 537-559.

Working paper

www.indiana.edu/~jslsoc/index.htm

Software

Stata 10 at www.stata.com with auxiliary ***SPost*** package (see Long and Freese 2005).

Stata 11 using **`margins`**.