Discrete Response Models
Count Models

H.C. Kongsted
Copenhagen Business School, SÍ
hck.si@cbs.dk

Agenda

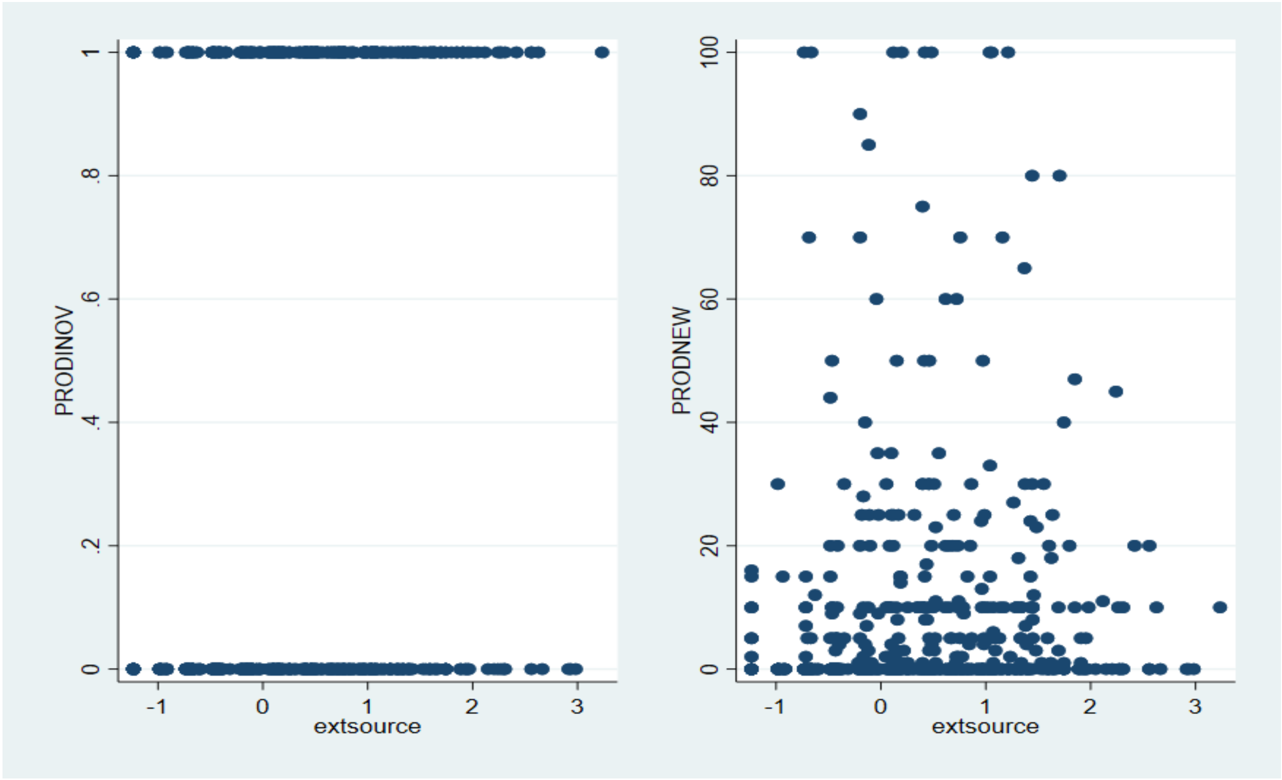
Copenhagen Business School
HANDELSHØJSKOLEN

- Final extension to basic LRM: non-linear transformations of variables
- How to deal with **dependent variables** that have categorical or count outcomes?
- We will look at three (common) cases:

1. Binary response: 0/1 outcome
    - Logit (and probit) models
    - Marginal effects
    - Example: Coefficients and marginal effects
    - Comparing probit and logit models
2. Multinomial responses: More than two (unordered) outcomes
3. Count data: The possible outcomes are the natural counts 0,1,2,3,…

Logit (and probit) models

Binary response: The UK CIS example revisited



How to deal with the binary variable PRODINOV (0/1) as an outcome?

Dichotomous dependent variables

Copenhagen Business School
HANDELSHØJSKOLEN

- Running a linear regression by OLS is still a feasible option.
- In fact, the linear regression model provides a valid model for the *expected value* of Y given X, when the first 4 OLS assumptions are satisfied:

$$E(Y|X) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

- The error term is heteroskedastic: We can fix that using robust s.e.
- Not always fully appropriate: the “expected” value for a particular unit could become negative, or larger than one.
- Readings:
  - Wooldridge, Introductory Econometrics, chapter 7.5, 17.1.
  - Hoetker SMJ 2006.

The logit model

Copenhagen Business School
HANDELSHØJSKOLEN

- A general binary response model:

$$P(y=1\,|\,\mathbf{x}) = G(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k) = G(\mathbf{x\beta}) = G(z)$$

- For the logit model, *G* is the logistic function:

$$G(z) = \exp(z)/[1+\exp(z)]$$

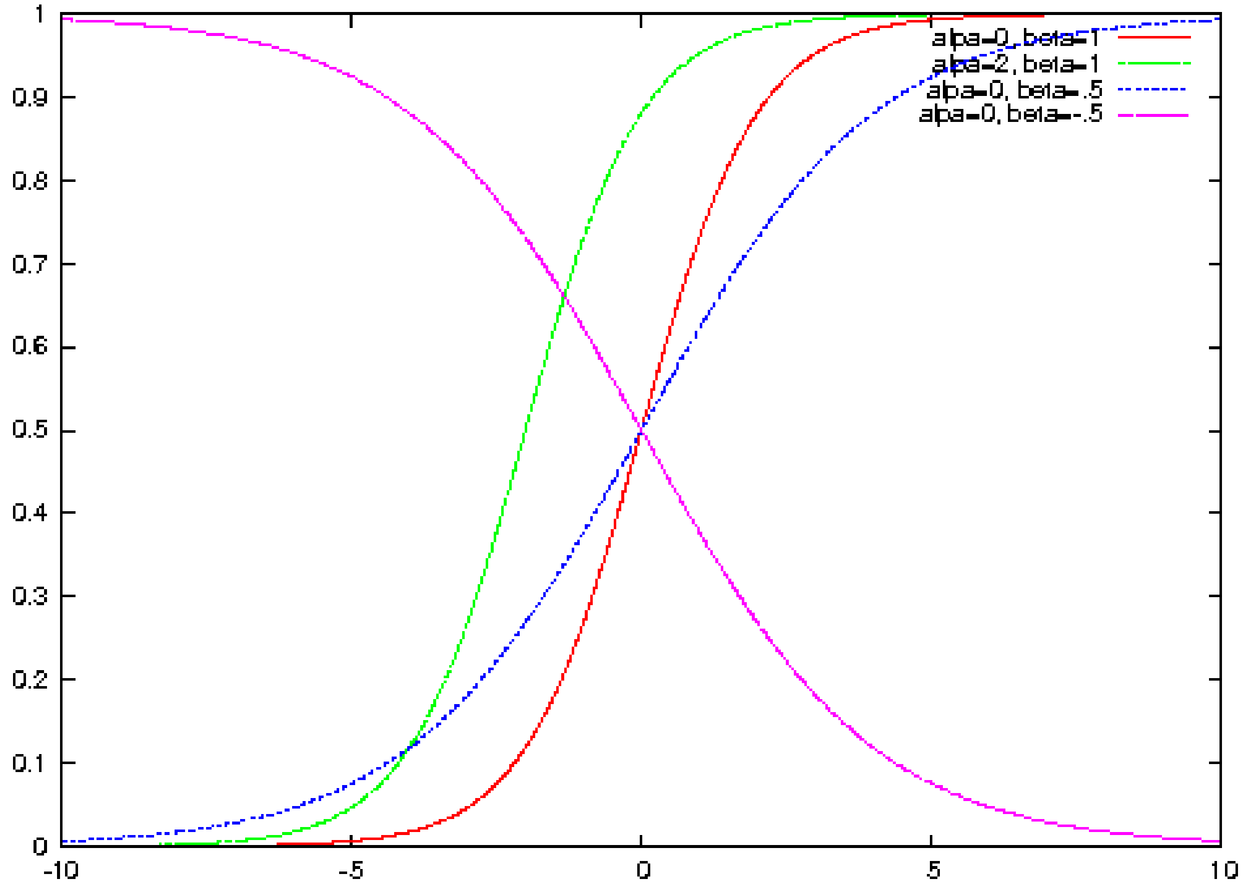
G is the cumulative distribution for a standard logistic random variable. *G* takes values strictly between 0 and 1: 0 < G(z) < 1.

- We use *G* as a model of the ***probability*** that *y* assumes the value 1 rather than 0. But this is also the expected value of Y:

$$E(y\,|\,\mathbf{x}) = 1\cdot P(y=1\,|\,\mathbf{x}) + 0\cdot P(y=0\,|\,\mathbf{x})$$
$$= P(y=1\,|\,\mathbf{x}) = G(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)$$

- The model is estimated using the maximum likelihood method: maximizes **the log-likelihood function** $lnL(\beta)$ -> finds the estimates of $\beta_0, \beta_1, \ldots, \beta_k$ that make the actual outcome of y_i in the sample *i=1,2,…,n* **most likely**, given the values of the x_i's.

Logistic Function for $z = \alpha + \beta x$



The probit model

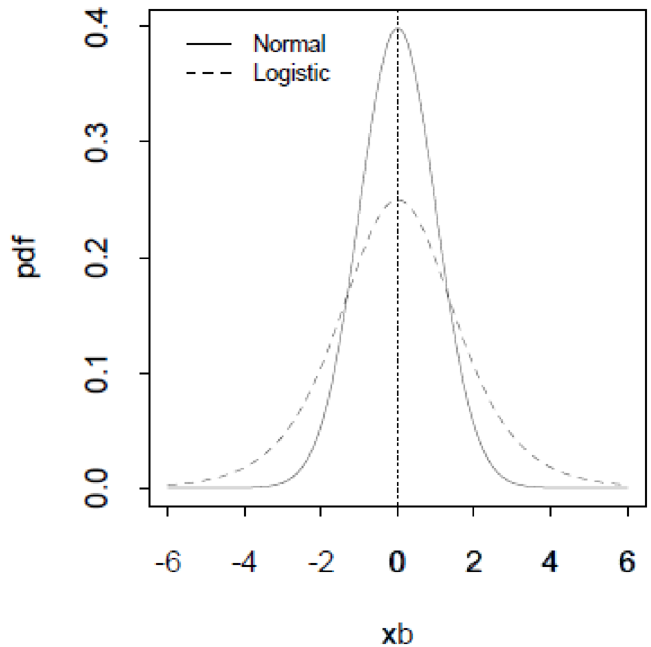
- Probit model: an alternative choice of probability distribution for a binary response model.
- Like logit, the focus is the *probability* that Y equals 1.
- In the probit model, *G* is the standard normal cumulative distribution function, which is expressed as an integral:

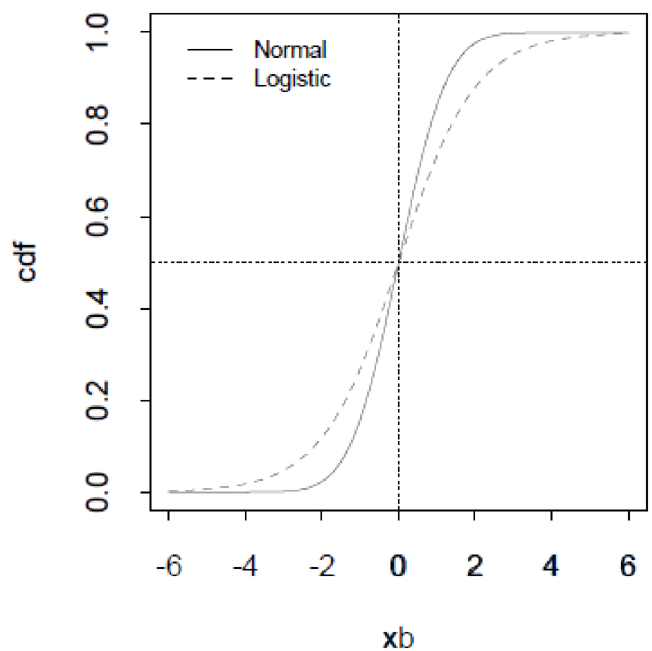
$$G(z) = \Phi(z) \equiv \int_{-\infty}^{z} \phi(v)dv$$

where $\varphi(z)$ is the standard normal density:

$$\phi(z) = (2\pi)^{-1/2} \exp(-z^2/2)$$

- Maximum likelihood is again used to estimate the β parameters.
- In many cases, the logit and probit results are "comparable." Let us focus on the logit case.





Both are symmetric; logistic has larger variance; tail probabilities differ.

Interpretation of coefficients

- Most often we are interested in the marginal effect of changing a given variable, x_j, on $E(Y|X)$, or equivalently, on $P(y=1|\, x)$.
- Because the logistic function is non-linear, the coefficients of the model (the β_j's) do not directly give the marginal effects.
- We can use the coefficients to tell us:
  - Sign: Whether there is a positive or negative effect of x_j, given that the relevant coefficient is statistically significant.
  - Significance: Whether there is an effect at all (is there a statistically significant effect? i.e., $\beta_j \neq 0$).

Example: logitexercise.do

- We estimate a logistic model, whereby product innovation (no=0/yes=1) is determined by whether or not a firm collaborates with a university and the size of the firm:

$$P(y=1|\mathbf{x}) = G(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$$

- For comparison, we also run the simple linear regression.
- *y* is the indicator for innovating (prodinov)
- x_1 measures the size of the firm expressed as the log of number of employees (lempl00).
- x_2 is a binary variable expressing whether or not the firm is collaborating for innovation with a university (unic).

```
/* Note: define unic as: */
gen unic=0
replace unic=1 if punivl==1 | punivn==1 | punive==1 | punivu==1 | punivo==1
```

Logit *vs* linear regression

```
. * OLS for comparison
. reg prodinov lempl00 i.unic, r

Linear regression                               Number of obs     =        906
                                                F(2, 903)         =      54.87
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1154
                                                Root MSE          =     .42903

------------------------------------------------------------------------------
             |               Robust
    prodinov | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     lempl00 |   .0789866   .0107481     7.35   0.000     .0578924    .1000808
      1.unic |   .3990625   .0669843     5.96   0.000     .2675995    .5305255
       _cons |  -.0434095   .0415253    -1.05   0.296    -.1249069    .0380879
------------------------------------------------------------------------------


. logit prodinov lempl00 i.unic

Iteration 0:   log likelihood = -548.43996
Iteration 1:   log likelihood = -498.17582
Iteration 2:   log likelihood = -497.59473
Iteration 3:   log likelihood = -497.59414
Iteration 4:   log likelihood = -497.59414

Logistic regression                                     Number of obs =    906
                                                        LR chi2(2)    = 101.69
                                                        Prob > chi2   = 0.0000
Log likelihood = -497.59414                             Pseudo R2     = 0.0927

------------------------------------------------------------------------------
    prodinov | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     lempl00 |   .4075079   .0568172     7.17   0.000     .2961482    .5188675
      1.unic |   1.798397   .3391331     5.30   0.000     1.133708    2.463086
       _cons |   -2.67987   .2538696   -10.56   0.000    -3.177446   -2.182295
------------------------------------------------------------------------------
```

Copenhagen Business School
HANDELSHØJSKOLEN

Logit: Marginal effects

- Logit coefficients do not directly measure marginal effects.
- The marginal effect corresponds to the effect of a one-unit increase (typically from the mean) of x_j on $P(y=1|\mathbf{x})$, holding other variables constant (typically at their means).
- Indeed, the marginal effect of x_j on the probability of obtaining y=1 (rather than y=0) depends on the value of the other variables in **x**.

Logit: Marginal effects

Copenhagen Business School
HANDELSHØJSKOLEN

- Consider the case where z = $\beta_0 + \beta_1 x_1 + \beta_2 x_2$

  That is:

  $$G(z) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)/[1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)],$$

  Assume that x_1 is continuous: If we differentiate (check out your high school math!) partially with respect to x_1, we get:

  $$\frac{\partial G(z)}{\partial x_1} = \beta_1 \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}{[1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)]^2}$$

- This is the formal expression for the marginal effect of x_1 in a logit model with two explanatory variables.
- From this expression we see that the marginal effect depends on the values of the x's, not only x_1, but also the value of x_2!!

Logit: Marginal effects

Copenhagen Business School
HANDELSHØJSKOLEN

- In other words, the marginal effect of x_1 will be different for different values of x_1 and x_2.
- We can examine the marginal effect at "representative" or "interesting values" of the independent variables.
- The "representative value" is normally taken to be the mean, so that the marginal effect of a given variable is examined for a one-unit increase **from the mean** with all other covariates **set to the their mean values**:
- "Conditional marginal effects" in Stata: `margins, dydx(*) atmeans`
- Or: we could calculate the marginal effect that applies to each firm with its specific values of the x's, and then average over firms:
- "Average marginal effect" in Stata: `margins, dydx(*)`

Same logit regression, but different ways to calculate marginal effects…

```
. logit prodinov lempl00 i.unic

------------------------------------------------------------------------------
    prodinov | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     lempl00 |   .4075079   .0568172     7.17   0.000     .2961482    .5188675
      1.unic |   1.798397   .3391331     5.30   0.000     1.133708    2.463086
       _cons |   -2.67987   .2538696   -10.56   0.000    -3.177446   -2.182295
------------------------------------------------------------------------------

. margins, dydx(*)

. margins, dydx(*) atmeans
```

Use the Stata file logitexercise.do to calculate the marginal effects, and compare.

Logit: Marginal effects

Copenhagen Business School
HANDELSHØJSKOLEN

- So far: looked at the effect of a *continuous* variable on the (expectation of the) binary DV.
- However, if an independent variable, say x_2, is *binary*, the meaningfull effect is the effect of a discrete change from 0 to 1.
- The effect of changing x_2 from 0 to 1, holding other covariates constant, is given as:

$$G(\beta_0 + \beta_1 x_1 + \beta_2) - G(\beta_0 + \beta_1 x_1)$$

- Note: still varies with x_1.
- Stata will recognize this if you mark x_2 as a “factor” variable:
  `logit prodinov lempl00 i.unic`
- Output will say: `Note: dy/dx for factor levels is the discrete change from the base level.`

Measures of model fit

Copenhagen Business School
HANDELSHØJSKOLEN

- The "LR chi2(*k*)" test in the Stata output is similar to F-test of overall significance in linear regression: tells us whether all the estimates in the model combined are (in)significant.
- Under the null that there is in fact no effect of any of the explanatory variables, the test follows a chi-square distribution with *k* degrees of freedom.
- Alternatively, we can look at a (pseudo) R2. However, there are several ones available.
- All measures of fit in models for discrete data have problems (see the Hoetker 2006 SMJ article)
- So do not over-interpret these numbers!
- No simple measure of model fit equivalent to the R^2 exists for models for discrete data.

Logit vs. probit

Copenhagen Business School
HANDELSHØJSKOLEN

- There are usually no compelling theoretical grounds for preferring one over the other.
- If the outcomes in the sample are divided between a large majority and a small minority, results can differ. This is because the observations are then concentrated in a tail of the distribution where the logit and probit functions are somewhat different.
- The coefficient estimates for the logit model are approximately 1.6 times the size of those in the probit model.
- **But the marginal effects are typically strikingly similar (and close to those obtained in a linear regression).**
- Use the Stata file logitexercise.do to calculate the marginal effects for logit and probit, and compare.

Copenhagen Business School
HANDELSHØJSKOLEN

Multinomial logit

When to use a multinomial model?

- The dependent variable is categorical with more than two (mutually exclusive) outcomes
- The variable is nominal, rather than on an ordinal scale ⇔ the order of its categories plays no role in the multinomial model (except for interpretation)

  ⇨ … if outcomes are ordered (bad, OK, good), there are specialized alternatives for that (ordered probit, ordered logit)

- Examples:
  - Occupational choice (paid employment, self-employment, inactive, etc.)
  - Transport mode (bike, car, metro, etc.)

Readings:

- Cameron & Trivedi, Microeconometrics: Methods and Applications, section 15.4.
- Reichstein, T. and A. Salter (2006), Investigating the sources of process innovation among UK manufacturing firms, *Industrial and Corporate Change,* 15, 653–682. **(applied example)**
- Wooldridge, Econometric analysis of cross-section and panel data, section 15.9.1.

The multinomial logit model

- Binary response model: Two outcomes: *y = {0,1}*

$$P(y=1\,|\,\mathbf{x}) = G(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k) = G(x\boldsymbol{\beta})$$

- More than two outcome categories → we need more than one logit function (more than one set of betas)
  - In fact, we generally need *J* different logit functions when there are *J+1* possible outcomes in the dependent variable

- *J+1* outcomes: Label as *y=0, y=1, …, y=J*

$$P(y=j\,|\,\mathbf{x}) = \frac{\exp(x\beta_j)}{1+\sum_{h=1}^{J}\exp(x\beta_h)},\ j=1,2,\ldots,J.$$

- We model the ***probability*** that *y* takes the value *j,* rather than any of the *J* other outcomes, conditional on *x, with a* ***reference outcome*** *(e.g., y = 0)*

Example when J=2 (log odds ratios)

Copenhagen Business School
HANDELSHØJSKOLEN

- If *J=2*, there are three categories → Two logit functions are to be estimated.

- The logit function is conveniently summarized by:

$$
\left.
\begin{aligned}
g_1(\boldsymbol{x}) &= ln\frac{P(Y=1|\boldsymbol{x})}{P(Y=0|\boldsymbol{x})} = \boldsymbol{x}'\boldsymbol{\beta_1} \\
g_2(\boldsymbol{x}) &= ln\frac{P(Y=2|\boldsymbol{x})}{P(Y=0|\boldsymbol{x})} = \boldsymbol{x}'\boldsymbol{\beta_2}
\end{aligned}
\right\}
$$

*Note: The coefficients reflect the **change in the log of the relative risk between y=j and the baseline option** when x changes by 1 unit.*

- Note the **special interpretation of *beta*** in the logit specification: The ***log-odds ratios* are linear in *x*.**

Response probabilities

- Conditional probabilities in the case of three outcomes (0,1,2) are:

$$P(Y=0|\boldsymbol{x}) = \frac{1}{1+e^{g_1(\boldsymbol{x})}+e^{g_2(\boldsymbol{x})}}$$

$$P(Y=1|\boldsymbol{x}) = \frac{e^{g_1(\boldsymbol{x})}}{1+e^{g_1(\boldsymbol{x})}+e^{g_2(\boldsymbol{x})}}$$

Note: Response probabilities must sum to 1.

$$P(Y=2|\boldsymbol{x}) = \frac{e^{g_2(\boldsymbol{x})}}{1+e^{g_1(\boldsymbol{x})}+e^{g_2(\boldsymbol{x})}}$$

- Parameter estimates are found through a **maximum likelihood estimation** technique: Obtained by selecting the values of beta-parameters to make the actual sample most likely.

Multinomial logit in research: Types and sources of process innovation

Copenhagen Business School
HANDELSHØJSKOLEN

- Reichstein, T. & A. Salter (2006), Industrial and Corporate Change
- **Three values** of outcome variable: **No/Incremental/Radical** (they do not to impose any ordering!)
- “No innovation” as the reference category.
- Data from the UK CIS for 2,800+ firms.
- Characteristics of firms and their innovation strategies as explanatory variables.

Copenhagen Business School
HANDELSHØJSKOLEN

Reichstein/Salter example

- Two parameters for each explanatory variable
- E.g., the first describes how the share of sales from products new to the market influences the probability of being an **incremental innovator** <u>compared to</u> not being innovative.
- Corresponding coefficient for the outcome of the firm being a **radical innovator.**
- Wald test to see whether β_{i1} and β_{i2} are significantly different from each other.
- Post estimation (after **mlogit**) in Stata: **test [1]var=[2]var**

Table 3 Determinants of process innovation, results of multinomial logistic regression

Variables	Model 1		
	Incremental versus not innovative	Radical versus not innovative	Wald test
Share of sales from products			
New to the market	0.0176**	0.0407***	+
	(0.01)	(0.01)	
New to the firm	0.0140***	0.0115*	
	(0.00)	(0.01)	
Significantly improved	0.0111***	0.0196***	
	(0.00)	(0.01)	
Cost factor	0.9718***	1.2478***	+
	(0.08)	(0.11)	
Product factor	0.1467*	0.2774**	
	(0.09)	(0.12)	
Suppliers	0.5936***	0.4641***	
	(0.07)	(0.09)	
Customers	−0.1708**	−0.3203***	
	(0.07)	(0.10)	
Consultants	−0.2262***	−0.0744	
	(0.08)	(0.10)	
Universities	−0.0195	−0.0168	
	(0.09)	(0.11)	
Standards and regulations	−0.0277	−0.0329	
	(0.03)	(0.04)	
R & D	0.2633*	0.3905**	
	(0.15)	(0.19)	
Log (size)	0.1587***	0.2386***	
	(0.05)	(0.06)	
Investment expenditure/sales	0.0375	0.0111	
	(0.10)	(0.14)	
Training expenditure/sales	−0.0646	−0.2899	
	(0.38)	(0.83)	
Collaboration	0.6727***	1.2134***	+
	(0.16)	(0.19)	
Intercept	−2.8369***	−4.3807***	
	(0.32)	(0.45)	
Industry dummies		Yes	
Observations		2885	
Likelihood ratio		−1517.1	
Pseudo R^2		0.29	
Maximum variance inflation factor		2.51	

Limitations (1/2): IIA Assumption

Copenhagen Business School
HANDELSHØJSKOLEN

Independence from Irrelevant Alternatives (IIA) assumption

- Adding another alternative or removing one of the outcomes will not change the **relative probabilities** of the others
- Often considered very restrictive (e.g., probability of choosing between car & red bus will not change if a blue bus is also introduced)
- You may use the **mlogtest** post-estimation command in Stata to test the validity of this assumption
- Type *findit mlogtest* (part of the spost9_ado); For a Hausman test type ***mlogtest, hausman base***; for additional tests: ***mlogtest, iia***

If IIA assumption is violated, we may consider other potential models: nested logit or mixed logit *(out of our scope)*

Limitations (2/2): Marginal effects…

Copenhagen Business School
HANDELSHØJSKOLEN

…are complicated:

$$\frac{\partial P(y=j\,|\,\mathbf{x})}{\partial x_k} = P(y=j\,|\,\mathbf{x})\left(\beta_{jk} - \sum_{h=1}^{J}\beta_{hk}\exp(x\beta_h)/(1+\sum_{h=1}^{J}\exp(x\beta_h))\right), j = 1,2,\ldots,J.$$

- There is not necessarily a 1-to-1 correspondence between the beta's and the effect of changing any particular *x* variable, not even in terms of sign.
- Stata example: High school program choice ⇨ `hs_example.do`
- High school students' choice between general, vocational, and academic programs, modelled by their writing score and socio-economic status.

General vs Academic Vocational vs Academic

- https://stats.idre.ucla.edu/stata/dae/multinomiallogistic-regression/

Copenhagen Business School
HANDELSHØJSKOLEN

Count models

Count models

1. **Core model: Poisson regression**
    a. Purpose & assumptions
    b. Outcome of the regression and interpretation issues
    c. Applied examples (papers & exercises)
2. Alternative count models (Negative Binomial, Zero-Inflated Models)

Readings:

- Winkelmann, Econometric analysis of count data (2008), chapter 3.
- Cameron & Trivedi, Microeconometrics: Methods and Applications, chapter 20.
- Kaiser, U., Kongsted, H. C., Rønde, T. (2015), “Does the mobility of R&D labor increase innovation?”, *Journal of Economic Behavior & Organization*, 110, 91-105. **(applied example)**
- Wooldridge, Econometric analysis of cross-section and panel data, chapter 19.

When to use a count model?

- When the dependent variable assumes only counts (non-negative integer values): the number of occurrences of an event within a fixed period of time.
    - Number of patents applied for by a firm in a year
    - Number of emergency room episodes per day
    - Number of trades in a minute
    - Number of traffic incidents in a month
- There is no natural a priori **upper bound**.
- The **outcome is = 0 for at least some members** of the population.

Why not OLS?

Drawbacks of OLS similar to those for binary responses:

- The distribution of the dependent variable is **highly skewed.**
- In count data, $y \geq 0$, so E(y|x) should be non-negative for all x, but:
- with OLS we can have $\widehat{x\beta}<0$ **(negative values).**
- OLS will **predict non-integer values.**

A solution could be to use the log transformation: log(y) and run an OLS. But in count data, y=0 is an important part of the data (as we will see). → How to recover E(y|x) from log(y), or even log (y+1)?

Look for other functional forms that suit such data better.

Distributions of Count Data I: Poisson

Poisson probability distribution:

$$P(Y=y|\lambda)=\frac{e^{-\lambda}\lambda^{y}}{y!} \quad \text{for } y=0, 1, 2\ldots$$

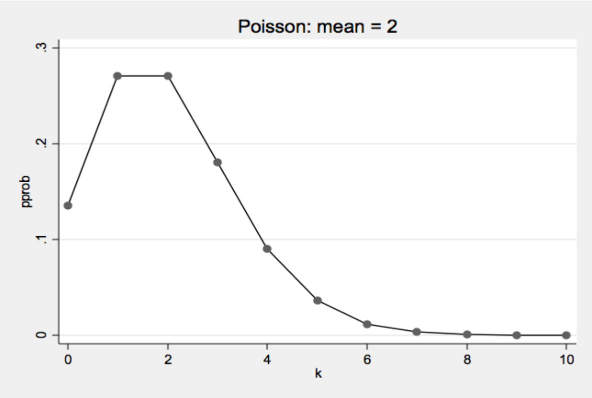
- **λ** is the **mean** or **expected value** of a Poisson distribution: $\boldsymbol{E(Y)} =$ **λ**.

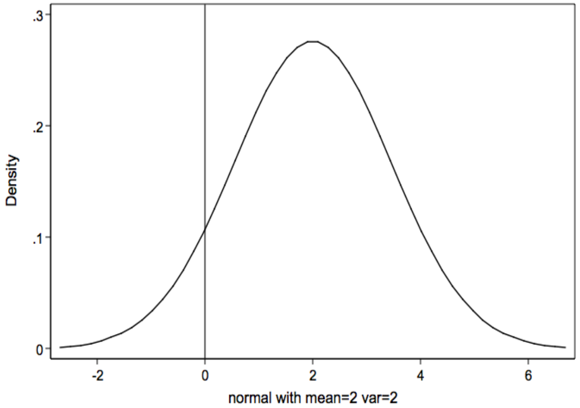
- **λ** is **also the variance** of a Poisson distribution: $\boldsymbol{Var(Y)} =$ **λ.**

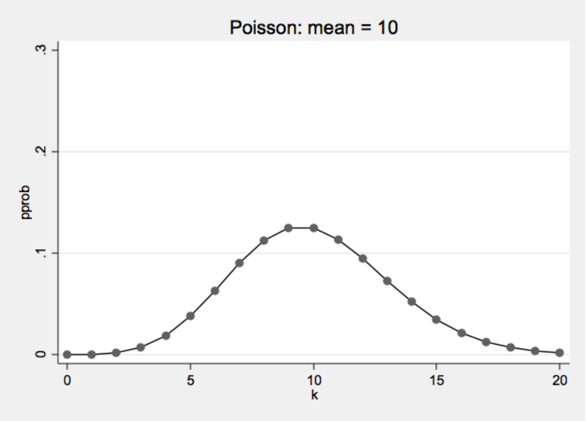
- Poisson is a one parameter distribution (λ).

Examples of Poisson Distribution

Copenhagen Business School
HANDELSHØJSKOLEN







The Poisson distribution implies that the <u>mean of the variable is equal to its variance</u>

Distributions of Count Data II: Negative Binomial

The Negative Binomial takes into account **"over-dispersion"**: The variance often exceeds the mean.

- The Stata default specification is

$$Var(y) = E(y)[1 + \alpha E(y)]$$

- α is the over-dispersion parameter

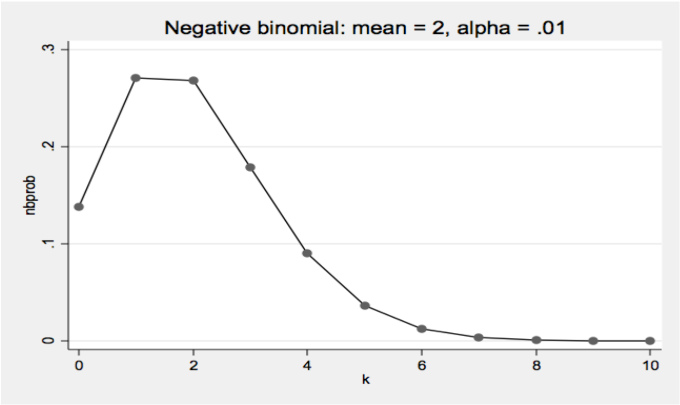
- When α = 0 the negative binomial distribution is the same as a Poisson distribution. Stata provides a test that α = 0. If the data are overdispersed, then a Poisson model will be mis-specified.

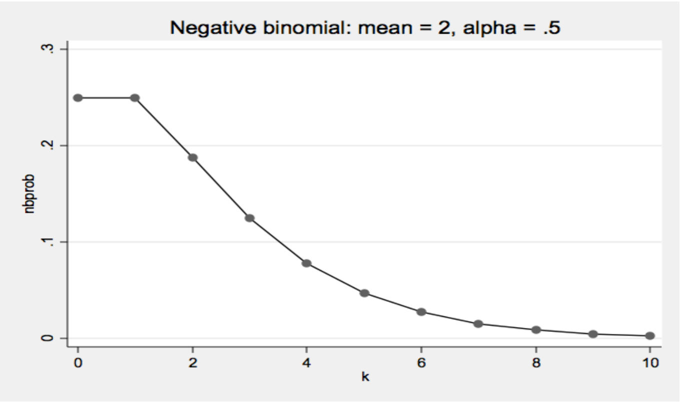
- But consistency of a negative binomial regression relies on correctly specifying the variance equation.

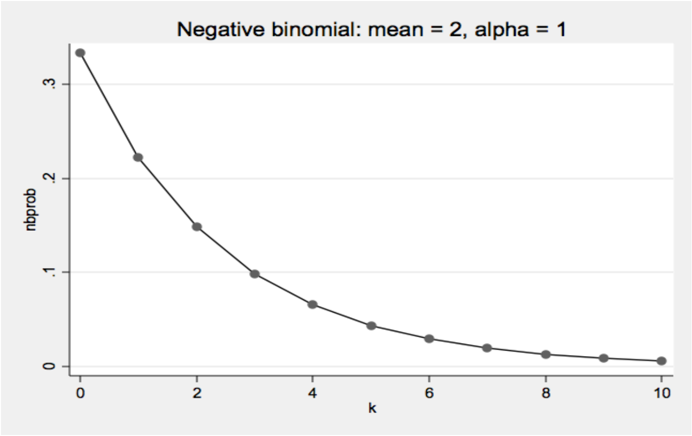
- Poisson consistently estimates the expected value parameter **λ** irrespective of overdispersion. But important to use robust standard errors in that case.

Examples of Negative Binomial Distribution

Copenhagen Business School
HANDELSHØJSKOLEN







Count Data Regressions

Copenhagen Business School
HANDELSHØJSKOLEN

Model the **mean function:**

$$\lambda_i = E[Y_i \mid X_i] = \exp(X_i'\beta) = \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)$$

- Parameterizes the mean function as function of a set of covariates X. Count models are estimated using maximum likelihood.
- Interpretation: Log-linear function → $\log E[Y \mid X] = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$
- Partial effects $\dfrac{\partial E[Y \mid X]}{\partial x_j} = \exp(x\beta)\beta_j$

*Note that **marginal changes differ across individuals.** Report marginal changes at $\bar{x}$ or particular values of x.*

$$\text{Continuous variable: } \%\Delta E(y \mid X) \approx 100\beta_j \Delta x_j$$

$$\text{Dummy variable: } \%\Delta E(y \mid X) = 100(\exp(\beta_j) - 1)$$

*If the explanatory variables are in logs, the coefficients can be interpreted as **elasticities.***

Look up an example to see what that really means

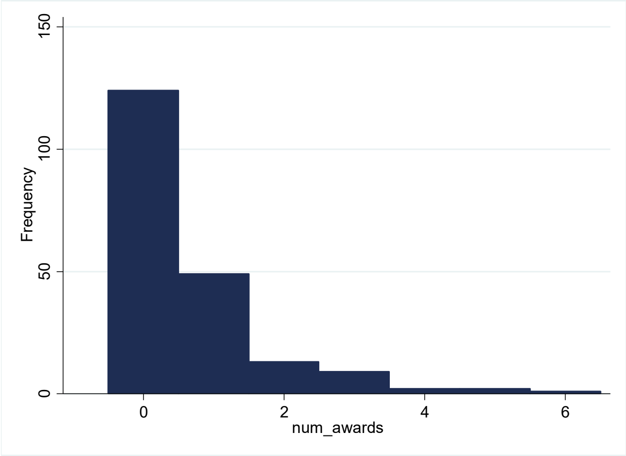
Stata Example 1: Poisson

DV: # Awards earned by students in a certain high school
Predictors: type of program (vocational, general, academic)
score in the final math exam

- Description of the data (Which distribution fits better?)
- Poisson regression (How to read the results?)
- Test the fit of the model (Does Poisson fit the data?)
- Marginal effects & Predicted counts
- Data generated to follow a Poisson distribution

Stata file `count.do`

Stata Example 1: Poisson



- The expected increase in log(# awards) for a 1-unit increase in math score is 0.07: Approx. 7 percent increase in expected number of awards.
- Compared to general program, the expected number of awards is exp(1.08) = 2.96 times the number expected for students in academic programs, or approx. 200% higher.

```
poisson num_awards i.prog math
```

```
Summary for variables: num_awards
     by categories of: prog (type of program)
```

prog	mean	sd	N
general	.2	.4045199	45
academic	1	1.278521	105
vocation	.24	.5174506	50
Total	.63	1.052921	200

```
Poisson regression                              Number of obs   =        200
                                                LR chi2(3)      =      98.22
                                                Prob > chi2     =     0.0000
Log likelihood = -182.75225                     Pseudo R2       =     0.2118
```

num_awards	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
prog						
academic	1.083859	.358253	3.03	0.002	.3816962	1.786022
vocation	.3698092	.4410703	0.84	0.402	-.4946727	1.234291
math	.0701524	.0105992	6.62	0.000	.0493783	.0909265
_cons	-5.247124	.6584531	-7.97	0.000	-6.537669	-3.95658

Goodness of Fit and Incidence Rate Ratios

Copenhagen Business School
HANDELSHØJSKOLEN

```
. poisson, irr

Poisson regression


Log likelihood = -182.75225
```

num_awards	IRR	Std. Err.	z
prog			
academic	2.956065	1.059019	3.03
vocation	1.447458	.6384309	0.84
math	1.072672	.0113695	6.62
_cons	.0052626	.0034652	-7.97

Incidence rate in academic programs is 3 times higher than in general programs, and increases 7% for every unit-increase in math scores.

Does the Poisson model fit our data?

```
estat gof

         Deviance goodness-of-fit =  189.4496
         Prob > chi2(196)         =    0.6182

         Pearson goodness-of-fit  =  212.1437
         Prob > chi2(196)         =    0.2040
```

In this case it does (no wonder, since the data are generated to fit the distribution).

If the test would be **significant**, it would indicate that the data did **not fit** the model well

(e.g., overdispersion?)

Marginal effects and predicted counts: example

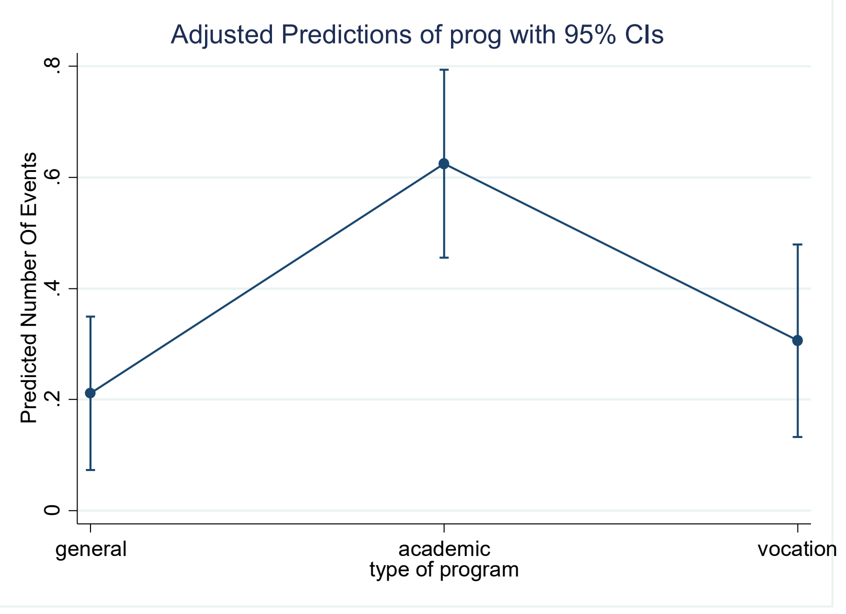
marginsplot

```
. margins prog, atmeans

Adjusted predictions                          Number of obs     =        200
Model VCE    : OIM

Expression   : Predicted number of events, predict()
at           : 1.prog          =        .225 (mean)
               2.prog          =        .525 (mean)
               3.prog          =         .25 (mean)
               math            =      52.645 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        prog |
     general |    .211411   .0705011     3.00   0.003     .0732314    .3495906
    academic |   .6249446   .0862812     7.24   0.000     .4558366    .7940526
    vocation |   .3060086   .0883371     3.46   0.001     .1328711     .479146
------------------------------------------------------------------------------
```



Predicted # awards for general (academic) programs is 0.21 (0.62), holding math scores at its mean level.

Note that 0.6249/0.2114 = 2.96 → IRR we obtained before.

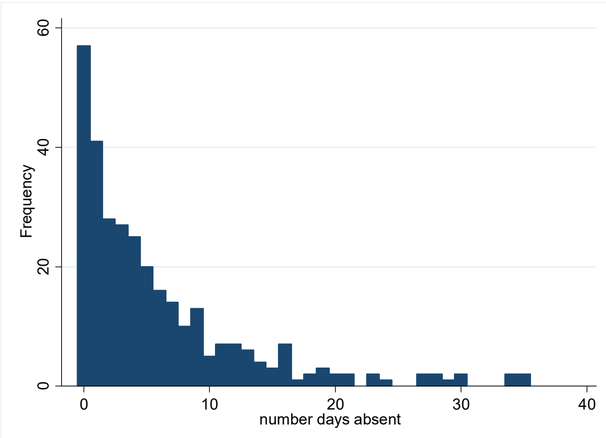
Stata Example 2: Negative Binomial

DV: Number of days absent during the school year (daysabs)

Predictors: Type of program and math scores, as before.

- Description of the data (Which distribution fits better?)
- Poisson & NegBin regression (How to read the results?)
- Test the fit of the model (Does Poisson fit the data?)
- Marginal effects & Predicted counts
- Data generated to follow a Negative Binomial distribution

Stata Example 2: Negative Binomial




```
. tabstat daysabs, by(prog) stats(mean v n)

Summary for variables: daysabs
     by categories of: prog

    prog |      mean  variance         N
---------+------------------------------
       1 |     10.65  67.25897        40
       2 |  6.934132  55.44744       167
       3 |  2.672897  13.93916       107
---------+------------------------------
   Total |  5.955414  49.51877       314
----------------------------------------
```

```
Negative binomial regression                    Number of obs =     314
                                                LR chi2(3)    =   61.69
Dispersion: mean                                Prob > chi2   = 0.0000
Log likelihood = -865.6289                      Pseudo R2     = 0.0344

------------------------------------------------------------------------------
     daysabs | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        math |   -.005993   .0025072    -2.39   0.017     -.010907   -.001079
             |
        prog |
          2  |    -.44076    .182576    -2.41   0.016    -.7986025   -.0829175
          3  |  -1.278651   .2019811    -6.33   0.000    -1.674526    -.882775
             |
       _cons |   2.615265   .1963519    13.32   0.000     2.230423    3.000108
-------------+----------------------------------------------------------------
    /lnalpha |  -.0321895   .1027882                     -.2336506    .1692717
-------------+----------------------------------------------------------------
       alpha |   .9683231   .0995322                      .7916384    1.184442
------------------------------------------------------------------------------
```

Stata `nbreg` will estimate (as the default) the variance relationship:

$$Var(y) = E(y)[1 + \alpha E(y)]$$

Estimates are consistent if this is the correct model for the variance.

Stata Example 2: Alternative: Apply Poisson w/ robust standard errors

```
Poisson regression                                Number of obs =      314
                                                  Wald chi2(3)  =    69.35
                                                  Prob > chi2   =   0.0000
Log pseudolikelihood = -1328.6425                 Pseudo R2     =   0.1431

------------------------------------------------------------------------------
             |               Robust
     daysabs | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        math |  -.0068084   .0023541    -2.89   0.004    -.0114223   -.0021944
             |
        prog |
          2  |  -.4398975   .1420843    -3.10   0.002    -.7183776   -.1614173
          3  |  -1.281364   .1822392    -7.03   0.000    -1.638546   -.9241819
             |
       _cons |   2.651974   .1473325    18.00   0.000     2.363207     2.94074
------------------------------------------------------------------------------
```

Technically: Poisson pseudo maximum likelihood w/ Huber-White
sandwich standard errors

Produces consistent estimates and ”correct” standard errors

In this case: Similar results.

Read more: Winkelmann (2008) chapter 3, pp. 63-126.

Too many zeros?

Copenhagen Business School
HANDELSHØJSKOLEN

When the data contain **“too many” zeros** (compared to Poisson/NegBin).

Two kinds of zeros generated by different processes? **True zeros** vs. **Excess zeros**

- In a **patenting example**, some firms active in R&D may get 0 because of **bad luck**, or they get 0 patents because **they were not active** active in R&D (”always zero”).

Zero-Inflated Models estimate the model in **two parts**:

- **Binary part**: 0 vs. 1

  (if 1, then ”always zero”; if 0, then Poisson or NegBin)

- **Count data part:** Poisson or NegBin distribution. **(zip vs zinb)**

Alternatively: restricting to subsample of ”active” firms based on independent variables (size, sector, pre-sample information).

Count models in applications: Kaiser et al. (2015)

Relationship between worker mobility and firms' inventive output (firm's patent applications per year)

	Poisson PSM		NegBin PSM	
	Coeff.	SE	Coeff.	SE
R&D worker shares				
Joiners from patenting firms	1.543***	0.400	1.608***	0.278
Joiners from non-patenting firms	0.506	0.385	0.362	0.336
Other joiners	1.238***	0.337	1.121***	0.274
Support	0.389	0.333	−0.109	0.203
Leavers to pat. firms	0.916**	0.464	0.668**	0.321
Leavers to non-pat. firms	−0.813	0.773	−0.486	0.424
Capital and R&D labor				
ln(total R&D workers)	0.384***	0.104	0.289***	0.059
ln(capital stock)	0.238***	0.068	0.138***	0.036
Lagged patent status and pre-sample variables				
Dummy patent $t-1$	2.026***	0.366	1.308***	0.138
Dummy patent $t-2$	1.080***	0.122	0.842***	0.107
ln(# pre-sample patents)	0.091	0.120	0.264***	0.087
Dummy pre-sample patent	−0.081	0.278	0.386	0.247
Number of observations and number of firms				
# of obs.	42,507		42,507	
# of firms	14,516		14,516	

Source: Kaiser et al. 2015, JEBO

What's next?

Friday at 9.00 on Zoom:

Workshop 3

Next lecture:

November 9 at 9.00:

Vera Rocha on Attrition and Selection Models