Lecture 10: Alternatives to OLS with limited dependent variables

- PEA vs APE
- Logit/Probit
- Poisson

PEA vs APE

- PEA: partial effect at the average
  - The effect of some x on y for a hypothetical case with sample averages for all x's.
  - This is obtained by setting all Xs at their sample mean and obtaining the slope of Y with respect to one of the Xs.
- APE: average partial effect
  - The effect of x on y averaged across all cases in the sample
  - This is obtained by calculating the partial effect for all cases, and taking the average.

PEA vs APE: different?

- In OLS where the independent variable is entered in a linear fashion (no squared or interaction terms), these are equivalent. In fact, it is an assumption of OLS that the partial effect of X does not vary across x’s.
- PEA and APE differ when we have squared or interaction terms in OLS, or when we use logistic, probit, poisson, negative binomial, tobit or censored regression models.

PEA vs APE in Stata

- The “margins” function can report the PEA or the APE. The PEA may not be very interesting because, for example, with dichotomous variables, the average, ranging between 0 and 1, doesn’t correspond to any individuals in our sample.
  - . “margins, dydx(x) atmeans” will give you the PEA for any variable x used in the most recent regression model.
  - . “margins, dydx(x)” gives you the APE

PEA vs APE

- In regressions with squared or interaction terms, the margins command will give the correct answer only if factor variables have been used
- http://www.public.asu.edu/~gasweete/crj604/misc/factor_variables.pdf

Limited dependent variables

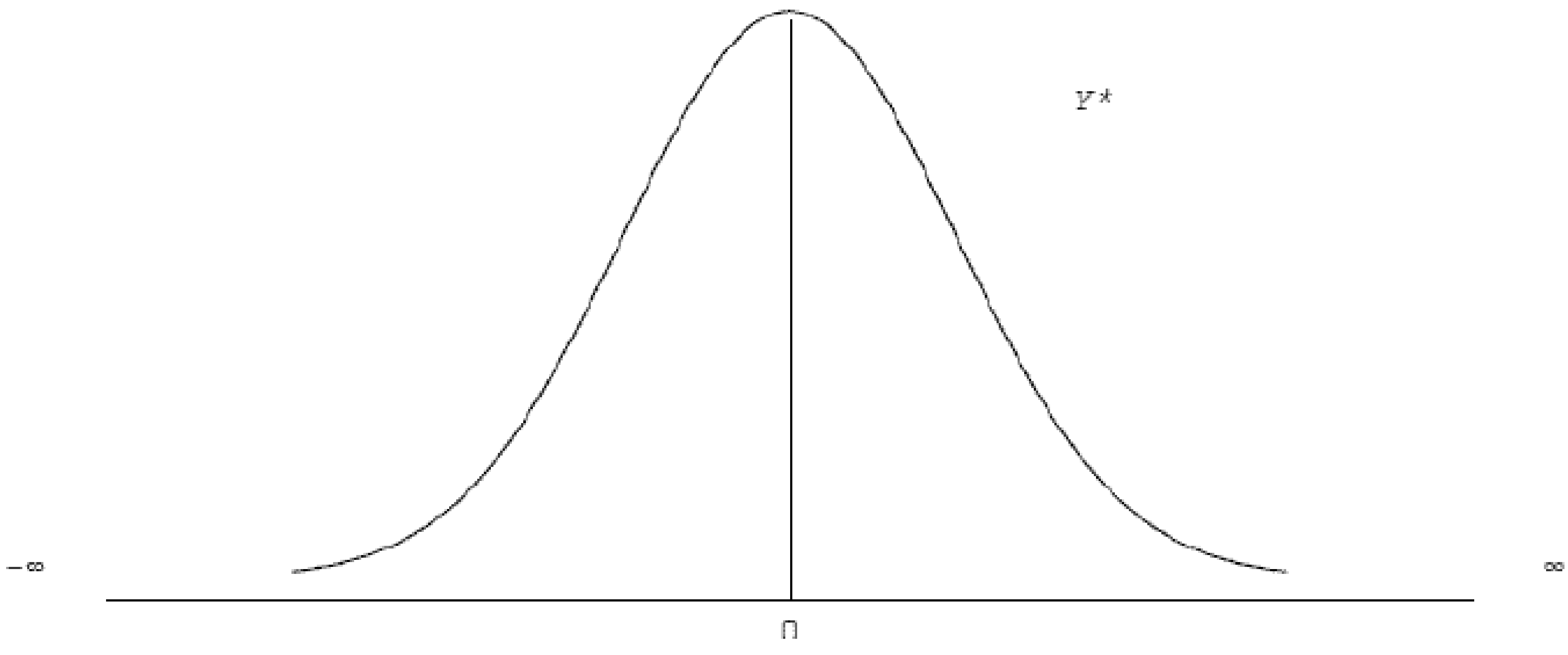
- Many problems in criminology require that we analyze outcomes with very limited distributions
  - Binary: gang member, arrestee, convict, prisoner
  - Lots of zeros: delinquency, crime, arrests
  - Binary & continuous: criminal sentences (prison or not & sentence length)
  - Censored: time to re-arrest
- We have seen that large-sample OLS can handle dependent variables with non-normal distributions. However, sometimes the predictions are nonsensical, and often they are hetoroskedastic.
- Many alternatives to OLS have been developed to deal with limited dependent variables.

Review of problems of the LPM

- Recall, the Linear probability Model uses OLS with a binary dependent variable. Each coefficient represents the expected change in the probability that Y=1, given a one point change in each x.
- While it is easy to interpret the results, there are a few problems.
  - Nonsensical predictions: above 1, below 0
  - Heteroskedasticity
  - Non-normality of errors: for any set of x’s the error term can take on only two values: y minus yhat, or negative yhat
  - Linearity assumption: requiring that X has equal effect across other Xs is not practical. There are diminishing returns approaching 0 or 1.

Binary response models (logit, probit)

- There exists an underlying response variable Y* that generates the observed Y (0,1).



Binary response models (logit, probit)

- Y* is continuous but unobserved. What we observe is a dummy variable Y, such that:

$$Y_i = \begin{cases} 1 \, if \; Y_i^* > 0 \\ 0 \, if \; Y_i^* \leq 0 \end{cases}$$

- When we incorporate explanatory variables into the model, we think of these as affecting Y*, which in turn, affects the probability that Y=1.

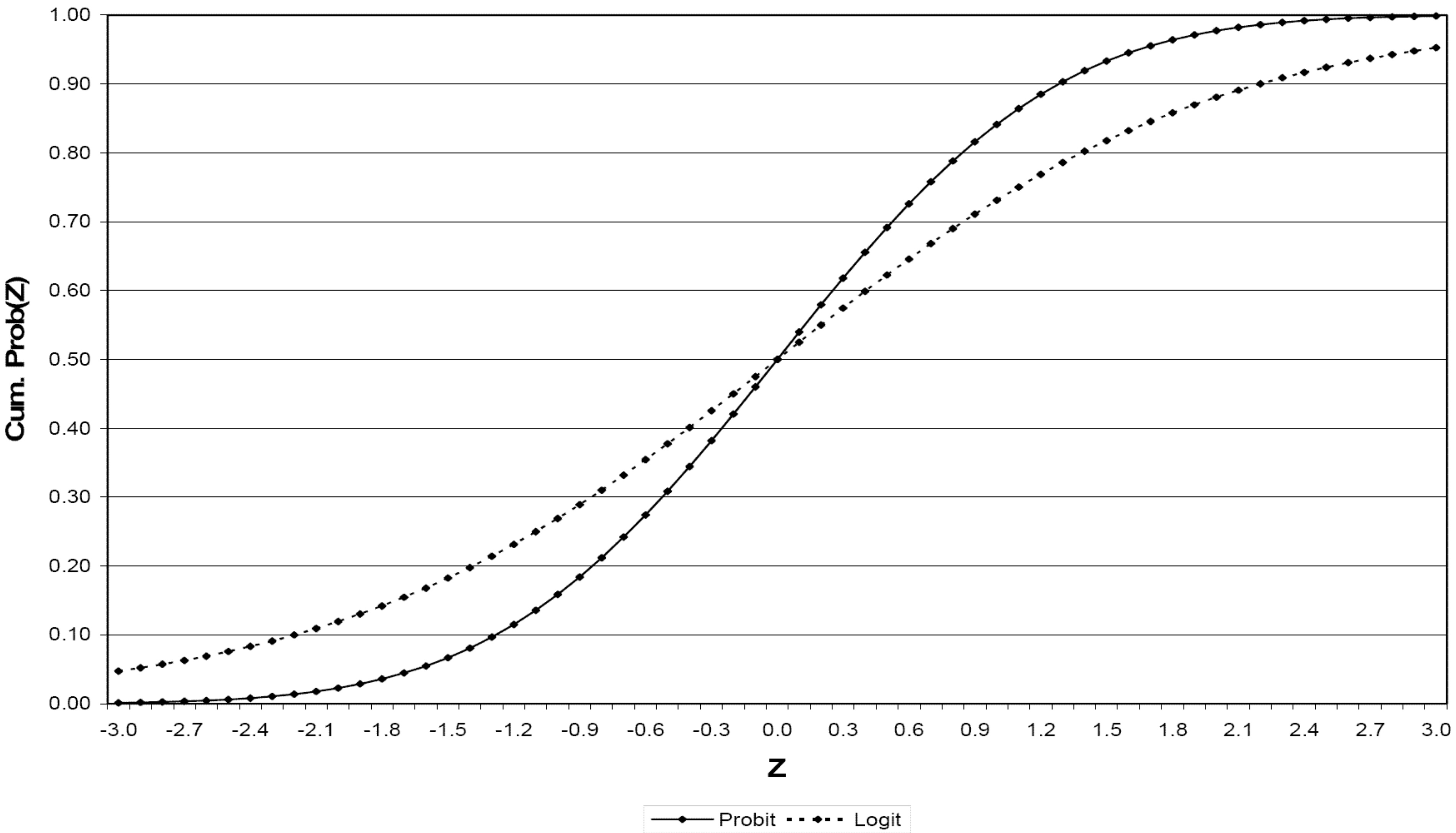
Binary response models (logit, probit)

- This leads to the following relationship:

$$E(Y) = P(Y = 1) = P(Y^* > 0)$$

- We generally choose from two options for modeling Y*
  - normal distribution (probit)
  - logistic distribution (logit)
- In each case, using the observed Xs, we model the area under the probability distribution function (max=1) up to the predicted value of Y*. This becomes P(Y=1) or the expected value of Y given Xs.

Probit and logit cdfs



Probit and logit models, cont.

- Clearly, the two distributions are very similar, and they'll yield very similar results.
- The logistic distribution has slightly fatter tails, so it's better to use when modeling very rare events.
- The function for the logit model is as follows:

$$\hat{y} = P(y = 1) = \frac{\exp(\hat{y}^*)}{1 + \exp(\hat{y}^*)}$$

Logit model reporting

- In Stata, at least two commands will estimate the logit model
  - Logit Y X reports the coefficients
  - Logistic Y X reports odds ratios
- What's an odds ratio?
- Back up, what's an odds?
- An odds is a ratio of two numbers. The first is the chances an event will happen, the second are the relative chances it won't happen.
- The odds that you roll a 6 on a six-sided die is 1:5, or .2
- The *probability* that you roll a 6 is 1/6 or about .167

Logit model reporting

- Probabilities and odds are directly related. If *p* is the probability that an event occurs, the odds are p/(1-p)
  - P=1, odds=undefined
  - P=.9, odds=.9/.1=9
  - P=.5, odds=.5/.5=1
  - P=.25, odds=.25/.75=1/3
- Likewise, if the odds of an event happening is equal to q, the probability p equals q/(1**+**q)
  - Odds=5, p=5/6=.833
  - Odds=1.78, p=1.78/2.78=.640
- Okay, now what's an odds ratio? Simply the ratio between two odds.

Logit model reporting

- Suppose we say that doing all the homework and reading doubles the odds of receiving an A in a course. What does this mean?
- Well, it depends on what the original odds of receiving an A in course.

Original odds	New odds	Original *p*	New *p*	Δp
5	10	.83	.91	.08
1	2	.50	.67	.17
.75	1.5	.43	.60	.17
.3333	.6666	.25	.40	.15
.01	.02	.0099	.0196	.0097

Logit model reporting

- So what does this have to do with logit model reporting?
- Raw coefficients, reported using the “logit” command in Stata, can be converted to odds ratios by exponentiating them: $\exp(\beta_j)$
- Let’s look at an example from Sweeten (2006), a model predicting high school graduation. Odds ratios are reported . . .

Table 2 Logistic regression estimates of the effect of official intervention on high school graduation (full sample, $N = 2501$)[a]

Independent variables	1	2	3	4
Arrest (16-17)	.37(−4.07)	—	.57(−1.67)	.65(−1.14)
Court (16-17)	—	.28(−4.15)	.46(−1.83)	.44(−1.73)
Delinquency scale (<16)	.76(−3.46)	.76(−3.49)	.76(−3.40)	.98(−0.22)
Delinquency scale (16-17)	.84(−2.32)	.84(−2.44)	.85(−2.17)	.91(−1.29)
Poverty	.39(−5.03)	.38(−5.21)	.39(−5.09)	.41(−4.35)
PIAT math score	1.02 (7.68)	1.02 (7.59)	1.02 (7.59)	1.01 (2.99)
African American	1.29 (1.16)	1.31 (1.21)	1.30 (1.19)	1.94 (2.57)
Hispanic	2.20 (3.51)	2.20 (3.50)	2.18 (3.47)	2.80 (3.88)
Male	1.22 (1.26)	1.25 (1.41)	1.25 (1.38)	1.65 (2.93)
Grade retention (<16)				.44(−2.79)
Grade retention (16-17)				.14(−7.10)
School suspension (<16)				.53(−3.17)
School suspension (16-17)				.56(−2.91)
Middle school g.p.a. (0-4)				2.17 (7.15)
Both biological parents				1.31 (1.49)
Waves 2-5				.92(−0.43)

[a]Odds ratios are reported, followed by t statistics. All observations are weighted, and standard errors are adjusted for design effects. Because a pseudo-likelihood procedure was used in estimation, likelihood ratios are not reported.

Marginal effects in logistic regression

- You have several options when reporting effect size in logistic regression.
- You can stay in the world of odds ratios, and simply report the expected change in odds for a one unit change in X. Bear in mind, however, that this is not a uniform effect. Doubling the odds of an event can lead to a 17 percentage point change in the probability of the event occurring, down to a near-zero effect.
- You can report the expected effect at the mean of the Xs in the sample. (margins command)

Marginal effects in logistic regression, cont.

- If there is a particularly interesting set of Xs, you can report the marginal effect of one X given the set of values for the other Xs.
- You can also report the average effect of X in the sample (rather than the effect at the average level of X). They are different.

Marginal effects in logistic regression, example

- Use the dataset from the midterm: mid14nlsy.dta
- Let’s predict the outcome “dpyounger” (dating partner is younger than self) using the following variables: male, age, age squared, in high school, in college, relationship quality
- What is the partial effect at the average for: male, age
- What is the average partial effect for these variables?
- What do these partial effects mean?

Goodness of fit

- Most stat packages report pseudo-r2. There are many different formulas for psuedo-r2. Generally, they are more useful in comparing models than in assessing how well the model fits the data.
- We can also report the percent of cases correctly classified, setting the threshold at $p>.5$, or preferably at the average *p* in the sample.
- Careful though, with extreme outcomes, it's very easy to get a model that predicts nearly all cases correctly without predicting the cases we want to predict correctly.

Goodness of fit, cont.

- For example, if only 3% of a sample is arrested, an easy way to get 97% accuracy in your prediction is to simply predict that nobody gets arrested.
- Getting much better than 97% accuracy in such a case can be very challenging.
- The “estat clas” command after a logit regression gives us detailed statistics on how well we predicted Y.
- Specificity: true negatives/total negatives, % of negatives identified, goes down as false positives go up
- Sensitivity: true positives/total positives, % of positives identified, goes down as false negatives go up

Goodness of fit, cont.

- “estat clas” also gives us the total correctly classified. All these number change depending on the threshold used.
- “lsens” shows the relationship between threshold, sensitivity and specificity.
- “lroc” shows the relationship between the false positive rate (X-axis) and the true positive rate (Y-axis). If you want to have more true positives, you need to accept more false positives. (roc=“receiver operating characteristic”)
- “lroc” also reports area under the curve. The maximum is 1, which is only attainable in a perfect model (100% true positives & 0% false positives). Generally, the closer you are to 1, the better the model is.

Probit model

- The probit model is quite similar to the logit model in the setup and post-estimation diagnostics. However, the coefficients are not exponentiated and interpreted as odds ratios.
- Rather, coefficients in probit models are interpreted as the change in the Z-score for the normal distribution associated with a one unit increase in x.
- Clearly, the magnitude of a change then depends on where you begin on the normal curve, which depends on the values of the other Xs. Also, at extreme values, the absolute effect of changes in X diminish.

Logit Stata exercise

- Use the midterm nlsy data:
- **http://www.public.asu.edu/~gasweete/crj604/midterm/mid14nlsy.dta**
- Calculate a model of predictors of discussing marriage using male, age, dating duration, relationship quality and an interaction term between male and dating duration
  - Report the odds ratio for male when dating duration is 2 years
  - Report the odds ratio for dating duration for females
  - Report the PEA/APE for age and male
  - What are the sensitivity and specificity using .5 as the threshold. How does this change when the sample mean for discussing marriage is used?
  - Is this a good model for predicting discussing marriage? Use the psuedo-r2 and lroc graph

Poisson model

- We may use a Poisson model when we have a dependent variable that takes on only nonnegative integer values [0,1,2,3, . . .]
- We model the relationship between the dependent an independent variables as follows

$$E(y \mid x_1, x_2, \ldots, x_k) = \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)$$

Poisson model, interpreting coefficients

- Individual coefficients can be interpreted a few different ways. First, we can multiply the coefficient by 100, and interpret it as an expected percent change in Y:

$$\% \Delta E(y \mid X) \approx (100\beta_j)\Delta x_j$$

- Second, we can exponentiate the coefficient, and interpret the result as the "incident rate ratio" – the factor by which the count is expected to change

$$e^{\beta_j} = IRR$$

Poisson model, interpreting coefficients example

- Let’s run a model using the midterm data “mid14nlsy.dta” predicting the number of days out of the past 30 that one has had alcohol.

```
. poisson ac30dy06 male age6 antipeer inhs6 college6

Iteration 0:   log likelihood =  -10755.27
Iteration 1:   log likelihood = -10755.253
Iteration 2:   log likelihood = -10755.253

Poisson regression                                Number of obs   =       2496
                                                  LR chi2(5)      =    1018.71
                                                  Prob > chi2     =     0.0000
Log likelihood = -10755.253                       Pseudo R2       =     0.0452

------------------------------------------------------------------------------
    ac30dy06 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   .3520078   .0201906    17.43   0.000     .3124349    .3915807
        age6 |    .123084   .0083049    14.82   0.000     .1068067    .1393614
    antipeer |   .0397006   .0067869     5.85   0.000     .0263985    .0530027
       inhs6 |  -.3275732   .0456022    -7.18   0.000    -.4169518   -.2381947
    college6 |   .0908153    .020916     4.34   0.000     .0498206      .13181
       _cons |  -1.365047    .167794    -8.14   0.000    -1.693917   -1.036177
------------------------------------------------------------------------------
```

Poisson model, interpreting coefficients example

- Let’s focus on the effect of age in this model. The coefficient on age is .123.
- Using the first method of interpretation, we multiply this coefficient by 100, and conclude that for each additional year, youths drink 12.3% more days.
- In the second method, we have to exponentiate .123 to obtain 1.13. Now we say that for each additional year, the expected number of days drinking alcohol increases by a factor of 1.13, or 13%.
    - The IRRs can also be obtained by using the “, irr” option with the poisson command.
- What about the PEA and the APE?

Poisson model, interpreting coefficients example

- The PEA and the APE can be obtained the same way they are obtained after any other regression.

```
. margins, dydx(age) atmeans

Conditional marginal effects                      Number of obs   =       2496
Model VCE    : OIM

Expression   : Predicted number of events, predict()
dy/dx w.r.t. : age6
at           : male            =     .4619391 (mean)
               age6            =      20.2311 (mean)
               antipeer        =     1.672676 (mean)
               inhs6           =     .1129808 (mean)
               college6        =     .4467147 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        age6 |   .4784335   .0319982    14.95   0.000     .4157182    .5411489
------------------------------------------------------------------------------
```

- The partial effect at the average is .48. So for the average individual in the sample, an additional year increases the number of days drinking alcohol by .48.

Poisson model, interpreting coefficients example

- The APE is slightly different: .50. This means that an additional year is expected to increase the expected count of days drank alcohol by .50.

```
. margins, dydx(age)

Average marginal effects                          Number of obs   =       2496
Model VCE    : OIM

Expression   : Predicted number of events, predict()
dy/dx w.r.t. : age6
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
age6	.5040725	.034375	14.66	0.000	.4366987	.5714463

Poisson model, interpreting coefficients example

- How does the average partial effect of .50 square with our initial interpretation that an additional year increases the expected count of days drank alcohol by 12.3 (or 13) percent?
- The average days drank alcohol in this sample is 4.09. A 12.3% increase over that would be .50. So the interpretation of the coefficient is the same – one is in percent terms and the other is in terms of actual units in the dependent variable.
- When reporting results of Poisson regressions, you may want to report effect sizes in one or more of these ways. I find the APE or PEA are the most concrete.
- You can also report the partial effect for specific examples:

Poisson model, interpreting coefficients example

- For somebody with a higher risk profile to begin with, age is even more consequential because they have a higher base rate which age is proportionally increasing.
- A 20 year old college male with antisocial peers is expected to increase his drinking days by .70 in a years time.

```
. margins, dydx(age) at(male=1 age6=20 antipeer=5 inhs6=0 college6=1)

Conditional marginal effects                      Number of obs   =       2496
Model VCE    : OIM

Expression   : Predicted number of events, predict()
dy/dx w.r.t. : age6
at           : male            =           1
               age6            =          20
               antipeer        =           5
               inhs6           =           0
               college6        =           1

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        age6 |   .6998226   .0416827    16.79   0.000      .618126    .7815191
------------------------------------------------------------------------------
```

Poisson model, exposure

- The standard Poisson model assumes equal “exposure.” Exposure can be thought of as opportunity or risk. The more opportunity, the higher the dependent variable. In the example, exposure is 30 days for every person. But it’s not always the same across units:
  - Delinquent acts since the last interview, with uneven times between interviews.
  - Number of civil lawsuits against corporations. The exposure variable here would be the number of customers.
  - Fatal use of force by police departments. Here the exposure variable would be size of the population served by the police department, or perhaps number of officers, or some other variable capturing opportunities to use force.

Poisson model, exposure

- Exposure can be incorporated into the model using the ", exposure(x)" option where "x" is you variable name for exposure.
  - This option inserts logged exposure into the model with a coefficient fixed to 1. It's not interpreted, but just adjusts your model so that exposure is taken into account.

Poisson model, the big assumption

- The poisson distribution assumes that the variance of Y equals the mean of Y. This is usually not the case.
- To test this assumption, we can run “estat gof” which reports two different goodness-of-fit tests for the Poisson model. If the p-value is small, our model doesn’t fit well, and we may need to use a different model.
- Often, we turn to a negative binomial regression instead, which relaxes the poisson distribution assumption.

Negative binomial model example

- The syntax and interpretation of the negative binomial model is nearly exactly the same. It has one additional parameter to relax the Poisson assumption.

```
. nbreg ac30dy06 male age6 antipeer inhs6 college6

Negative binomial regression                      Number of obs   =       2496
                                                  LR chi2(5)      =     107.16
Dispersion     = mean                             Prob > chi2     =     0.0000
Log likelihood = -5996.5149                       Pseudo R2       =     0.0089

------------------------------------------------------------------------------
    ac30dy06 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   .3492844    .062688     5.57   0.000     .2264182    .4721507
        age6 |   .1224952    .026124     4.69   0.000      .071293    .1736973
    antipeer |   .0326278   .0210445     1.55   0.121    -.0086187    .0738742
       inhs6 |  -.3466635   .1162288    -2.98   0.003    -.5744677   -.1188593
    college6 |   .0991895   .0665571     1.49   0.136    -.0312599     .229639
       _cons |  -1.341267   .5215261    -2.57   0.010    -2.363439   -.3190943
-------------+----------------------------------------------------------------
    /lnalpha |   .7308955    .037055                      .6582692    .8035219
-------------+----------------------------------------------------------------
       alpha |    2.07694   .0769609                      1.931446    2.233393
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) = 9517.48 Prob>=chibar2 = 0.000
```

Negative binomial model

- “Alpha” is the additional parameter, which is used in modeling dispersion in the dependent variable. If alpha equals zero, you should just use a Poisson model.
- Stata tests the hypothesis that alpha equals zero so that you can be sure that the negative binomial model is preferable to the Poisson (when the null hypothesis is rejected).
- Another option is a Zero-Inflated Poisson model, which is essentially a two-part model: a logit model for zero-inflation and a poisson model for expected count.
  - We won’t go into this model in detail, but it’s the “zip” command if you’re interested.
  - More info here: http://www.ats.ucla.edu/stat/stata/output/Stata_ZIP.htm

Tobit models

- Tobit models are appropriate when the outcome y is naturally limited in some way. The example in the book is spending on alcohol. For many people, spending on alcohol is zero because they don’t consume alcohol, and for those who do spend money on alcohol, spending generally follows a normal curve.
- There are two processes of interest here: the decision to spend money on alcohol, and how much money is spent on alcohol.

Censored regression

- Use censored regression when the true value of the dependent variable is unobserved above or below a certain known threshold.
  - Censoring is a data collection problem. In the tobit model, we observe the true values of y, but their distribution is limited at certain thresholds.
  - In stata, “cnreg” will give censored regression results. It requires that you create a new variable with the values of 0 for uncensored cases, 1 for right censored cases, and -1 for left censored. If this variable were called “apple”, for example, you’d write: “cnreg y x, censored(apple)”

Truncated regression

- Use truncated regression when the sample itself is a subset of the population of interest. Some cases are missing entirely.
- The truncreg command in Stata will produce truncated regression estimates
- All the same postestimation commands are available

Sample selection correction

- Truncated regression is used when cases above or below a certain threshold in y are unobserved.
- Sample selection correction is sometimes necessary when cases are dropped by more complicated selection processes.
- Often the analysis sample is not the same as the sample originally drawn from the population of interest. Listwise deletion of independent and dependent variables is a common problem that can lead to dramatically smaller samples.
- If the analysis sample is limited in systematic ways, model estimates are no longer representative of the population.

Matching

- Matching is analogous to regression, used for the purpose of identifying the effect of a binary "treatment" variable on some outcome of interest.
- In the language of Heckman & Hotz (1999), regression and matching methods are "selection on observables" strategies. They both assume that the lion's share of "selection" into the treatment of interest is observed. That is, it's measured by variables to which you have access. The main difference between matching and regression methods are:
  - Parametric assumptions dropped
  - Assumption of common support dropped
  - ATT, ATE and ATU effects estimated