Ordinal Regression Models

Categorical and Limited Dependent Variables
Paul A. Jargowsky

Probit: Latent Variable Approach

$$Y_i^* = \mathbf{x}_i \boldsymbol{\beta} + u_i \qquad u_i \sim N\left(0, \sigma^2\right)$$

For example: $$Y_i^* = \beta_1 + \beta_2 X_i + u_i$$

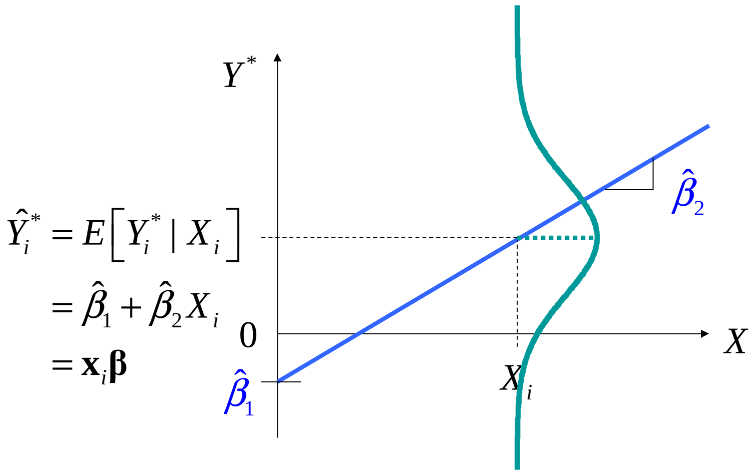
But Y_i* is a latent (unobserved) variable, so we can assume any scale. Assume variance = 1.

$$u_i \sim N\left(0,1\right)$$

However, only Y_i is observed:
- If $Y_i^*>0$ then $Y_i=1$
- If $Y_i^*\leq 0$ then $Y_i=0$

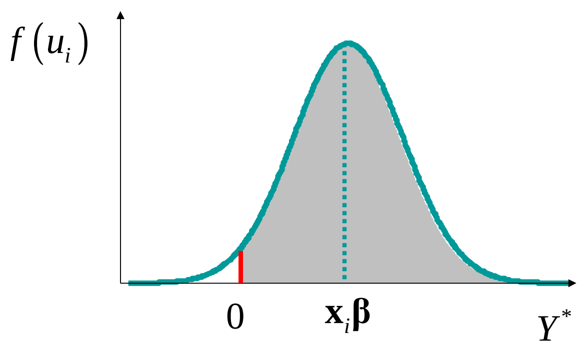
Examples: either you are married or not, but different people have more or less propensity to be married.

Graphic View



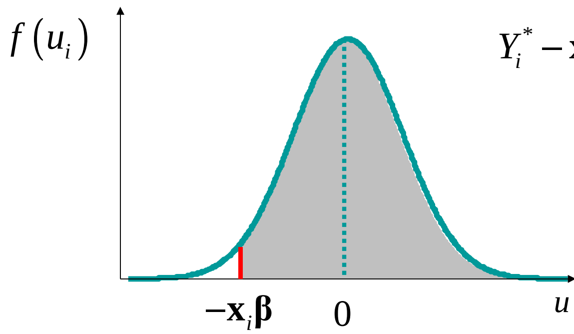
Given a specific X value, the probability of Y=1 depends on the distribution of the disturbance term.

Probability Distribution of u_i



$$\Pr\left(Y_i = 1 \mid X_i\right) = \Pr\left(\mathbf{x}_i\boldsymbol{\beta} + u_i > 0\right)$$

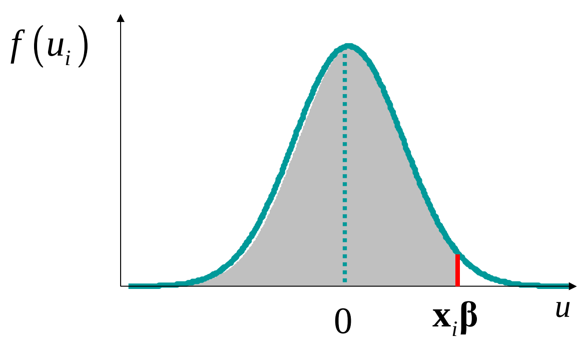
Subtract $\mathbf{X_i\beta}$



$$Y_i^* - \mathbf{x}_i\boldsymbol{\beta} = u_i$$

$$\begin{aligned}\Pr\left(Y_i = 1 \mid X_i\right) &= \Pr\left(\mathbf{x}_i\boldsymbol{\beta} + u_i > 0\right)\\ &= \Pr\left(u_i > -\mathbf{x}_i\boldsymbol{\beta}\right)\end{aligned}$$

Exploit Symmetry



$$\begin{aligned}\Pr\left(Y_i = 1 \mid X_i\right) &= \Pr\left(\mathbf{x}_i\boldsymbol{\beta} + u_i > 0\right)\\ &= \Pr\left(u_i > -\mathbf{x}_i\boldsymbol{\beta}\right)\\ &= \Pr\left(u_i < \mathbf{x}_i\boldsymbol{\beta}\right) = \Phi\left(\mathbf{x}_i\boldsymbol{\beta}\right)\end{aligned}$$

In General

$$\Pr\left(Y_i = 1 \mid \mathbf{x}_i\right) = \Pr\left(u_i < \mathbf{x}_i\boldsymbol{\beta}\right)$$
$$= \Phi\left(\mathbf{x}_i\boldsymbol{\beta}\right)$$

$$Note: \mathbf{x}_i = \left[X_{2i}, \ldots, X_{Ki}\right]$$

- This is the standard equation for probit.
- This time we used an assumption about the distribution of the error term to derive it.
- If you assume a logistic distribution for the disturbance term, you get the logit model.

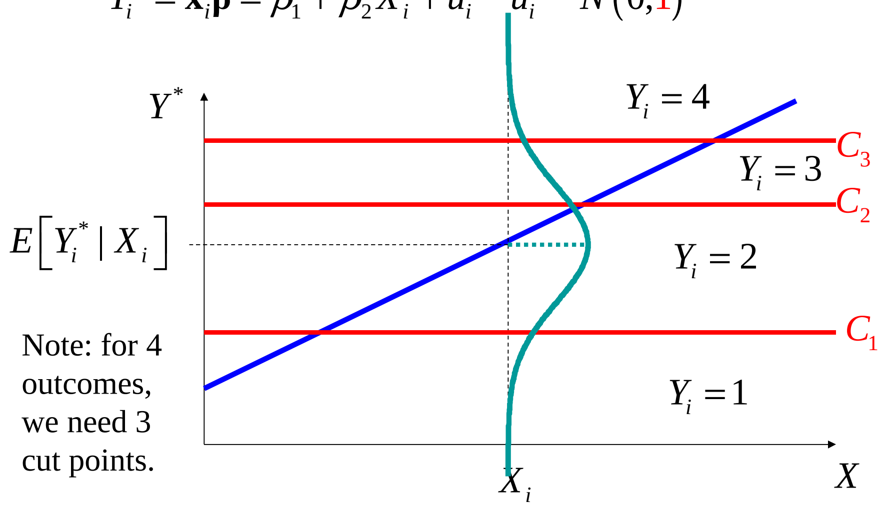
Ordinal Variables

- Definition?
- Examples?
  - Education (level attained vs. years)
  - Sears tool grades: good, better, best
- What about color? occupation?

$$Assume\ Y_i = \begin{cases} 1 & A & \text{Drop out} & \ldots \\ 2 & B & \text{HS Grad} & \ldots \\ 3 & C & \text{College Degree} & \ldots \\ 4 & D & \text{Graduate Degree} & \ldots \end{cases}$$

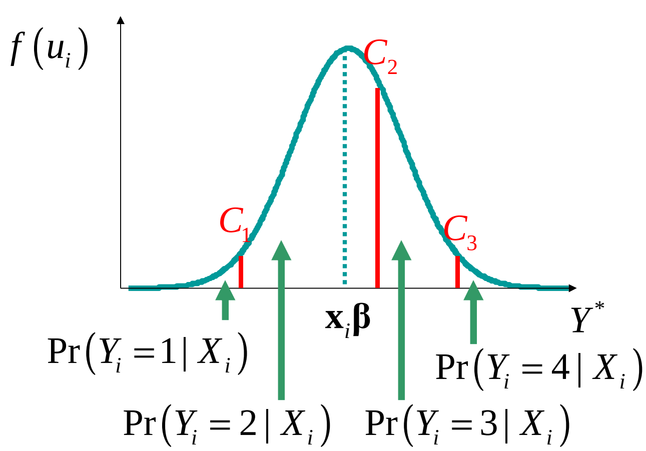
Latent variable approach

$$Y_i^* = \mathbf{x}_i\boldsymbol{\beta} = \beta_1 + \beta_2 X_i + u_i \quad u_i \sim N(0,1)$$



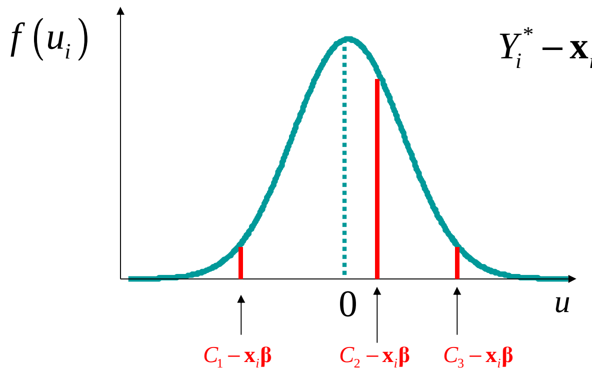
Note: for 4 outcomes, we need 3 cut points.

Probability Distribution of *u*



$$\textit{Define } P_{ji} = \Pr(Y_i = j \mid X_i) \qquad \sum_{j=1}^{4} P_{ji} = 1$$

Subtract $\mathbf{X}_i\mathbf{B}$

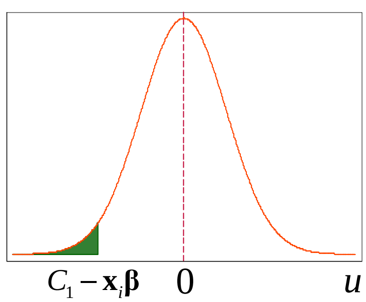


$$Y_i^* - \mathbf{x}_i\boldsymbol{\beta} = u_i$$

By assumption, all u_i follow a standard normal distribution.

Probability of outcome $Y_i = 1$

$$\begin{aligned} P_{1i} &= \Pr(Y_i = 1 \mid X_i) \\ &= \Pr(\mathbf{x}_i\boldsymbol{\beta} + u_i < C_1) \\ &= \Pr(u_i < C_1 - \mathbf{x}_i\boldsymbol{\beta}) \end{aligned}$$

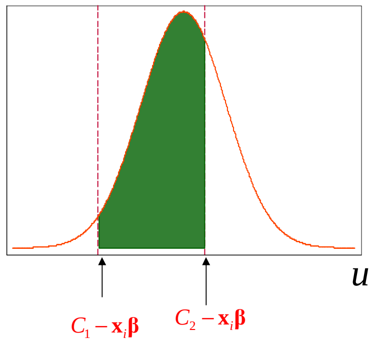


$$Since\ u_i \sim N(0,1)$$

$$\begin{aligned} \Pr(Y_i = 1 \mid X_i) &= \Pr(u_i < C_1 - \mathbf{x}_i\boldsymbol{\beta}) \\ &= \Phi(C_1 - \mathbf{x}_i\boldsymbol{\beta}) \end{aligned}$$

Probability of Y_i =2

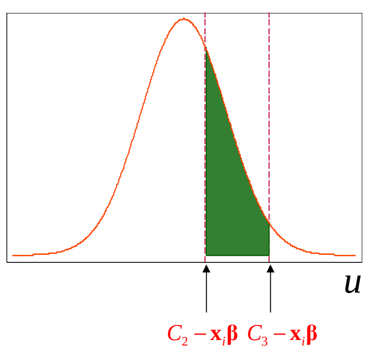
$$\Pr\left(Y_i = 2 \mid X_i\right) = ??$$



$$\Pr\left(Y_i = 2 \mid X_i\right) = \Pr\left[\left(C_1 - \mathbf{x}_i\boldsymbol{\beta}\right) < u_i < \left(C_2 - \mathbf{x}_i\boldsymbol{\beta}\right)\right]$$
$$= \Phi\left(C_2 - \mathbf{x}_i\boldsymbol{\beta}\right) - \Phi\left(C_1 - \mathbf{x}_i\boldsymbol{\beta}\right)$$

Probability of Y_i = 3

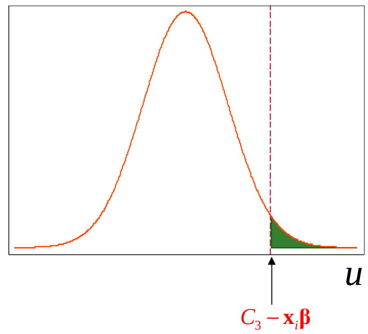
$$\Pr\left(Y_i = 3 \mid X_i\right) = ??$$



$$\Pr\left(Y_i = 3 \mid X_i\right) = \Pr\left[\left(C_2 - \mathbf{x}_i\boldsymbol{\beta}\right) < u_i < \left(C_3 - \mathbf{x}_i\boldsymbol{\beta}\right)\right]$$
$$= \Phi\left(C_3 - \mathbf{x}_i\boldsymbol{\beta}\right) - \Phi\left(C_2 - \mathbf{x}_i\boldsymbol{\beta}\right)$$

Probability of $Y_i = 4$

$$\Pr\left(Y_i = 4 \mid X_i\right) = ??$$



$$\Pr\left(Y_i = 4 \mid X_i\right) = \Pr\left[u_i > \left(C_3 - \mathbf{x}_i\boldsymbol{\beta}\right)\right]$$

$$= 1 - \Phi\left(C_3 - \mathbf{x}_i\boldsymbol{\beta}\right)$$

In general, for choices j=1,2,…,J

For $j = 1$:

$$P_{1i} = \Pr\left(Y_i = 1 \mid \mathbf{x}_i\right) = \Phi\left(C_1 - \mathbf{x}_i\boldsymbol{\beta}\right) \quad \textit{Note}: \mathbf{x}_i = \left[X_{2i}, \ldots, X_{Ki}\right]$$

For $j = 2$:

$$P_{2i} = \Pr\left(Y_i = 2 \mid \mathbf{x}_i\right) = \Phi\left(C_2 - \mathbf{x}_i\boldsymbol{\beta}\right) - \Phi\left(C_1 - \mathbf{x}_i\boldsymbol{\beta}\right)$$

For $j = 2, 3, \ldots, (J-1)$:

$$P_{ji} = \Pr\left(Y_i = j \mid \mathbf{x}_i\right) = \Phi\left(C_j - \mathbf{x}_i\boldsymbol{\beta}\right) - \Phi\left(C_{j-1} - \mathbf{x}_i\boldsymbol{\beta}\right)$$

For $j = J$:

$$P_{Ji} = \Pr\left(Y_i = J \mid \mathbf{x}_i\right) = 1 - \Phi\left(C_{J-1} - \mathbf{x}_i\boldsymbol{\beta}\right)$$

In general, for any $\mathbf{x}_i$

$$\text{Given } j = \{1, 2, \ldots, J\}$$

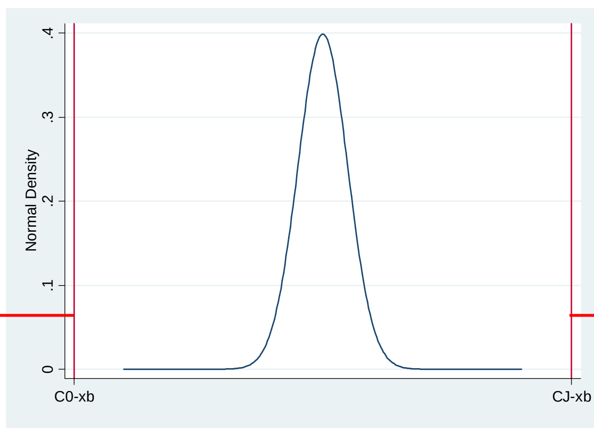
$$P_{ji} = \Pr(Y_i = j \mid \mathbf{x}_i)$$
$$= \Phi(C_j - \mathbf{x}_i\boldsymbol{\beta}) - \Phi(C_{j-1} - \mathbf{x}_i\boldsymbol{\beta})$$

To cover the extreme cases, we define:

$$C_0 = -\infty \qquad P_{1i} = \Phi(C_1 - \mathbf{x}_i\boldsymbol{\beta}) - \Phi(-\infty - \mathbf{x}_i\boldsymbol{\beta})$$
$$= \Phi(C_1 - \mathbf{x}_i\boldsymbol{\beta})$$

$$C_J = \infty \qquad P_{Ji} = \Phi(\infty - \mathbf{x}_i\boldsymbol{\beta}) - \Phi(C_{J-1} - \mathbf{x}_i\boldsymbol{\beta})$$
$$= 1 - \Phi(C_{J-1} - \mathbf{x}_i\boldsymbol{\beta})$$

Infinity is a Long Way



$-\infty$ $\quad$ ∞

$\Phi(-\infty) = 0$ $\quad$ $\Phi(\infty) = 1$

Estimation by MLE

$$Y_i = \{1, 2, ..., J\} \quad P_i = \Pr(Y_i \mid \mathbf{x}_i)$$

$$\mathcal{L} = \prod_{i=1}^{n}[P_i] = \prod_{i=1}^{n}\left[\Phi\left(C_{Y_i} - \mathbf{x}_i\boldsymbol{\beta}\right) - \Phi\left(C_{Y_i-1} - \mathbf{x}_i\boldsymbol{\beta}\right)\right]$$

$$\ln \mathcal{L} = \sum_{i=1}^{n}\ln\left[\Phi\left(C_{Y_i} - \mathbf{x}_i\boldsymbol{\beta}\right) - \Phi\left(C_{Y_i-1} - \mathbf{x}_i\boldsymbol{\beta}\right)\right]$$

$$\ln L = \sum_{i=1}^{n}\ln\left[\Phi\left(C_{Y_i} - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right) - \Phi\left(C_{Y_i-1} - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right)\right]$$

$Max(\ln L)$ with respect to: $\hat{\beta}_k, C_j$

We assume independence and estimate the model using ML.

Identification Problem

$$\begin{aligned}
\Phi\left(C_j - \mathbf{x}_i\boldsymbol{\beta}\right) &= \Phi\left(C_j - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - ...\right) \\
&= \Phi\left(\left(C_j + \delta\right) - \left(\hat{\beta}_1 + \delta\right) - \hat{\beta}_2 X_i - ...\right) \\
&= \Phi\left(C_j' - \hat{\beta}_1' - \hat{\beta}_2 X_{2i} - ...\right)
\end{aligned}$$

- In other words, if we shift the intercept and cutoffs by the same amount, the probabilities and likelihoods are unaffected.
- Must make an identifying assumption.
- Most common to assume either that β_1=0 or C_1=0.
- Stata assumes β_1=0 (no constant).
- Slopes are unaffected by the assumption.

Identification

- In other words, if we shift the intercept and cutoffs by the same amount, the probabilities and likelihoods are unaffected.
- Must make an identifying assumption.
- Most common to assume either that β_1=0 or C_1=0.
- Stata assumes β_1=0 (no constant).
- Slopes are unaffected by the assumption.

Example: Challenger Explosion

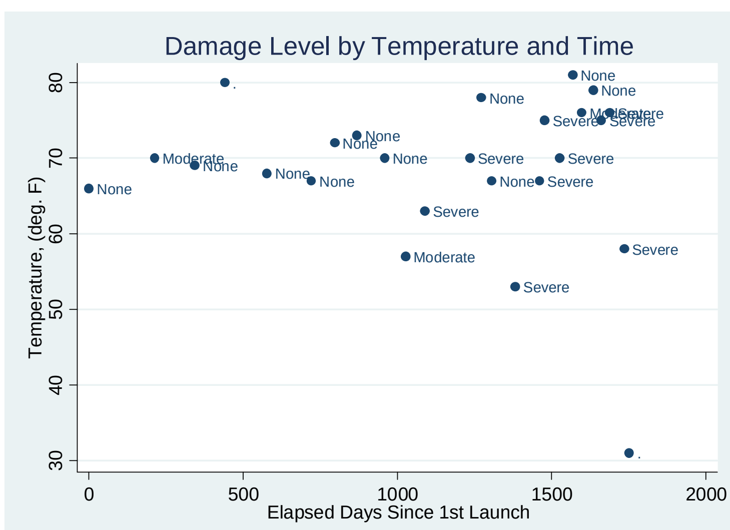
CNN Video of Challenger Disaster

shuttle.dta

name	date	time	temp	damage
1	12apr1981	0	66	None
2	12nov1981	214	70	Moderate
3	22mar1982	344	69	None
4	27jun1982	441	80	.
5	11nov1982	578	68	None
6	04apr1983	722	67	None
7	18jun1983	797	72	None
8	30aug1983	870	73	None
9	28nov1983	960	70	None
41B	03feb1984	1027	57	Moderate
41C	**06apr1984**	**1090**	**63**	**Severe**
41D	30aug1984	1236	70	Severe
41G	05oct1984	1272	78	None
51A	08nov1984	1306	67	None
51C	24jan1985	1383	53	Severe
51D	12apr1985	1461	67	Severe
51B	29apr1985	1478	75	Severe
51G	17jun1985	1527	70	Severe
51F	29jul1985	1569	81	None
51I	27aug1985	1598	76	Moderate
51J	03oct1985	1635	79	None
61A	30oct1985	1662	75	Severe
61B	26nov1985	1689	76	Severe
61C	12jan1986	1736	58	Severe
51L	28jan1986	1752	31	.

Source: Lawrence C. Hamilton, Statistics with Stata 3, p. 140 (citing Presidents Report on Challenger Accident 1986)

$$damage = \begin{cases} 1 & \text{No damage} \\ 2 & \text{Moderate} \\ 3 & \text{Severe} \end{cases}$$




```
scatter temp time, mlabel(damage) title("Damage Level by Temperature and Time")
```

Ordinal Probit

```
. oprobit damage time temp

Iteration 0:   log likelihood = -22.668661
Iteration 1:   log likelihood = -16.938046
Iteration 2:   log likelihood = -16.676629
Iteration 3:   log likelihood = -16.674263
Iteration 4:   log likelihood = -16.674263

Ordered probit regression                         Number of obs   =         23
                                                  LR chi2(2)      =      11.99
                                                  Prob > chi2     =     0.0025
Log likelihood = -16.674263                       Pseudo R2       =     0.2644

------------------------------------------------------------------------------
      damage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   .0019099   .0006703     2.85   0.004     .0005962    .0032236
        temp |  -.0987003    .047242    -2.09   0.037    -.1912929   -.0061077
-------------+----------------------------------------------------------------
       /cut1 |  -4.848873   3.132957                     -10.98936     1.29161
       /cut2 |   -4.32447   3.098891                     -10.39818    1.749245
------------------------------------------------------------------------------
```

Initial log likelihood

```
. tab damage
     Damage |
      Level |      Freq.     Percent        Cum.
------------+-----------------------------------
       None |         11       47.83       47.83
   Moderate |          3       13.04       60.87
     Severe |          9       39.13      100.00
------------+-----------------------------------
      Total |         23      100.00
```

$$\mathcal{L} = \prod_{i=1}^{n} P_i \qquad \ln \mathcal{L} = \sum_{i=1}^{n} \ln P_i = \sum_{Y_i=1} \ln P_1 + \sum_{Y_i=2} \ln P_2 + \sum_{Y_i=3} \ln P_3$$

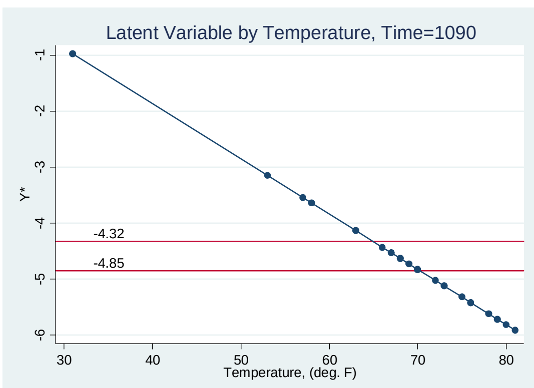
$$\ln L_0 = 11\left[\ln(0.4783)\right] + 3\left[\ln(0.1304)\right] + 9\left[\ln(0.3913)\right]$$
$$= -22.669$$

The List of Things

- *Different than* Binary Logit/Probit
  - A: Predicted Values
  - C: Marginal Effects
  - E: Odds (only in Binary/Ordinal *Logit*)
- *Same as* Binary Logit/Probit
  - B: Significance of Coefficients
  - D: Discrete Changes
  - F: Goodness of Fit
  - G: Hypothesis Testing

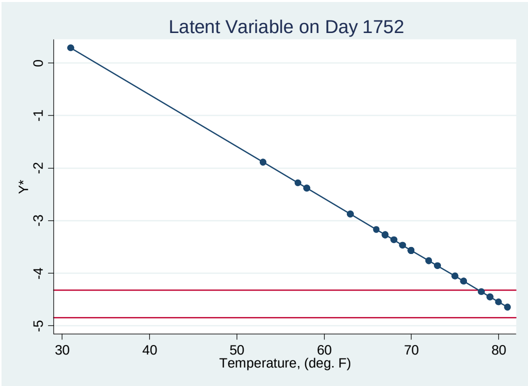
A: Predicted Values of *Y** for Mission 41C

$$\begin{aligned}\hat{Y}_i^* &= \mathbf{x}_i\hat{\boldsymbol{\beta}} = \hat{\beta}_2 time_i + \hat{\beta}_3 temp_i \\ &= 0.00191(1090) - 0.0987(63) \\ &= -4.14\end{aligned}$$



A: Predicted Values of *Y** for Challenger

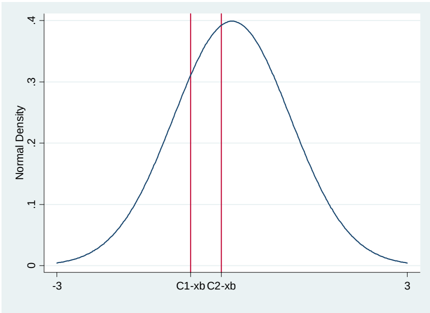
$$\hat{Y}_i^* = \mathbf{x}_i\hat{\boldsymbol{\beta}} = \hat{\beta}_2 time_i + \hat{\beta}_3 temp_i$$
$$= 0.00191(1752) - 0.0987(31)$$
$$= -0.27$$



A: Predicted Probabilities for Mission 41C

$$\mathbf{x}_i\hat{\boldsymbol{\beta}} = -4.14$$

$$\Pr(1) = \Phi(C_1 - \mathbf{x}_i\boldsymbol{\beta})$$



$$\Pr(2) = \Phi(C_2 - \mathbf{x}_i\boldsymbol{\beta}) - \Phi(C_1 - \mathbf{x}_i\boldsymbol{\beta})$$

$$\Pr(3) = 1 - \Phi(C_2 - \mathbf{x}_i\boldsymbol{\beta})$$

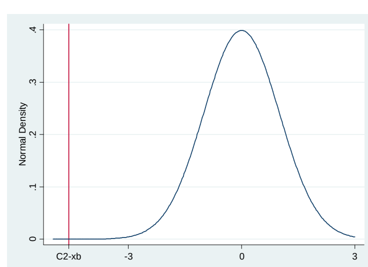
FYI, Code to Produce the Graphs

```
graph twoway function y = normalden(x),      ///
   xline(-0.71 -0.18) range(-3 3)            ///
   xscale(range(-3.1 3.1))                   ///
   xlabel(-3 -0.71 "C1-xb" -0.18 "C2-xb" 3) ///
   ytitle("Normal Density")                  ///
   xtitle(" ")                               ///
   text(0.05 -1.4  "0.239")                  ///
   text(0.05 -0.45 "0.19")                   ///
   text(0.05  0.75 "0.571")
```

A: Predicted Probability of Severe Damage at 31 Degrees and Day 1752

$$\begin{aligned}\hat{Y}_i^* &= \mathbf{x}_i\hat{\boldsymbol{\beta}} = \hat{\beta}_2 time_i + \hat{\beta}_3 temp_i \\ &= 0.00191(1752) - 0.0987(31) \\ &= 0.27\end{aligned}$$

$$\Pr(Y_i = 3) = 1 - \Phi\left(C_2 - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right)$$



A: Stata Tools

Built in:

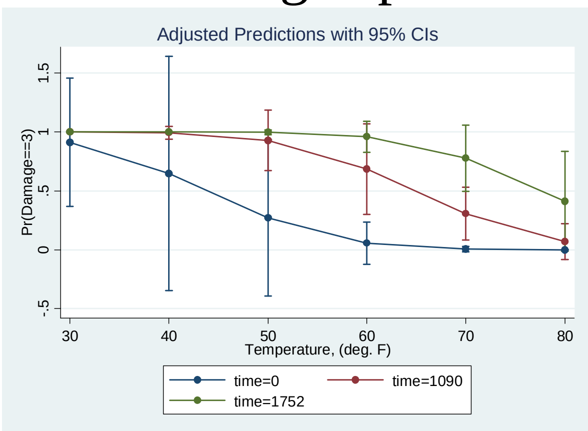
```
predict p1 p2 p3
predict xb, xb
margins
marginsplot
```

Predicted Values
Linear predictor

Long and Freese (spost13):

```
mchange
mchangeplot
mtable
```

Marginsplot

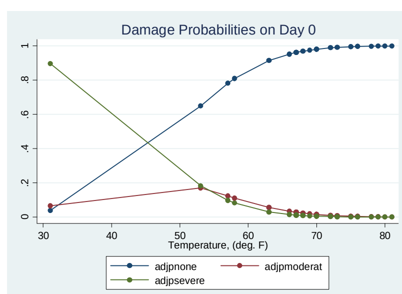


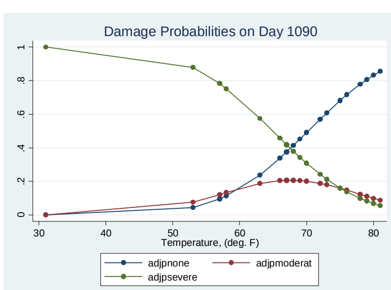

```
margins, at(temp=(30 40 50 60 70 80) ///
               time=(0 1090 1752)) ///
               predict(outcome(3))
marginsplot
```

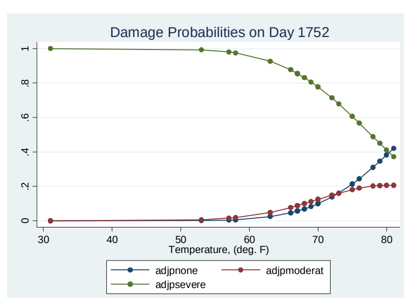
Long and Freese: mtable

```
. mtable, at(temp=(30(10)80) time=(0 1090 1752))

Expression: Pr(damage), predict(outcome())
          |      time       temp       None  Moderate    Severe
----------+-------------------------------------------------------
        1 |         0         30      0.030     0.057     0.914
        2 |         0         40      0.184     0.169     0.647
        3 |         0         50      0.534     0.195     0.271
        4 |         0         60      0.858     0.087     0.055
        5 |         0         70      0.980     0.015     0.005
        6 |         0         80      0.999     0.001     0.000

        7 |      1090         30      0.000     0.000     1.000
        8 |      1090         40      0.001     0.006     0.993
        9 |      1090         50      0.023     0.048     0.929
       10 |      1090         60      0.157     0.158     0.686
       11 |      1090         70      0.491     0.201     0.308
       12 |      1090         80      0.833     0.099     0.068

       13 |      1752         30      0.000     0.000     1.000
       14 |      1752         40      0.000     0.000     1.000
       15 |      1752         50      0.001     0.003     0.997
       16 |      1752         60      0.012     0.029     0.960
       17 |      1752         70      0.099     0.124     0.777
       18 |      1752         80      0.382     0.207     0.411
```

```
foreach t in 0 1090 1752 {
 capture drop adj*
 * linear part, holding time constant
 gen adjxb = _b[time]*`t'+_b[temp]*temp
 * adjusted probabilities
 gen adjpnone = normal(_b[/cut1]-adjxb)
 gen adjpmoderat = normal(_b[/cut2]-adjxb) ///
    -normal(_b[/cut1]-adjxb)
 gen adjpsevere = 1 - normal(_b[/cut2]-adjxb)
 * graph probabilities
 graph twoway scatter adjp* temp, ///
   sort connect(l l l) ///
   title("Damage Probabilities on Day `t'") ///
   name(day`t', replace)
   }
```

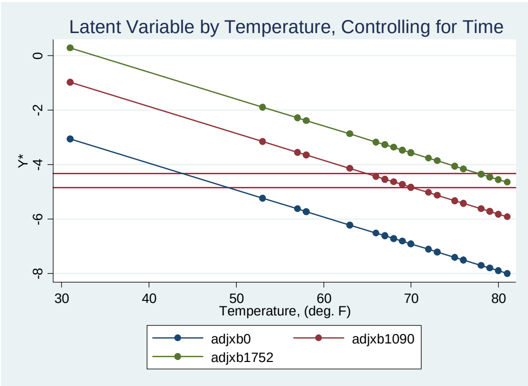






C: Marginal Effects (on Y^*)

$$\hat{Y}_i^* = \mathbf{x}_i\hat{\boldsymbol{\beta}} = \hat{\beta}_2 time_i + \hat{\beta}_3 temp_i$$
$$= 0.00191(time_i) - 0.0987(temp_i)$$



The latent variable is continuous and linear. Effects on the latent variable are easy to calculate.

The underlying latent variable is related to temperature. The actual outcome depends and the latent variable and the disturbance term. The whole relationship rises over time. Eventually, the program hit a fourth category: catastrophic damage.

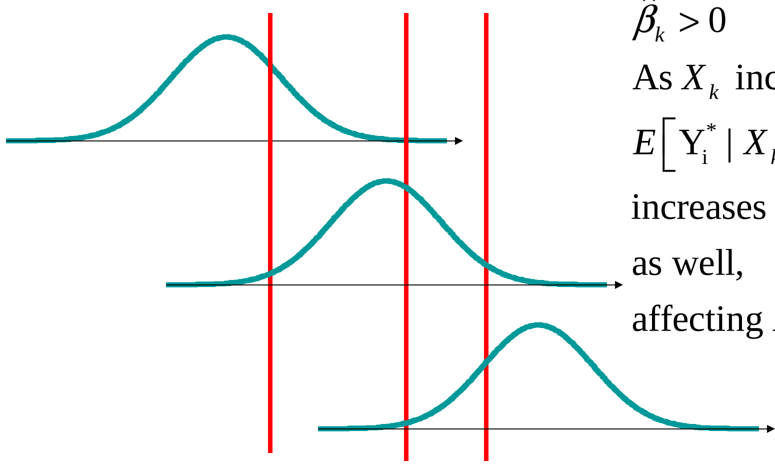
C: Marginal Effects (Probabilities)

$$\hat{P}_j = \Phi\left(C_j - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right) - \Phi\left(C_{j-1} - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right)$$

Effects on the probabilities are more important.

$$\frac{\partial \hat{P}_j}{\partial X_k} = \frac{\partial \Phi\left(C_j - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right)}{\partial X_k} - \frac{\partial \Phi\left(C_{j-1} - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right)}{\partial X_k}$$
$$= \phi\left(C_j - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right)\left(-\hat{\beta}_k\right) - \phi\left(C_{j-1} - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right)\left(-\hat{\beta}_k\right)$$
$$= \hat{\beta}_k\left[\phi\left(C_{j-1} - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right) - \phi\left(C_j - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right)\right]$$

C: Interpreting Effects



$\hat{\beta}_k > 0$

As X_k increases,

$E\left[\mathrm{Y}_{\mathrm{i}}^* \mid X_k \right]$

increases

as well,

affecting P_j.

In the intermediate categories, the marginal effect can change signs! P rises, then falls.

C: Example: Marginal Effects for Mission 41C

$$\mathbf{x}_i\hat{\boldsymbol{\beta}} = 0.00191(1090) - 0.0987(63) = -4.14$$

$$\boxed{\frac{\partial \hat{P}_j}{\partial X_k} = \hat{\beta}_k \left[\phi\left(C_{j-1} - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right) - \phi\left(C_j - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right) \right]}$$

$$\begin{aligned}\frac{\partial \hat{P}_1}{\partial temp} &= (-0.0987)\left[-\phi(-4.85 + 4.14)\right] \\ &= (0.0987)\phi(-0.71) \\ &= (0.0987)(0.31) \\ &= +0.0306\end{aligned}$$

How to calculate this?

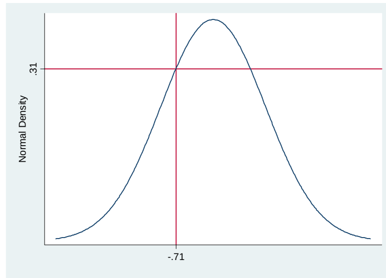
Standard Normal Density Function

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{1}{2}\right)z^2}$$

$$\phi(-0.71) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{1}{2}\right)\left(-0.71^2\right)}$$

$$= (0.4)e^{0.252}$$

$$= 0.31$$




```
. display normalden(-0.71)
.31006028

. display normalden(0.71)
.31006028
```

Not the same as capital phi (Φ). Note the symmetry.

$$\frac{\partial \hat{P}_2}{\partial temp} = (-0.0987)\left[\phi(-4.85+4.14) - \phi(-4.32+4.14)\right]$$

$$= (-0.0987)\left[\phi(-0.71) - \phi(-0.18)\right]$$

$$= (-0.0987)(0.31 - 0.39)$$

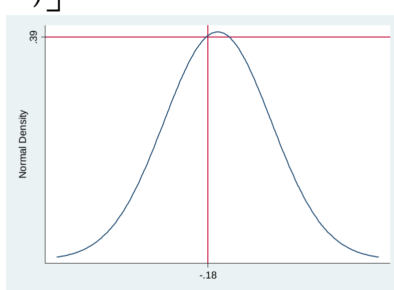
$$= 0.0079$$

$$\frac{\partial \hat{P}_3}{\partial temp} = (-0.0987)\left[\phi(-4.32+4.14)\right]$$

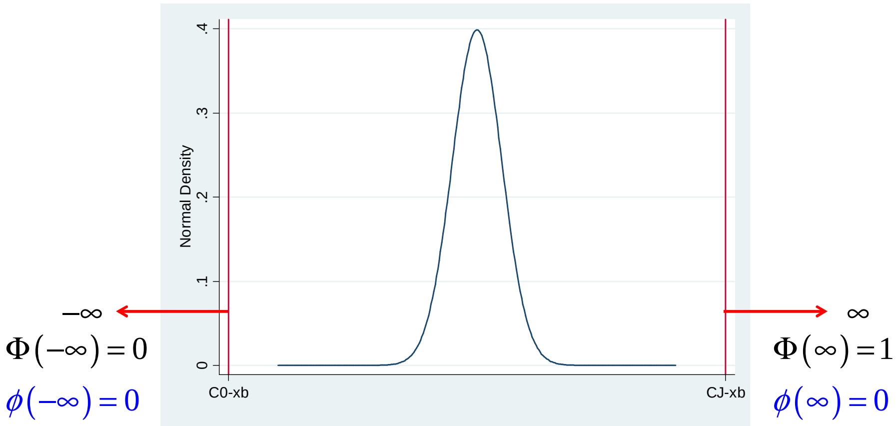
$$= (-0.0987)\left[\phi(-0.18)\right]$$

$$= (-0.0987)(0.39)$$

$$= -0.0385$$



Infinity is a Long Way Away



C: Marginal Effect *at the Means*

```
margins, predict(outcome(3)) dydx(time temp) atmeans
Conditional marginal effects                      Number of obs     =         23
Model VCE    : OIM

Expression   : Pr(damage==3), predict(outcome(3))
dy/dx w.r.t. : time temp
at           : time            =     1137.13 (mean)
               temp            =    69.56522 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   .0007116   .0002535     2.81   0.005     .0002147    .0012084
        temp |  -.0367727   .0182605    -2.01   0.044    -.0725627   -.0009827
------------------------------------------------------------------------------
```

Average Marginal Effects (Arguably better)

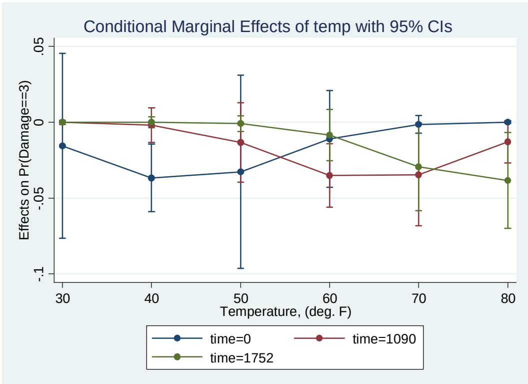
```
margins, predict(outcome(3)) dydx(time temp)

Average marginal effects                          Number of obs     =         23
Model VCE    : OIM

Expression   : Pr(damage==3), predict(outcome(3))
dy/dx w.r.t. : time temp

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   .0005114   .0001303     3.93   0.000     .0002561    .0007668
        temp |  -.0264291   .0094527    -2.80   0.005    -.0449561   -.0079022
------------------------------------------------------------------------------
```

At Specific Values

```
margins, dydx(temp) ///
                 at(temp=(30 40 50 60 70 80) ///
                 time=(0 1090 1752)) predict(outcome(3))

Conditional marginal effects                      Number of obs     =         23
Expression   : Pr(damage==3), predict(outcome(3))
dy/dx w.r.t. : temp

1._at        : time            =           0
               temp            =          30
2._at        : time            =           0
               temp            =          40
…
18._at       : time            =        1752
               temp            =          80
       |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
temp         |
         _at |
           1  |  -.0155434    .031077    -0.50   0.617    -.0764531    .0453663
           2  |  -.0366821   .0114117    -3.21   0.001    -.0590487   -.0143155
…
          18  |  -.0383878   .0161171    -2.38   0.017    -.0699768   -.0067988
------------------------------------------------------------------------------
```

```
. mchange

oprobit: Changes in Pr(y) | Number of obs = 23
Expression: Pr(damage), predict(outcome())
             |      None   Moderate     Severe
-------------+--------------------------------
time         |
          +1 |    -0.001      0.000      0.001
     p-value |     0.000      0.731      0.000
         +SD |    -0.240     -0.022      0.262
     p-value |     0.000      0.443      0.000
    Marginal |    -0.001      0.000      0.001
     p-value |     0.000      0.729      0.000
temp         |
          +1 |     0.027     -0.001     -0.026
     p-value |     0.013      0.684      0.004
         +SD |     0.190     -0.023     -0.167
     p-value |     0.004      0.414      0.001
    Marginal |     0.027     -0.001     -0.026
     p-value |     0.014      0.755      0.005

Average predictions
             |      None   Moderate     Severe
-------------+--------------------------------
  Pr(y|base) |     0.462      0.145      0.393
```

C: Output of marginsplot, dydx(temp)

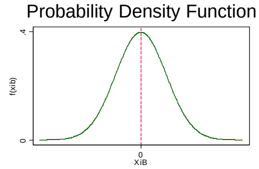



```
margins, dydx(temp) at(temp=(30 40 50 60 70 80) ///
        time=(0 1090 1752)) predict(outcome(3))
marginsplot
```

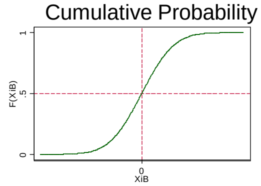
Logit vs. Probit

The logit function can be used in a similar way.

Probability Density Function

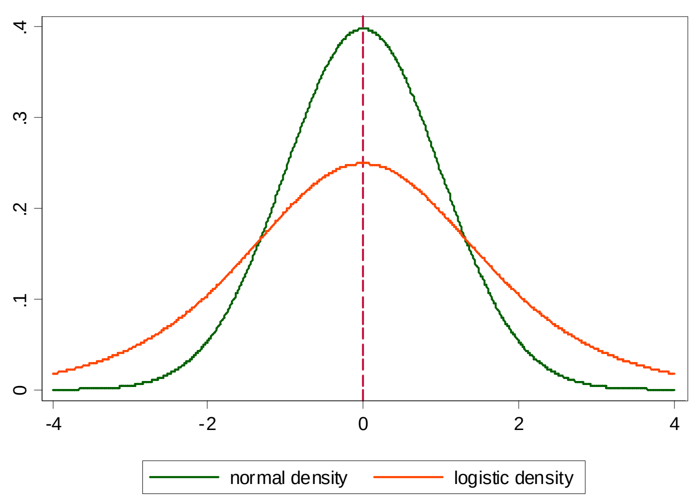


Cumulative Probability



Probit	Logit
$\phi(\mathbf{x}_i\boldsymbol{\beta}) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{1}{2}\right)(\mathbf{x}_i\boldsymbol{\beta})^2}$	$\lambda(\mathbf{x}_i\boldsymbol{\beta}) = \frac{e^{\mathbf{x}_i\boldsymbol{\beta}}}{\left(1+e^{\mathbf{x}_i\boldsymbol{\beta}}\right)^2}$
$P_i = \Phi(\mathbf{x}_i\boldsymbol{\beta}) = \textit{see table}$	$P_i = \Lambda(\mathbf{x}_i\boldsymbol{\beta}) = \frac{e^{\mathbf{x}_i\boldsymbol{\beta}}}{1+e^{\mathbf{x}_i\boldsymbol{\beta}}}$

Logistic vs. Normal Density Functions



Interesting Notes about Logit

$$P(1-P) = \Lambda(\mathbf{x}_i\boldsymbol{\beta})(1-\Lambda(\mathbf{x}_i\boldsymbol{\beta}))$$

$$= \left(\frac{e^{\mathbf{x}_i\boldsymbol{\beta}}}{1+e^{\mathbf{x}_i\boldsymbol{\beta}}}\right)\left(\frac{1}{1+e^{\mathbf{x}_i\boldsymbol{\beta}}}\right) = \frac{e^{\mathbf{x}_i\boldsymbol{\beta}}}{\left(1+e^{\mathbf{x}_i\boldsymbol{\beta}}\right)^2}$$

$$= \lambda(\mathbf{x}_i\boldsymbol{\beta})$$

So the logit marginal effects formula is actually comparable to the probit one.

$$\frac{\partial P_{ji}}{\partial X_k} = \phi(\mathbf{x}_i\boldsymbol{\beta})\beta_k$$

$$\frac{\partial P_{ji}}{\partial X_k} = \lambda(\mathbf{x}_i\boldsymbol{\beta})\beta_k$$

$$Var(normal) = 1,\ SD_{normal} = 1$$

$$Var(logit) = \frac{\pi^2}{3} = 3.29,\ SD_{logit} = 1.8$$

Ordinal Logit: Formulation

$$Given\ \ j = \{1, 2, ..., J\}$$

$$P_{ji} = \Pr(Y_i = j \mid X_i)$$

$$= \Lambda\left(C_j - \mathbf{x}_i\boldsymbol{\beta}\right) - \Lambda\left(C_{(j-1)} - \mathbf{x}_i\boldsymbol{\beta}\right)$$

$$\Lambda\left(C_j - \mathbf{x}_i\boldsymbol{\beta}\right) = \frac{e^{\left(C_j - \mathbf{x}_i\boldsymbol{\beta}\right)}}{1+e^{\left(C_j - \mathbf{x}_i\boldsymbol{\beta}\right)}} = \frac{1}{1+e^{-\left(C_j - \mathbf{x}_i\boldsymbol{\beta}\right)}}$$

$$\frac{\partial P_{ji}}{\partial X_k} = \hat{\beta}_k\left[\lambda\left(C_{(j-1)} - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right) - \lambda\left(C_j - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right)\right]$$

Ordinal Logit of Shuttle Problem

```
. ologit damage time temp
Iteration 0:   log likelihood = -22.668661
Iteration 1:   log likelihood = -17.142763
Iteration 2:   log likelihood = -16.723266
Iteration 3:   log likelihood = -16.700438
Iteration 4:   log likelihood = -16.700337

Ordered logistic regression                       Number of obs   =         23
                                                  LR chi2(2)      =      11.94
                                                  Prob > chi2     =     0.0026
Log likelihood = -16.700337                       Pseudo R2       =     0.2633
------------------------------------------------------------------------------
      damage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   .0035179   .0014664     2.40   0.016     .0006438     .006392
        temp |  -.1687129   .0855747    -1.97   0.049    -.3364363   -.0009896
-------------+----------------------------------------------------------------
       /cut1 |  -7.923562   5.448151                     -18.60174    2.754618
       /cut2 |  -7.031049   5.368454                     -17.55303    3.490928
------------------------------------------------------------------------------
```

A. Predicted Probabilities

$$\begin{aligned}\mathbf{x}_i\hat{\boldsymbol{\beta}} &= \hat{\beta}_2 time_i + \hat{\beta}_3 temp_i \\ &= 0.00352(1752) - 0.169(31) \\ &= 0.93\end{aligned}$$

$$\begin{aligned}\Pr(Y_i = 3) &= 1 - \Lambda\left(C_2 - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right) \\ &= 1 - \Lambda(-7.03 - 0.93) \\ &= 1 - \frac{1}{1 + e^{-(-7.96)}} \approx 1\end{aligned}$$

$\Pr(Y_i = 1) = ?\quad \Pr(Y_i = 2) = ?$

Cumulative Probabilities

$Y_i = \{1,2,3,4\}$ $\quad F = \Phi(\ )$ or $\Lambda(\ )$

$\Pr(Y_i = 4 \mid X_i) = 1 - F(C_3 - \mathbf{x}_i\boldsymbol{\beta})$

$\Pr(Y_i = 3 \mid X_i) = F(C_3 - \mathbf{x}_i\boldsymbol{\beta}) - F(C_2 - \mathbf{x}_i\boldsymbol{\beta})$

$\Pr(Y_i = 2 \mid X_i) = F(C_2 - \mathbf{x}_i\boldsymbol{\beta}) - F(C_1 - \mathbf{x}_i\boldsymbol{\beta})$

$\Pr(Y_i = 1 \mid X_i) = F(C_1 - \mathbf{x}_i\boldsymbol{\beta})$

$$\Pr(Y_i \le 4 \mid X_i) = 1$$
$$\Pr(Y_i \ge 3 \mid X_i) = F(C_3 - \mathbf{x}_i\boldsymbol{\beta})$$
$$\Pr(Y_i \le 2 \mid X_i) = F(C_2 - \mathbf{x}_i\boldsymbol{\beta})$$
$$\Pr(Y_i \le 1 \mid X_i) = F(C_1 - \mathbf{x}_i\boldsymbol{\beta})$$

These regressions have the same slope but different intercepts. Hence, they are known as parallel logits or probits. This is an implicit assumption of these ordinal models.

E. Ordinal Logit Only: Odds

$$\Omega_{Y_i \le j} = \frac{\Pr(Y_i \le 3)}{1 - \Pr(Y_i \le 3)} = \frac{\Lambda(C_3 - \mathbf{x}_i\boldsymbol{\beta})}{1 - \Lambda(C_3 - \mathbf{x}_i\boldsymbol{\beta})}$$

let $C_3 - \mathbf{x}_i\boldsymbol{\beta} = \theta$

$$= \frac{\left(\frac{e^\theta}{1+e^\theta}\right)}{1 - \left(\frac{e^\theta}{1+e^\theta}\right)} = \frac{e^\theta}{1 + e^\theta - e^\theta} = e^\theta$$

So the odds of "up to category j" vs. "higher than j" is: $\Omega_j = e^{C_j - \mathbf{x}_i\boldsymbol{\beta}}$

The odds of "higher than j" vs. "up to category j" is one over that:

$$\Omega_{Y>j} = \frac{1}{\Omega_{Y \le j}} = \frac{1}{e^{C_j - \mathbf{x}_i\boldsymbol{\beta}}} = e^{-(C_j - \mathbf{x}_i\boldsymbol{\beta})}$$

Odds Ratios in Ordinal Logit:

$$C_j - \mathbf{x}_i\boldsymbol{\beta} = C_j - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki} - \ldots - \beta_K X_{Ki}$$

Now increase X_k by one unit $(\mathbf{x}_i')$:

$$C_j - \mathbf{x}_i'\boldsymbol{\beta} = C_j - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k\left(X_{ki} + 1\right) - \ldots - \beta_K X_{Ki}$$

$$= C_j - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki} - \beta_k - \ldots - \beta_K X$$

Odds of ***j* or less** vs. **more than *j*** :

$$\frac{\Omega'_{Y\le j}}{\Omega_{Y\le j}} = \frac{e^{C_j - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki} - \beta_k - \ldots - \beta_K X_{Ki}}}{e^{C_j - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki} - \ldots - \beta_K X_{Ki}}} = e^{-\beta_k}$$

The minus is confusing, so…

Odds of **more than *j*** vs. ***j* or less** :

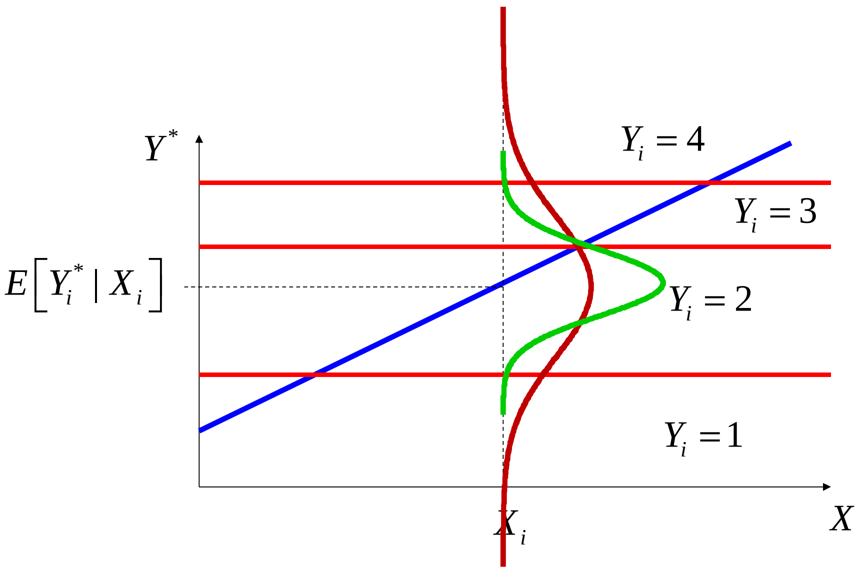
$$\frac{\Omega'_{Y> j}}{\Omega_{Y> j}} = \frac{e^{-\left(C_j - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki} - \beta_k - \ldots - \beta_K X_{Ki}\right)}}{e^{-\left(C_j - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki} - \ldots - \beta_K X_{Ki}\right)}} = e^{\beta_k}$$

That's better. And it doesn't matter which j.

And Don't Forget…

- *Same as* Binary Logit/Probit (and all MLE)
  - B: Significance of Coefficients
    - Asymptotically normal, z test for H_0: β_k=0
  - D: Discrete Changes: $\Delta P = P_2 - P_1$
    - 1 unit change in X_k
    - Dummy = 1 vs. 0
  - F: Goodness of Fit – Pseudo R^2
  - G: Hypothesis Testing
    - Test for the Model
    - Test for subset of Coefficient

What if two individuals had the same expected value but different variances?



Getting Wild: Heteroskedastic Ordered Probit

$$Y_i^* = \mathbf{x}_i\boldsymbol{\beta} + u_i \qquad u_i \sim N\left(0, \sigma_i^2\right)$$

$$\sigma_i = e^{\mathbf{z}\boldsymbol{\gamma}} \qquad \mathbf{z}\boldsymbol{\gamma} = \gamma_1 + \gamma_2 z_{2i} + \ldots + \gamma_k z_{Ki}$$

$$P_{ji} = \Phi\left(\frac{C_j - \mathbf{x}_i\boldsymbol{\beta}}{e^{\mathbf{z}\boldsymbol{\gamma}}}\right) - \Phi\left(\frac{C_{j-1} - \mathbf{x}_i\boldsymbol{\beta}}{e^{\mathbf{z}\boldsymbol{\gamma}}}\right)$$

$$\mathcal{L} = \ldots\ldots$$

See Alvarez and Brehm (1998). "Speaking in Two Voices: American Equivocation About the Internal Revenue Service," American Journal of Political Science 42: 418-452.