Analyses of ordered logit and probit models

Presentation · September 2016
DOI: 10.13140/RG.2.2.10458.13761

1 author:

Afees Salisu
University of Ibadan

Ordered Logit and Probit Models

Afees A. Salisu

Centre for Econometric & Allied Research

University of Ibadan

adebare1@yahoo.com

08034711769

Outline

- Introduction
- Ordered Logit Model
  - ➢ Specification
  - ➢ Estimation
  - ➢ Diagnostics
- Ordered Probit Model
  - ➢ Specification
  - ➢ Estimation
  - ➢ Diagnostics

Introduction

- Some discrete outcomes can be ordered to elicit more robust & representative information about the subject under consideration.
- Examples include:

➢ Rating systems (excellent, very good, good, fair, poor)

➢ For instance, self assessed English proficiency:

(i) excellent, (ii) very good, (iii) good,

(iv) fair, (v) poor

Intro…

- ➢ Do you agree with the following statement
    - ❖ Strongly agree, agree, disagree, strongly disagree
- For instance:
    - ❖ The Nigerian banking system is efficient in promoting SMEs.

    (i) strongly agree, (ii) agree, (iii) disagree, (iv) strongly disagree

Intro…

- Grades (A, B, C, D, E)
- Employment (unemployed, part time, full time)
- Economic/Social status (low, medium and high)
- Educational experience (elementary school graduate, high school graduate, College graduate).
- Bond ratings (AAA, AA, A, B, etc.)

Intro…

- Coding the responses:
- For instance, the options - Excellent, very good, good, fair, poor can be coded as 5,4,3,2,1
- They can also be coded as 1,2,3,4,5
- However, it is easier to explain the former than the latter in terms of highest to lowest value representing highest to lowest rating.
- Similarly, for the options - Strongly Agree, Agree, Disagree, Strongly Disagree options, we may also code as 5,4,3,2,1.

Intro…

- **Please note**:
- The numbers 1-5/ 5-1 mean nothing in terms of their value, just an ordering to show you the lowest to highest/ highest to lowest.
- Even though we can order these from lowest to highest, the spacing between the values may not be the same across the categories of the ordered variable.

Intro…

- In essence, these categories are not equally spaced.
- For instance, if we are modelling the predictors of differences in economic/social status at the household level in Nigeria, we can assign scores 1, 2, and 3 to the low, medium and high levels of economic and social status.

Intro…

- However, the difference between categories one and two (low and medium) is probably much bigger than the difference between categories two and three (medium and high).

Intro…

- In Statistics, variables described this way are classified as Ordinal variables/Ordered outcomes/ Polychotomous responses (as opposed to Dichotomous responses in the case of Binary outcomes).

Intro…

- Like the Binary case, when such a variable appears on the left-hand side [Dependent variable] of a statistical model, it is obvious that Least Squares regression will suffer from some short-comings such as heteroskedasticity, predicted probabilities lying outside the unit interval, etc.

Intro…

- Thus, the appropriate models for analysis in such a situation are the Logit and Probit models.
- The logit and probit models involving ordered outcomes are described as Ordered Logit and Ordered Probit models respectively.

Generalization

- Just like the binary choice models, the central idea behind the ordinal outcomes is that there is a latent continuous metric (defined as y^*) underlying the observed responses by the analyst.
- As previously explained, y^* is an unobserved variable, we only know when it crosses thresholds.

Generalization

- For instance, if we are modeling the predictors of bank performance; once y^* crosses a certain value we report poor, then good, then very good, then excellent performance.

Generalization

- Theoretical explanation:
- Let us consider a latent variable model given as:

$$y_i^* = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_k x_k + e$$

$$y_i^* = x_i' \alpha + e_i$$

$$y_i = j \quad \text{if} \quad u_{j-1} < y_i^* \le u_j$$

where $i = 1, \ldots, N$

Generalization

- The probability that observation i will select alternative j is:
- $p_{ij} = p(y_i = j) = p\left(u_{j-1} < y_i^* \ \leq u_j\right)$
  $= F\left(u_j - x_i'\alpha\right) - F\left(u_{j-1} - x_i'\alpha\right)$

For the Ordered logit, F is the logistic cdf

$$F(z) = e^z/(1 + e^z)$$

For the Ordered probit, F is the standard normal cdf.

Generalization

- An Illustration:
- Lets assume $y_i = (1,2,3,4,5)$ for (poor, fair, good, very good, excellent)
- Choice rule:

$$
\begin{aligned}
y_i &= 1 \text{ if } \quad y_i^* \le u_1 \\
y_i &= 2 \text{ if } u_1 < y_i^* \le u_2 \\
y_i &= 3 \text{ if } u_2 < y_i^* \le u_3 \\
y_i &= 4 \text{ if } u_3 < y_i^* \le u_4 \\
y_i &= 5 \text{ if } \quad y_i^* > u_4
\end{aligned}
$$

Generalization

- Using the generic representation, the respective probabilities for the five categories are derived as:
- Pr $(y_i = 1) = F(u_1 - x_i'\alpha)$
- Pr $(y_i = 2) = F(u_2 - x_i'\alpha) - F(u_1 - x_i'\alpha)$
- Pr $(y_i = 3) = F(u_3 - x_i'\alpha) - F(u_2 - x_i'\alpha)$
- Pr $(y_i = 4) = F(u_4 - x_i'\alpha) - F(u_3 - x_i'\alpha)$
- Pr $(y_i = 5) = 1 - F(u_4 - x_i'\alpha)$
- How did we come about these probabilities?

Generalization

- Proof:
- For Category 1:

$$\Pr\,(y_i = 1) = \Pr(y_i^* \leq u_1\,)$$

Recall: $y_i^* = x_i'\alpha + e_i$ and substitute into the above expression:

$$\Pr\,(y_i = 1) = \Pr(x_i'\alpha + e_i \leq u_1\,)$$

$$= \Pr(e_i \leq u_1\, - x_i'\alpha)$$

$$= F(u_1 - x_i'\alpha)$$

Generalization

- For Category 2:

$$\begin{aligned}\Pr(y_i = 2) &= \Pr(u_1 < \ y_i^* \leq u_2) \\ &= \Pr(u_1 < x_i'\alpha + e_i \leq u_2) \\ &= \Pr(u_1 - x_i'\alpha < e_i \leq u_2 \ - x_i'\alpha) \\ &= F(u_2 - x_i'\alpha) - F(u_1 - x_i'\alpha)\end{aligned}$$

Generalization

- For Category 3:

$$
\begin{aligned}
\Pr(y_i = 3) &= \Pr(u_2 < \ y_i^* \leq u_3) \\
&= \Pr(u_2 < x_i'\alpha + e_i \leq u_3) \\
&= \Pr(u_2 - x_i'\alpha < e_i \leq u_3 \ - x_i'\alpha) \\
&= F(u_3 - x_i'\alpha) - F(u_2 - x_i'\alpha)
\end{aligned}
$$

Generalization

- For Category 4:

$$\Pr(y_i = 4) = \Pr(u_3 < \ y_i^* \leq u_4)$$
$$= \Pr(u_3 < x_i'\alpha + e_i \leq u_4)$$
$$= \Pr(u_3 - x_i'\alpha < e_i \leq u_4 \ - x_i'\alpha)$$
$$= F(u_4 - x_i'\alpha) - F(u_3 - x_i'\alpha)$$

Generalization

- For Category 5:

$$\Pr(y_i = 5) = \Pr(y_i^* > u_4)$$
$$= \Pr(x_i'\alpha + e_i > u_4)$$
$$= \Pr(e_i > u_4 - x_i'\alpha)$$
$$= 1 - F(u_4 - x_i'\alpha)$$

Generalization

- As previously noted and also in line with probit and logit models, the $F(\cdot)$ is determined by the assumed distribution of e_i.

Generalization

- Things to note when estimating ordered logit/probit models:
  - The ordered ordered logit/probit model with j alternatives will have one set of coefficients with $(j-1)$ intercepts. You can recognize an ordered choice model by the multiple intercepts.
  - The ordered logit/probit model with j alternatives will have j sets of marginal effects.

Generalization

- Interpretation of coefficients (α_k):
  - The sign of α_k shows whether the latent variable y^*increases with the regressor (x_k).
  - The magnitude of α_k will be different by a scale factor between the probit and logit models.

Generalization

- Computation of Marginal effects for the ordered logit/probit models:

❑ All the regressors are continuous variables

- $\partial p_{ij}/\partial x_{ki} = \left\{F'\left(u_{j-1} - x_i'\alpha\right) - F'\left(u_j - x_i'\alpha\right)\right\}\alpha$
- Consider $y_i = 1$

$$\partial \text{Pr}\,(y_i = 1)/\partial x_i = \partial F(u_1 - x_i'\alpha)/\partial x_i$$
$$= F'\left(u_{j-1} - x_i'\alpha\right)\alpha$$

Generalization

- Consider y_i =2
- $\partial \Pr\,(y_i = 1)/\partial x_i = \partial F(u_2 - x_i'\alpha) - F(u_1 - x_i'\alpha)/\partial x_i$

$$= \{F'(u_1 - x_i'\alpha) - F'(u_2 - x_i'\alpha)\}\ \alpha$$

- It is the same procedure for the remaining outcomes.

Generalization

- ❑ Marginal Effects for Discrete X's

The procedure is similar to binary logit/probit models.

Let us assume that:

$$x_i\alpha = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \cdots + \alpha_k x_{ki}$$

If x_{2i} is binary in nature (1 or 0)

Generalization

Using Outcome (1) to illustrate, the Marginal Effect is computed thus:

$\partial \Pr(y_i = 1)/\partial x_{2i}$ = Change in the probabilities when x_{2i}=1 and x_{2i}=0

$$= F(\alpha_0 + \alpha_1 x_{1i} + \alpha_2 . 1 + \alpha_3 x_{3i} + \cdots + \alpha_k x_{ki}) - F(\alpha_0 + \alpha_1 x_{1i} + 0 + \alpha_3 x_{3i} + \cdots + \alpha_k x_{ki})$$

Generalization

- ❑Marginal effects interpretation:
  - Each unit increase in the independent variable increases/decreases the probability of selecting alternative *j* by the marginal effect expressed as a percent.

Application

- Research Title:

  Modelling the predictors of health status

- Data:

  Data are from Deb and Trivedi (2002).

  Deb, P. and P. K. Trivedi. "The Structure of Demand for Health Care: Latent Class versus Two-part Models, Journal of Health Economics 21: 601-625, 2002.

Application

- Research Objective:

The main objective of this study is to examine the factors influencing the health status.

- Model

$$y_{ij} = f(Age_i, Income_i, Diseases_i)$$

$$i = 1,2,3, \dots, 5574; \;\; j = 1,2,3$$

i is for observations *and*

j is for alternative outcomes

Application

- 1= fair
- 2= good
- 3= excellent
- N=5574

❑Estimator:

- Maximum Likelihood Estimator (MLE)

❑ Statistical sofware: Stata Software

Application

- Hints on results:
  - There will be 1 set of coefficients with two intercepts.
  - There will be 3 sets of marginal effects, one for each category.

Application

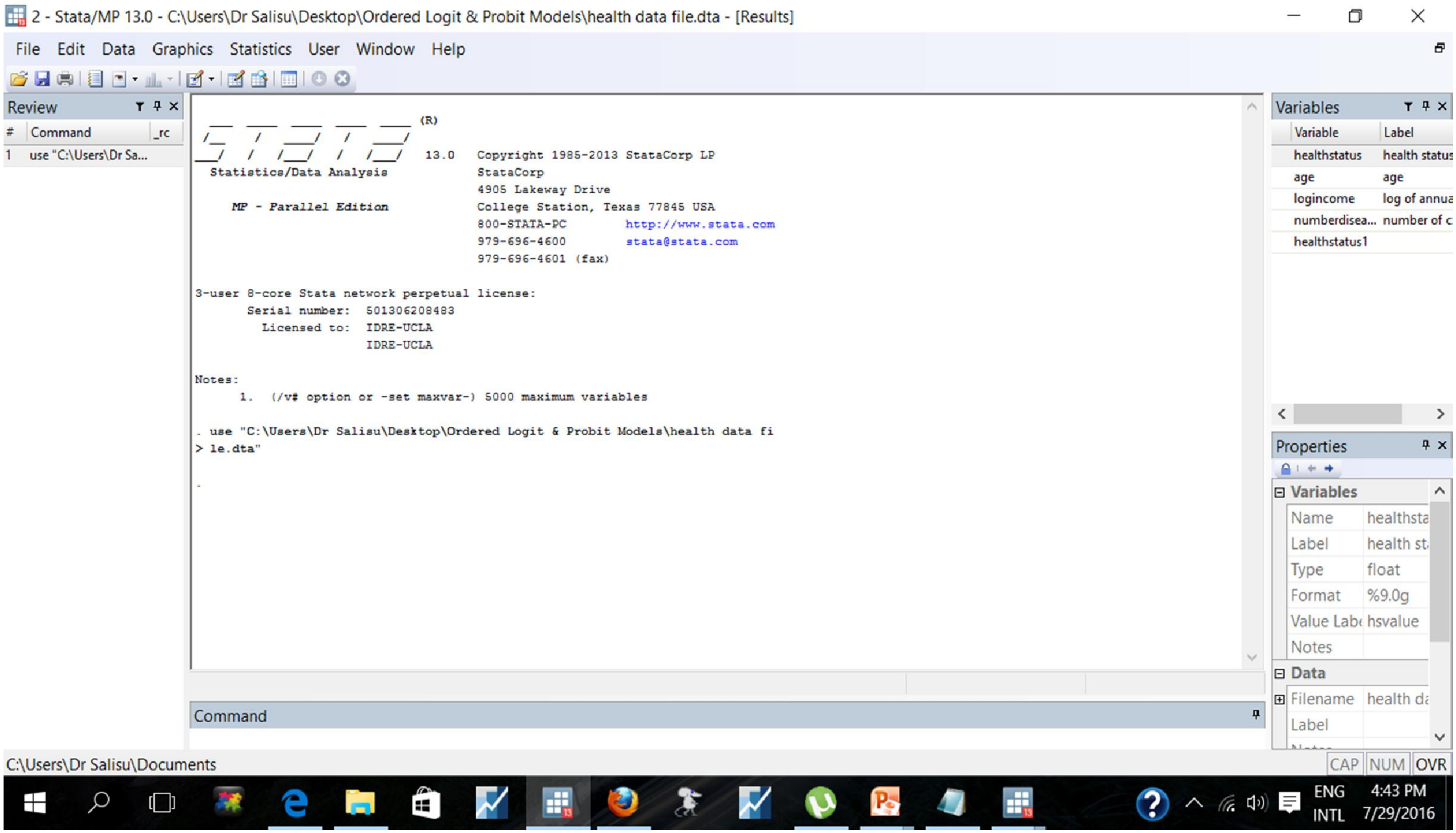
- You may consider the following steps when modelling with Ordinal dependent variables.

- Step 1: Load your data into Stata window

Since we already have our data in .dta format,

you may use stata command or menu approach to load the required data.

Application



Application

- Step 2:Compute relevant descriptive statistics for all the variables in your model.

(a) Compute summary statistics for continuous variables

```
. su age logincome numberdiseases
```

Variable	Obs	Mean	Std. Dev.	Min	Max
age	5574	25.57613	16.73011	.0253251	63.27515
logincome	5574	8.696929	1.220592	0	10.28324
numberdise~s	5574	11.20526	6.788959	0	58.6

- Some highlights of the descriptive statistics:
  - The average age of the respondents covered in the survey is about 25.
  - While their income levels average 8.69; they reported an average number of 11 diseases.
  - We find that the variation in income and no. of diseases is minimal while that of age is relatively high.

Application

(b1) You are to tabulate discrete variables and not to summarize. Why?

```
. tab healthstatus
```

health status (fair, good, excellent)	Freq.	Percent	Cum.
fair	523	9.38	9.38
good	2,034	36.49	45.87
excellent	3,017	54.13	100.00
Total	5,574	100.00	

Application

(b2) Still on discrete variables

```
. tab healthstatus1
```

healthstatus1	Freq.	Percent	Cum.
1	523	9.38	9.38
2	2,034	36.49	45.87
3	3,017	54.13	100.00
Total	5,574	100.00	

- Some highlights of the statistics:
  - About 9.4% (equivalent to 523), 36.5% (equivalent to 2034) and 54% (equivalent to 3017) of the respondents report fair, good and excellent health status respectively.
  - In other words, more than half of the respondents report excellent health status.

Application

- You may one to probe further to determine the distribution of age, income and number of diseases across the three categories.

Case I: Category 1

```
. su age logincome numberdiseases if healthstatus==1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
age	523	34.73173	17.91884	1.336756	62.9384
logincome	523	8.02038	2.059494	0	10.09323
numberdise~s	523	15.06425	9.675768	0	58.6

Case II: Category 2

```
. su age logincome numberdiseases if healthstatus==2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
age	2034	28.94548	16.47197	1.156057	63.02327
logincome	2034	8.706175	1.125837	0	10.28324
numberdise~s	2034	12.10255	7.108902	0	44.8

Case III: Category 3

```
. su age logincome numberdiseases if healthstatus==3
```

Variable	Obs	Mean	Std. Dev.	Min	Max
age	3017	21.71745	15.54491	.0253251	63.27515
logincome	3017	8.807976	1.035706	0	10.28324
numberdise~s	3017	9.931373	5.490731	0	41.4

- Highlights of the breakdown:
  - People with excellent health status have the lowest mean values of incidence of diseases and age but highest mean value of income
  - The reverse is the case for fair health status while good health status is in-between.
  - Thus, heath status seems to improve with lower incidence of diseases, higher income and lower age.

Step 3: Estimation
(a) 1. Estimation of the Ordered Logit Model

```
. ologit healthstatus age logincome numberdiseases

Iteration 0:   log likelihood = -5140.0463
Iteration 1:   log likelihood =  -4776.008
Iteration 2:   log likelihood = -4769.8693
Iteration 3:   log likelihood = -4769.8525
Iteration 4:   log likelihood = -4769.8525

Ordered logistic regression                     Number of obs   =       5574
                                                LR chi2(3)      =     740.39
                                                Prob > chi2     =     0.0000
Log likelihood = -4769.8525                     Pseudo R2       =     0.0720
```

healthstatus	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	-.0292944	.001681	-17.43	0.000	-.0325891	-.0259996
logincome	.2836537	.0231098	12.27	0.000	.2383593	.3289481
numberdiseases	-.0549905	.0040692	-13.51	0.000	-.0629661	-.047015
/cut1	-1.39598	.2061301			-1.799987	-.9919722
/cut2	.9513097	.2054301			.5486741	1.353945

- Highlights of the results:
  - Note that the results are log odds.
  - Each additional year of age decreases the log odds of reporting better health status (from fair, good to excellent) by 0.029 points holding other variables constant.
  - Also, the log odds in favour of an improved health status will increase by 0.283 points with a unit increase in income after controlling for age and no. of diseases.

- Similarly, a unit increase in the incidence of diseases reduces the log odds by 0.055 points after controlling for age and income.

- In sum, the health status is better (from fair to good to excellent) with lower age, higher income and lower number of diseases.

- You will find the cut-off points below the regression coefficients.
- They are statistically different from each other so the three categories should not be combined into one.
- We can as well compute the Odds ratio. See the next slide.

2. Computation of Odds Ratio

```
. ologit healthstatus age logincome numberdiseases, or

Iteration 0:   log likelihood = -5140.0463
Iteration 1:   log likelihood =  -4776.008
Iteration 2:   log likelihood = -4769.8693
Iteration 3:   log likelihood = -4769.8525
Iteration 4:   log likelihood = -4769.8525

Ordered logistic regression                       Number of obs   =       5574
                                                  LR chi2(3)      =     740.39
                                                  Prob > chi2     =     0.0000
Log likelihood = -4769.8525                       Pseudo R2       =     0.0720
```

healthstatus	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.9711306	.0016325	-17.43	0.000	.9679362	.9743355
logincome	1.327973	.0306892	12.27	0.000	1.269165	1.389506
numberdiseases	.9464941	.0038515	-13.51	0.000	.9389753	.9540731
/cut1	-1.39598	.2061301			-1.799987	-.9919722
/cut2	.9513097	.2054301			.5486741	1.353945

- Interpretation:
  - Note that the results are odds ratios.
  - The possibility of reporting better health status (from fair to good to excellent) increases by 33% [(1.33 -1)*100] with a unit increase in income while controlling for the other variables.
  - Also, the option of reporting better health status (from fair, good to excellent) deteriorates by 2.9% [(0.971 -1)*100] with each additional year of age holding other variables constant.

- In the same vein, for each additional unit increase in the no. of diseases, the likelihood of reporting improved health status declines by 5.4% [(0.946 -1)*100] after controlling for age and no. of diseases.

- We are done with the coefficients and odds ratios. Let us now determine the slope (marginal effect) for each of the regressors across the three categories.
- In fact, it is more convenient to interpret the marginal effects than the coefficients (log odds) and odds ratios.

(b) Marginal Effects for the Ordered Logit (Fair Health Status)

```
. margins, dydx(*) atmeans predict(outcome(1))

Conditional marginal effects                      Number of obs   =       5574
Model VCE    : OIM

Expression   : Pr(healthstatus==1), predict(outcome(1))
dy/dx w.r.t. : age logincome numberdiseases
at           : age             =    25.57613 (mean)
               logincome       =    8.696929 (mean)
               numberdise~s    =    11.20526 (mean)
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.002058	.0001333	15.44	0.000	.0017969	.0023192
logincome	-.0199278	.0017344	-11.49	0.000	-.0233272	-.0165284
numberdiseases	.0038633	.0003056	12.64	0.000	.0032643	.0044623

- Some highlights of the results.
- Each additional year of age increases the probability of reporting fair by 0.2%.
- Each additional unit increase in the no. of diseases increases the probability of reporting fair by 0.4%.
- However, for income, the probability of reporting fair declines by 2% for every additional unit increase in income.

(c) Marginal Effects for the Ordered Logit (Good Health Status)

```
. margins, dydx(*) atmeans predict(outcome(2))

Conditional marginal effects                      Number of obs   =       5574
Model VCE    : OIM

Expression   : Pr(healthstatus==2), predict(outcome(2))
dy/dx w.r.t. : age logincome numberdiseases
at           : age             =    25.57613 (mean)
               logincome       =    8.696929 (mean)
               numberdise~s    =    11.20526 (mean)
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.0052244	.0003258	16.04	0.000	.0045859	.0058629
logincome	-.0505872	.0043054	-11.75	0.000	-.0590256	-.0421489
numberdiseases	.0098071	.000768	12.77	0.000	.0083018	.0113124

- Some highlights of the results.
- Note that the results for category 2 are quite similar to category 1 particularly in terms of the signs and statistical significance.
- Thus, the coefficients are interpreted the same way.

- (d) Marginal Effects for the Ordered Logit (Excellent Health Status)

```
. margins, dydx(*) atmeans predict(outcome(3))

Conditional marginal effects                      Number of obs   =       5574
Model VCE    : OIM

Expression   : Pr(healthstatus==3), predict(outcome(3))
dy/dx w.r.t. : age logincome numberdiseases
at           : age             =    25.57613 (mean)
               logincome       =    8.696929 (mean)
               numberdise~s    =    11.20526 (mean)
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	-.0072824	.0004179	-17.43	0.000	-.0081014	-.0064634
logincome	.070515	.0057527	12.26	0.000	.05924	.0817901
numberdiseases	-.0136704	.0010126	-13.50	0.000	-.015655	-.0116858

- Some highlights of the results.
- The results under category 3 (excellent health status) differ from categories 1 & 2 particularly in terms of the signs.
- Each additional year of age reduces the chance of reporting excellent by 0.7%.
- A unit decrease in the no. of diseases increases the probability of reporting excellent by 1.4%.
- While for every additional unit increase in income, the probability of reporting excellent increases by 7.1%.

(e) Computation of predicted probabilities

(1) Predicted probabilities for the three categories for the ordered logit model

```
. predict p1ologit p2ologit p3ologit, pr

. su p1ologit p2ologit p3ologit
```

Variable	Obs	Mean	Std. Dev.	Min	Max
p1ologit	5574	.0946903	.0843148	.0233629	.859022
p2ologit	5574	.3651672	.0946158	.1255265	.5276064
p3ologit	5574	.5401425	.1640575	.0154515	.7999009

- Some highlights of the results
- Given the mean values of the regressors, the average probability values of reporting fair, good and excellent are 9%, 37% and 54% respectively.
- Let us compare the predicted with the actual.

(2) Actual percentage distribution

```
. tab healthstatus
```

health status (fair, good, excellent)	Freq.	Percent	Cum.
fair	523	9.38	9.38
good	2,034	36.49	45.87
excellent	3,017	54.13	100.00
Total	5,574	100.00	

Observation: The model reasonably fits the data. The predicted probabilities are similar to the actual.

- (f) Estimation of Ordered Probit Model

```
. oprobit healthstatus age logincome numberdiseases

Iteration 0:   log likelihood = -5140.0463
Iteration 1:   log likelihood = -4771.7555
Iteration 2:   log likelihood = -4771.0299
Iteration 3:   log likelihood = -4771.0298

Ordered probit regression                         Number of obs   =       5574
                                                  LR chi2(3)      =     738.03
                                                  Prob > chi2     =     0.0000
Log likelihood = -4771.0298                       Pseudo R2       =     0.0718
```

healthstatus	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	-.0171681	.0009898	-17.34	0.000	-.0191082	-.0152281
logincome	.1654079	.01286	12.86	0.000	.1402028	.190613
numberdiseases	-.0315288	.0023848	-13.22	0.000	-.0362029	-.0268548
/cut1	-.7945455	.115108			-1.020153	-.5689379
/cut2	.5459371	.1148228			.3208886	.7709857

- Some highlights of the results.
  - Like the binary probit models, the results here are z-scores.
  - Thus, the interpretation here is not different from the binary probit models except that the ordering is reflected in this case.
  - For instance, using income, the result shows that a unit increase in income will lead to an increase in the z-score in favour of better health status by 0.165 points.

- Note that the z-scores are similar to the log odds in terms of sign and significance (income is positive while age and no. of diseases are negative).
- Therefore, the option of reporting better health status increases with a higher income, lower age and declining incidence of diseases.

(g) Marginal Effects for the Ordered Probit (Outcome (1) - Fair Health Status)

```
. margins, dydx(*) atmeans predict(outcome(1))

Conditional marginal effects                      Number of obs   =       5574
Model VCE    : OIM

Expression   : Pr(healthstatus==1), predict(outcome(1))
dy/dx w.r.t. : age logincome numberdiseases
at           : age             =    25.57613 (mean)
               logincome       =    8.696929 (mean)
               numberdise~s    =    11.20526 (mean)
```

	Delta-method dy/dx	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.0024261	.0001545	15.70	0.000	.0021233	.002729
logincome	-.0233749	.0019304	-12.11	0.000	-.0271584	-.0195914
numberdiseases	.0044556	.0003587	12.42	0.000	.0037525	.0051586

(h) Marginal Effects for the Ordered Probit (Good Health Status)

```
. margins, dydx(*) atmeans predict(outcome(2))

Conditional marginal effects                    Number of obs   =       5574
Model VCE    : OIM

Expression   : Pr(healthstatus==2), predict(outcome(2))
dy/dx w.r.t. : age logincome numberdiseases
at           : age             =    25.57613 (mean)
               logincome       =    8.696929 (mean)
               numberdise~s    =    11.20526 (mean)
```

	Delta-method dy/dx	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.0043886	.0002786	15.75	0.000	.0038426	.0049347
logincome	-.0422827	.0034702	-12.18	0.000	-.0490842	-.0354812
numberdiseases	.0080596	.0006459	12.48	0.000	.0067937	.0093256

(i) Marginal Effects for the Ordered Probit (Excellent Health Status)

```
. margins, dydx(*) atmeans predict(outcome(3))

Conditional marginal effects                      Number of obs   =       5574
Model VCE    : OIM

Expression   : Pr(healthstatus==3), predict(outcome(3))
dy/dx w.r.t. : age logincome numberdiseases
at           : age             =    25.57613 (mean)
               logincome       =    8.696929 (mean)
               numberdise~s    =    11.20526 (mean)
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	-.0068148	.0003929	-17.34	0.000	-.0075849	-.0060447
logincome	.0656577	.0051072	12.86	0.000	.0556476	.0756677
numberdiseases	-.0125152	.0009472	-13.21	0.000	-.0143717	-.0106587

- Note that the marginal effects for ordered probit models are quite similar to ordered logit models.
- Also, their slopes are interpreted the same way.
- Thus, the interpretation of the latter suffices for the former here.

(j) Computation of Predicted Probabilities

(1) Predicted probabilities for the three categories for the ordered logit model

```
. predict p1oprobit p2oprobit p3oprobit, pr
```

```
. summarize p1oprobit p2oprobit p3oprobit
```

Variable	Obs	Mean	Std. Dev.	Min	Max
p1oprobit	5574	.0941691	.0897707	.0155477	.8556623
p2oprobit	5574	.3641854	.0802058	.1361741	.497296
p3oprobit	5574	.5416455	.1570388	.0081637	.7925686

(2) Actual percentage distribution

```
. tab healthstatus
```

health status (fair, good, excellent)	Freq.	Percent	Cum.
fair	523	9.38	9.38
good	2,034	36.49	45.87
excellent	3,017	54.13	100.00
Total	5,574	100.00	

Observation: The model reasonably fits the data. The predicted probabilities are similar to the actual.

Step 4: Some scenario analyses

- Note that the mean values were used in the computation of the marginal effects for both ordered logit and probit models.
- However, one may be interested in computing marginal effects for specific values of one or more of the regressors.
- You will recall that we did something similar to that in binary logit and probit models.

Case I: at(age=26/30)

```
. margins, at(age=(26/30)) predict(outcome(3)) atmeans

Adjusted predictions                              Number of obs   =       5574
Model VCE    : OIM

Expression   : Pr(healthstatus==3), predict(outcome(3))

1._at        : age             =          26
               logincome       =    8.696929 (mean)
               numberdise~s    =    11.20526 (mean)

2._at        : age             =          27
               logincome       =    8.696929 (mean)
               numberdise~s    =    11.20526 (mean)

3._at        : age             =          28
               logincome       =    8.696929 (mean)
               numberdise~s    =    11.20526 (mean)

4._at        : age             =          29
               logincome       =    8.696929 (mean)
               numberdise~s    =    11.20526 (mean)

5._at        : age             =          30
               logincome       =    8.696929 (mean)
               numberdise~s    =    11.20526 (mean)
```

	Margin	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
_at						
1	.534388	.0070448	75.86	0.000	.5205805	.5481954
2	.5270922	.0070719	74.53	0.000	.5132316	.5409528
3	.5197848	.0071208	73.00	0.000	.5058284	.5337413
4	.512469	.0071911	71.26	0.000	.4983748	.5265632
5	.5051478	.0072821	69.37	0.000	.4908751	.5194205

- The higher the age, the lower the probability of reporting excellent health status.
- In other words, younger people are more likely to report excellent health than older people.

- **Case II: at(income=5/8)**

```
. margins, at( logincome=(5/8)) predict(outcome(3)) atmeans

Adjusted predictions                              Number of obs   =       5574
Model VCE    : OIM

Expression   : Pr(healthstatus==3), predict(outcome(3))

1._at        : age             =    25.57613 (mean)
               logincome       =           5
               numberdise~s    =    11.20526 (mean)

2._at        : age             =    25.57613 (mean)
               logincome       =           6
               numberdise~s    =    11.20526 (mean)

3._at        : age             =    25.57613 (mean)
               logincome       =           7
               numberdise~s    =    11.20526 (mean)

4._at        : age             =    25.57613 (mean)
               logincome       =           8
               numberdise~s    =    11.20526 (mean)
```

	Margin	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
_at						
1	.2893668	.0187194	15.46	0.000	.2526775	.3260561
2	.3509633	.0158177	22.19	0.000	.3199612	.3819654
3	.4179599	.0119799	34.89	0.000	.3944798	.44144
4	.4881272	.0082732	59.00	0.000	.471912	.5043423

- The higher the level of income, the higher the probability of reporting excellent health status.
- In other words, rich people are more likely to report excellent health than poor people.

• **Case III: at(no. of diseases=23 26 29)**

```
. margins, at( numberdiseases =(23 26 29)) predict(outcome(3)) atmeans

Adjusted predictions                              Number of obs   =       5574
Model VCE    : OIM

Expression   : Pr(healthstatus==3), predict(outcome(3))

1._at        : age             =    25.57613 (mean)
               logincome       =    8.696929 (mean)
               numberdise~s    =          23

2._at        : age             =    25.57613 (mean)
               logincome       =    8.696929 (mean)
               numberdise~s    =          26

3._at        : age             =    25.57613 (mean)
               logincome       =    8.696929 (mean)
               numberdise~s    =          29
```

	Margin	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
_at						
1	.3779138	.0132116	28.60	0.000	.3520196	.403808
2	.3399798	.0150399	22.61	0.000	.3105022	.3694574
3	.3039928	.0165583	18.36	0.000	.2715392	.3364464

- Get ready for class exercise

- Thank you