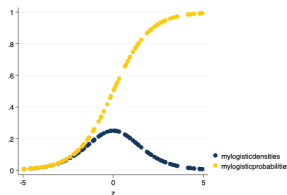# Logistic Regression

Andy Grogan-Kaylor

11 May 2020

Key Concepts and Commands

- Fitting a Curve to 2 Possible Values

- Linear models, probit and logit
- `y x1 x2 ...` $\leftarrow \rightarrow F(y) = \beta_0 + \beta x_1 + \beta x_2...$
- `regress y x1 x2` OLS; Linear Model
- `logit y x1 x2` Logistic Regression
- `probit y x1 x2` Probit Regression
- `glm ...`

Limited Dependent Variables

- Categorical Dependent Variable
- Binary Dependent Variable
- Limited Dependent Variable

General Social Survey

```
. use "/Users/agrogan/Box Sync/DATA WAREHOUSE/General Social Survey Panel Data/GSS_panel2010w12
> 3_R6 - stata.dta", clear
( )

. codebook happy_3 // what does this variable look like?
```

happy_3	happy_3: GENERAL HAPPINESS

```
                 type:  numeric (byte)
```

```
                    label:  HAPPY_3
                    range:  [1,3]                          units:  1
            unique values:  3                          missing .:  0/2,044
           unique mv codes:  3                         missing .*:  742/2,044

              tabulation:  Freq.   Numeric  Label
                            391        1    VERY HAPPY
                            758        2    PRETTY HAPPY
                            153        3    NOT TOO HAPPY
                              1        .d   DK
                            740        .i   IAP
                              1        .n   NA
```

Data Management

```
. recode happy_3 (1/2 = 1)(3=0), generate(happy_3_D)
(911 differences between happy_3 and happy_3_D)

. tabulate happy_3 happy_3_D // double check
```

happy_3: GENERAL HAPPINESS	RECODE of happy_3 (happy_3: GENERAL HAPPINESS)		Total
	0	1	
VERY HAPPY	0	391	391
PRETTY HAPPY	0	758	758
NOT TOO HAPPY	153	0	153
Total	153	1,149	1,302

```
. keep happy_3 happy_3_D coninc_3 // keep only some variables

. save GSSsmall.dta, replace
file GSSsmall.dta saved
```

Visualize

```
. twoway scatter happy_3_D coninc_3, scheme(burd) jitter(5)

. graph export happiness-income.png, width(500) replace
(file happiness-income.png written in PNG format)
```

Linear Probability Model

```
. regress happy_3_D coninc_3
```

Source	SS	df	MS		Number of obs	=	1,223
					F(1, 1221)	=	22.87
Model	2.26477708	1	2.26477708		Prob > F	=	0.0000
Residual	120.937185	1,221	.099047654		R-squared	=	0.0184
					Adj R-squared	=	0.0176
Total	123.201962	1,222	.100819936		Root MSE	=	.31472

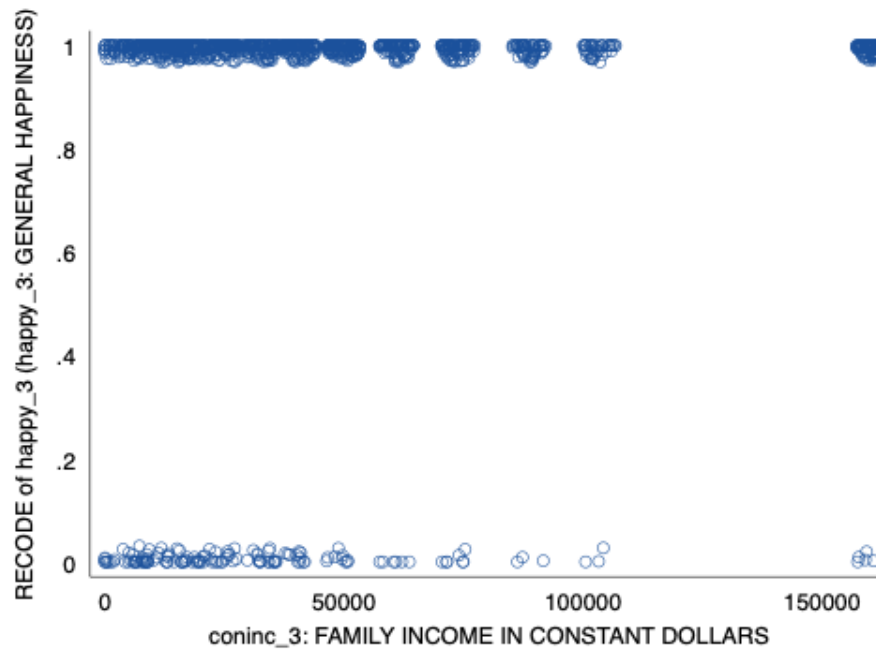| happy_3_D	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
coninc_3	9.69e-07	2.03e-07	4.78	0.000	5.72e-07	1.37e-06
_cons	.8368664	.0137133	61.03	0.000	.8099621	.8637706

Figure 1: Happiness and Income

Normal and Cumulative Normal Distribution

```
. clear all

. set obs 100 // 100 observations
number of observations (_N) was 0, now 100

. generate z = runiform(-5, 5) // randomly distributed z scores

. generate mynormaldensities = normalden(z) // normal densities

. generate myprobabilities = normal(z) // cumulative normal probabilities

. twoway scatter mynormaldensities myprobabilities z, scheme(michigan)

. graph export normal.png, width(500) replace
(file normal.png written in PNG format)
```

The Probit Model

```
. use GSSsmall.dta, clear
( )

. probit happy_3_D coninc_3

Iteration 0:   log likelihood = -433.05123
Iteration 1:   log likelihood = -419.92819
Iteration 2:   log likelihood = -419.73499
Iteration 3:   log likelihood = -419.73484
Iteration 4:   log likelihood = -419.73484

Probit regression                               Number of obs    =       1,223
                                                LR chi2(1)       =       26.63
                                                Prob > chi2      =      0.0000
```
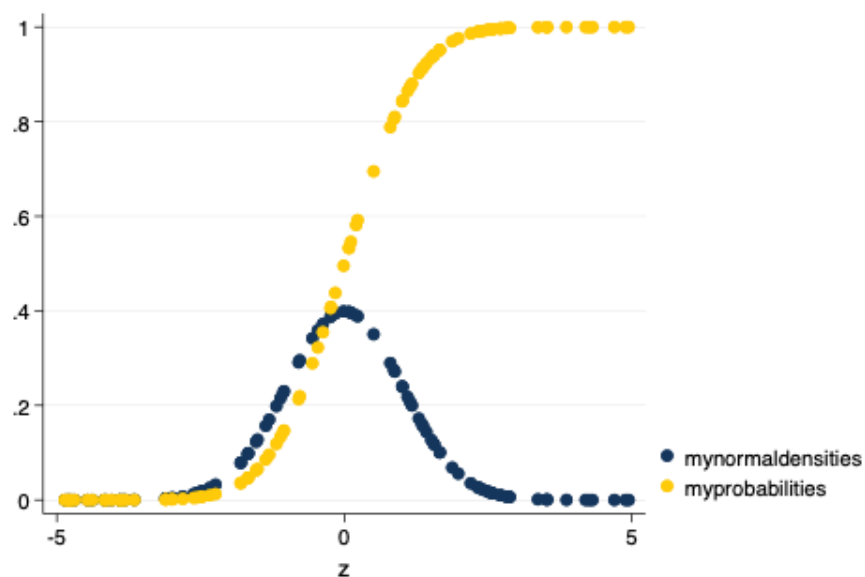
Figure 2: Standard and Cumulative Normal Curves

```
Log likelihood = -419.73484                    Pseudo R2       =      0.0308
```

happy_3_D	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]
coninc_3	6.43e-06	1.35e-06	4.76	0.000	3.78e-06	9.08e-06
_cons	.9244086	.0721521	12.81	0.000	.7829931	1.065824

The Logistic Distribution

```
. clear all

. set obs 100 // 100 observations
number of observations (_N) was 0, now 100

. generate z = runiform(-5, 5) // randomly distributed z scores

. generate mylogisticdensities = logisticden(z) // logistic densities

. generate mylogisticprobabilities = logistic(z) // cumulative logistic probabilities

. twoway scatter mylogisticdensities mylogisticprobabilities z, scheme(michigan)

. graph export logistic.png, width(500) replace
(file logistic.png written in PNG format)
```

The Logit (Logistic) Model

```
. use GSSsmall.dta, clear
( )

. logit happy_3_D coninc_3
```
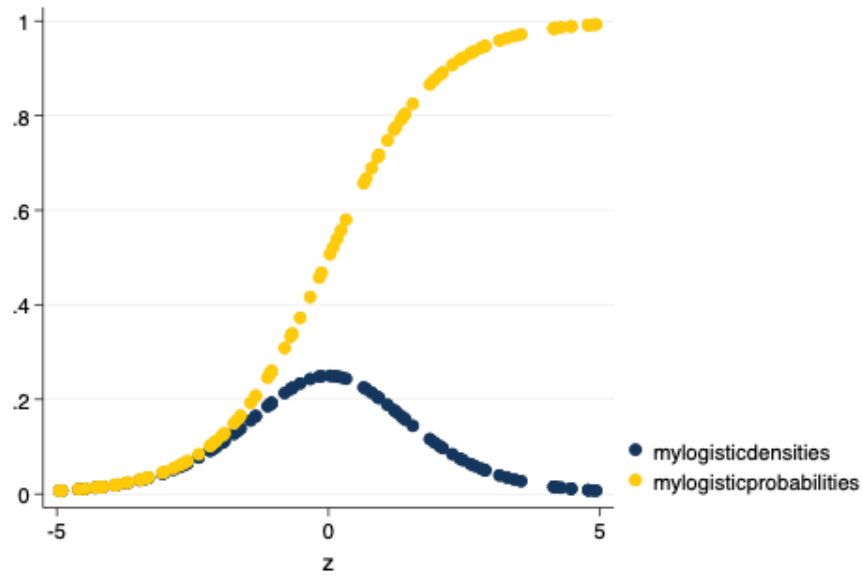
4

Figure 3: Standard and Cumulative Logistic Curves

```
Iteration 0:    log likelihood = -433.05123
Iteration 1:    log likelihood = -420.07608
Iteration 2:    log likelihood = -419.28644
Iteration 3:    log likelihood = -419.28513
Iteration 4:    log likelihood = -419.28513

Logistic regression                              Number of obs   =       1,223
                                                 LR chi2(1)      =       27.53
                                                 Prob > chi2     =      0.0000
Log likelihood = -419.28513                      Pseudo R2       =      0.0318
```

happy_3_D	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]
coninc_3	.0000134	2.93e-06	4.58	0.000	7.68e-06	.0000192
_cons	1.484066	.1381599	10.74	0.000	1.213277	1.754854

Comparison of LPM, Probit and Logistic Coefficients

NB: Negative vs. positive β.

```
. quietly probit happy_3_D coninc_3

. est store myprobit

. quietly logit happy_3_D coninc_3

. est store mylogit

. est table myprobit mylogit
```

Variable	myprobit	mylogit
coninc_3	6.430e-06	.00001343
_cons	.92440857	1.4840659

Logistic Model (2)

Derivation of logistic model from linear probability model. Using instructor notes

Interpretation of Odds Ratios (Robert Mare)

$$0 < OR < 1$$

indicates that an increase in x is associated with a decrease in y.

$$1 < OR < \infty$$

indicates that an increase in x is associated with an increase in y.

A Poem About Logistic Regression

Complete Determination

See handout

Rare Events

- Statistical power
- Complete determination

Predicted Probabilities

Discussion

The General Linear Model

Interaction Terms

See interactive demo, or example script.

https://agrogan1.github.io/multilevel/logistic-interactions/logistic-interactions.html