Logistic Regression

Andy Grogan-Kaylor

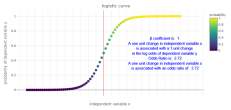
9 May 2020

Last Updated

```
. display "$S_TIME  $S_DATE"
17:40:15   9 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) 0	1	Total
VERY HAPPY	0	391	391
PRETTY HAPPY	0	758	758
NOT TOO HAPPY	153	0	153
Total	153	1,149	1,302

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)
```

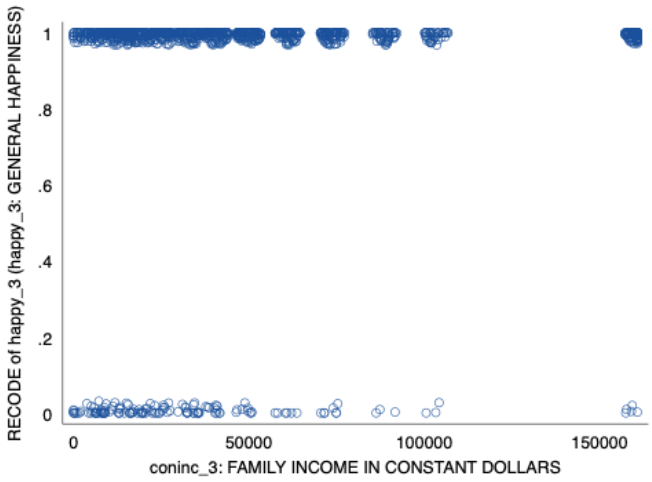


Figure 1: Happiness and Income

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

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)
```

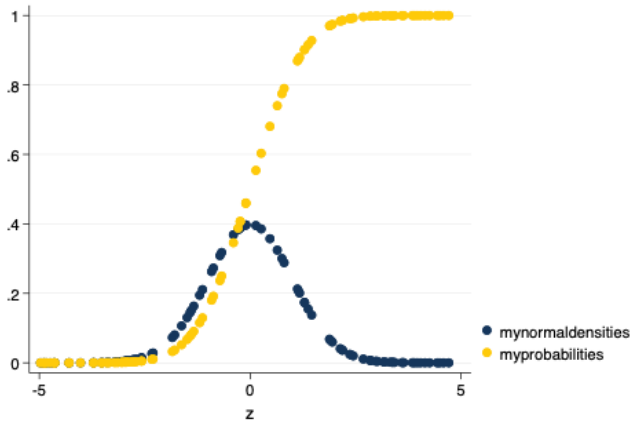


Figure 2: Standard and Cumulative Normal Curves