Complete Separation

Andy Grogan-Kaylor

31 Jul 2020

Adapted from an example at IDRE @ UCLA

```
. use complete-separation.dta, clear

. twoway scatter y x1, scheme(michigan)

. graph export scatter1.png, width(500) replace
(file scatter1.png written in PNG format)
```

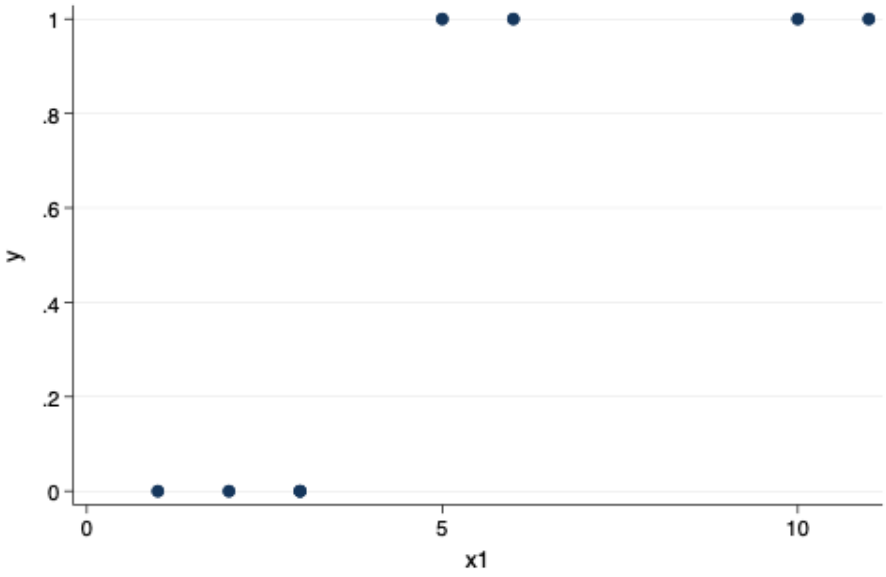

Figure 1: y by x1

```
. twoway scatter y x2, scheme(michigan)

. graph export scatter2.png, width(500) replace
(file scatter2.png written in PNG format)
```

From IDRE:

"What happens when we try to fit a logistic regression model of Y on X1 and X2 using our small sample data shown above? Well, the maximum likelihood estimate on the parameter for X1 does not exist. In particular with this example, the larger the coefficient for X1, the larger the likelihood. In other words, the coefficient for X1 should be as large as it can be, which would be infinity!"

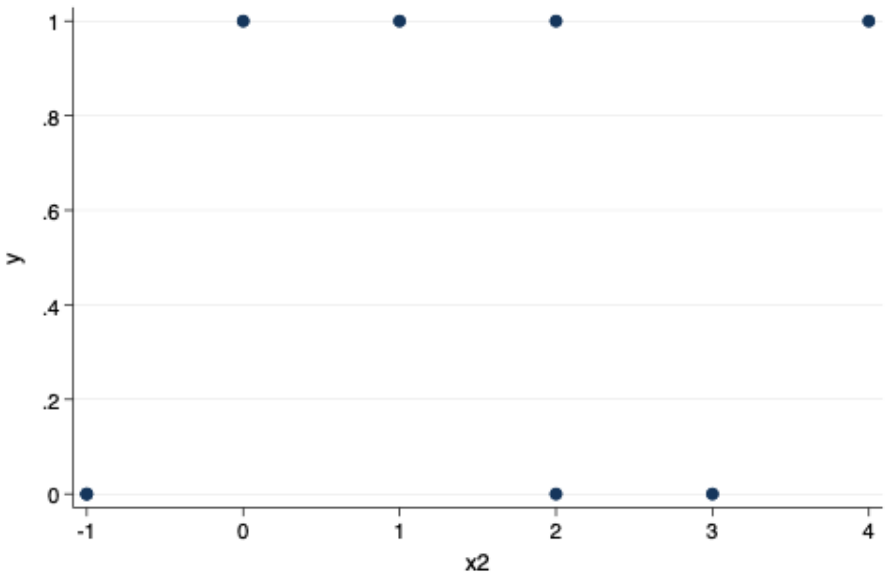

Figure 2: y by x2

```
. capture noisily logit y x1 x2
outcome = x1 > 3 predicts data perfectly
```

Stata provides a warning here, and would not estimate the model. We used `capture` to *capture* the error code and keep running the do file despite the error. `noisily` ensured that we saw any error messages.

R would still estimate the model, but will provide a *somewhat hidden* warning.