Generalized Linear Models

Andy Grogan-Kaylor

19 Jun 2020 18:17:38

Introduction

This discussion closely follows the Stata help for the *generalized linear model*, see `help glm`.

Briefly, per Stata documentation, in the *generalized linear model* framework, we consider models of the form:

$$g(E(y)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ...$$

where y is distributed as F *i.e.* $y \sim F$.

$g(y)$ is called the link function; F is called the distribution. Hence:

Link	Distribution	Standard Command	glm
identity	normal	`regress y x`	`glm y x, link(identity) family(gaussian)`
logit	bernoulli	`logit y x`	`glm y x, link(logit) family(binomial)`
probit	bernoulli	`probit y x`	`glm y x, link(probit) family(binomial)`

Palmer Penguins

Thes examples use the *Palmer Penguins* data set: https://github.com/allisonhorst/palmerpenguins.

```
. clear all

. use penguins.dta, clear
```

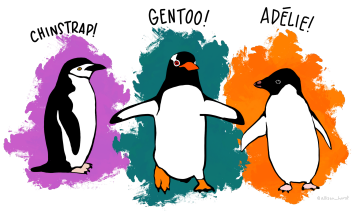

Figure 1: Palmer Penguins Illustration from @allison_horst

Models

I use the Stata prefix `quietly` to run the models without output. I then store the results using `estimates store`. Finally, I present all the results together in compact form using `estimates table`.

What Predicts Culmen Depth?

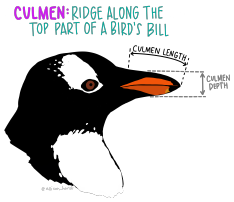

Figure 2: Culmen Depth from @allison_horst

```
. quietly: regress culmen_depth_mm body_mass_g flipper_length_mm

. est store usual_OLS // store estimates usual OLS

. quietly: glm culmen_depth_mm body_mass_g flipper_length_mm, link(identity) family(gaussian)

. est store glm_OLS // store estimates glm OLS
```

What Predicts That A Penguin Lives on Dream Island?

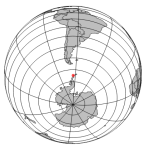

Figure 3: Location of Dream Island

```
. tabulate island
```

island	Freq.	Percent	Cum.
Biscoe	168	48.84	48.84
Dream	124	36.05	84.88
Torgersen	52	15.12	100.00
Total	344	100.00	

```
. generate dream = island == 2

. label variable dream "Penguin Lives on Dream Island"

. quietly: logit dream flipper_length_mm body_mass_g

. est store usual_logit // store estimates usual logit

. quietly: glm dream flipper_length_mm body_mass_g, link(logit) family(binomial)

. est store glm_logit // store estimates glm logit
```

```
. quietly: probit dream flipper_length_mm body_mass_g

. est store usual_probit // store estimates usual probit

. quietly: glm dream flipper_length_mm body_mass_g, link(probit) family(binomial)

. est store glm_probit // store estimates glm probit
```

Results

```
. est table usual_OLS glm_OLS usual_logit glm_logit usual_probit glm_probit, star
```

Variable	usual_OLS	glm_OLS	usual_logit	glm_logit
_				
body_mass_g	.00037535			
flipper_le~m	-.1006443***			
_cons	35.794997***			
culmen_dep~m				
body_mass_g		.00037535		
flipper_le~m		-.1006443***		
_cons		35.794997***		
dream				
flipper_le~m			-.0160116	-.0160116
body_mass_g			-.0013785***	-.0013785***
_cons			8.193819**	8.193819**

legend: * p<0.05; ** p<0.01; *** p<0.001

Variable	usual_probit	glm_probit
_		
body_mass_g		
flipper_le~m		
_cons		
culmen_dep~m		
body_mass_g		
flipper_le~m		
_cons		
dream		
flipper_le~m	-.01114532	-.01114532
body_mass_g	-.00082575***	-.00082575***
_cons	5.2018764**	5.2018764**

legend: * p<0.05; ** p<0.01; *** p<0.001