

# Generalized Linear Models

Andy Grogan-Kaylor

25 Oct 2020 08:12:13

## 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:

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

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

$g$  is called the link function;  $F$  is called the distribution. Hence<sup>1</sup>:

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

Though not the subject of this discussion, it is worth noting here, that *count regression* models follow a similar logic.

		Link	Distribution	Standard Command	glm
log	poisson	<code>poisson y x</code>		<code>glm y x, link(log)</code>	<code>family(poisson)</code>
log	nbinomial	<code>nbinomial y x</code>		<code>glm y x, link(log)</code>	<code>family(nbinomial)</code>

## Palmer Penguins

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

```
. clear all  
  
. use penguins.dta, clear
```

---

<sup>1</sup>This table and the table below draw heavily on the Stata documentation.

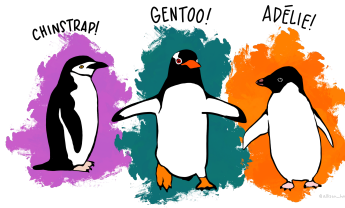


Figure 1: Palmer Penguins Illustration from @allison\_horst

## 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?

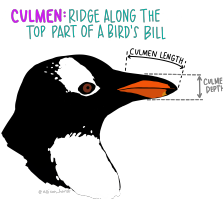


Figure 2: Culmen Depth from @allison\_horst

```
. 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?

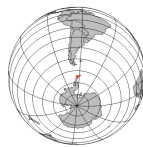


Figure 3: Location of Dream Island

```
. 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	usual_probit
body_mass_g	.00037535				
flipper_length_mm	-.1006443***				
_cons	35.794997***				
culmen_depth_mm					
body_mass_g		.00037535			
flipper_length_mm		-.1006443***			
_cons		35.794997***			
dream					
flipper_length_mm			-.0160116	-.0160116	-.01114532
body_mass_g			-.0013785***	-.0013785***	-.00082575***
_cons			8.193819**	8.193819**	5.2018764**

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

Variable	glm_probit
body_mass_g	
flipper_length_mm	
_cons	
culmen_depth_mm	
body_mass_g	
flipper_length_mm	
_cons	
dream	
flipper_length_mm	-.01114532
body_mass_g	-.00082575***
_cons	5.2018764**

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