LOGISTIC REGRESSION MODELS
Chapman & Hall/CRC

Joseph M. Hilbe

Published May 11, 2009

ERRATA AND ADDITIONS: 2nd Printing

Note: Printing number identified by rightmost number under copyright date and nation where printed on page before the Table of Contents; eg, 10 9 8 7 6 5 4 3 2

CHAPTER 2

Page 35: Add following text to paragraph directly above Equation 2.20:
"Although we do not use it in our calculations here, the standard error of an odds ratio is determined by using the **delta method**; i.e. exp(β)*se(β)."

CHAPTER 3

Page 52: Equation 3.6, missing = sign.

$$\partial^2 L = \frac{\partial^2 L}{\partial \beta \partial \beta'}$$

Page 53: Equation 3.10, and the sentence directly following: the term left of the = sign should not read $\partial L/\partial \beta$, but rather:

$$L(\theta,\phi; y) = \text{<rest is OK>}$$

"By the chain rule:" to now read:
"Solving for L with respect to β by using the chain rule, we have"

Page 53: Equation 3.11, a ∑ symbol should be to the right of the = sign,
Page 53: Equation 3.12, delete θ in $y\theta$ (in the numerator to the right of the ∑ symbol)
Page 53: Equation 3.17, far left term should read $\partial\mu/\partial\eta$, not the inverse.
Page 54: top line should read:
"where y and μ are the response and fitted values respectively, x is …"
Page 54: 3rd line on page: change to: "**Solving for $\partial^2 L$ − Fisher Scoring**"
Page 56: Equation 3.38, replace ∑ to ∏, and have subscripts, to now appear as:

$$f(y_i; \theta_i,\phi) = \prod_{i=1}\{ (y_i\theta_i - b(\theta_i))/\alpha(\phi) + C(y_i;\phi) \} \qquad (3.38)$$

CHAPTER 4

Page 63: The final sentence before Equation 4.1 should read:
"Given these terms, the Bernoulli PDF can be expressed as:"

CHAPTER 5

Page 77: Top line on page: Substitute "three" for "two".
Page 107: 4th line of 1st full paragraph: Amend to read:
"… For instance, consider a response, e.g. *death*, that we are attempting to …"

Page 133:
CREATE TWO RANDOM VARIATES
The word "uniform" should be "runiform". Add one more) to end of formula. Read as

```
. gen x1 = abs(invnorm(runiform()))
. gen x2 = abs(invnorm(runiform()))
```

Page 133 Change => CREATE BINARY LOGISTIC RESPONSE WITH DEFINED DATA
BINOMIAL DENOMINATOR
TO =>: CREATE BINOMIAL LOGISTIC RESPONSE WITH DEFINED DATA

Page 154: 1st full paragraph, Change Long and Freese (2006a) to: Long and Freese (2006)

CHAPTER 6

Page 193: 3rd line from bottom. Coefficient of β_3 should be negative; ie −1.846994

CHAPTER 7

Page 270: 3rd line under Eq 7.25: The y^ should be $\hat{y}$.
Page 272: Eq 7.33 and Eq 7.34 are mistaken, Please correct to read as:

$$d = \text{sqrt}[2\sum\{\ln(1/\mu)\}] \quad \text{if } y = 1 \qquad (7.33)$$
$$d = \text{sqrt}[2\sum\{\ln(1/(1-\mu)\}] \quad \text{if } y = 0 \qquad (7.34)$$

Page 279: 3rd line from top (under 7.4.1.6 Likelihood Residuals), and line directly above Eq 7.47, should read: "…deviance residuals, and is defined as:"

CHAPTER 9

Page 337 Top full paragraph plus text through `glm y x1 x2 x3, <…>`

CHANGE FROM THIS (CURRENTLY IN TEXT)=>

The model below is supplied with the negative binomial heterogeneous or ancillary parameter value (.0357547). This value was previously obtained by modeling a maximum likelihood negative binomial, which estimated the ancillary parameter. In Stata, maximum likelihood negative binomial estimates are obtained using the **nbreg** command. The SAS GENMOD procedure estimates the ancillary parameter. See Hilbe (2007a) for a thorough discussion of this subject.

RATE NEGATIVE BINOMIAL

```
[. nbreg y x1 x2 x3, nolog exp(d)] /// obtain ML estimate of
                                      ancillary parameter

. glm y x1 x2 x3, fam(nb .0357547) lnoffset(d) nolog
```

CHANGE TO THIS=>

The model below obtains the negative binomial heterogeneity or ancillary parameter value of .0357547 from a maximum likelihood negative binomial algorithm called from within the **glm** program. It supplies the value and employs it as a constant to the **glm** estimating equations. See Hilbe (2007a) for a comprehensive discussion of this subject.

RATE NEGATIVE BINOMIAL

```
. glm y x1 x2 x3, fam(nb ml) lnoffset(d) nolog
  <rest of page the same>
------------------------------------------------------------------------------
```

CHAPTER 10

Page 357 : formula under CATEGORY OR LEVEL 3 should read (now mistake in subscript)
"Logit = $\ln[(p_1 + p_2 + p_3)/(1 - p_1 + p_2 + p_3)]$"

Page 364: Add sentence below to last sentence on page (before **distinct** command):
"The **distinct** command below is from Longton and Cox
(http://fmwww.bc.edu/repec/bocode/d/distinct.ado**).**"

CHAPTER 11

Page 391: Change last sentence on the page and add another:
"We use the **prtab** command to do our work (Long, 1997). **prtab** is in **spost9_ado**
(http://www.indiana.edu/~jslsoc/stata) "

CHAPTER 12

Page 414, top most programming code: The second line of the "recode" command should read (5 42/52=57), not (4 42/52=57) as in the book.
Page 423: 2nd line of text from bottom. Change Long and Freese (2006a) to: (2006)

CHAPTER 15

Page 548: 4th line from top. The word "converge" should be "convergence"
Page 558: Exercise 15.2, start of second line. Change Exercise 12.3 to 5.3.

APPENDIX G

Page 601: right column, item : prtab. The author is Long, not Williams.

APPENDIX H

Page 611: First paragraph, add date to Long and Freese. "Long and Freese (2006) have …"

REFERENCE

Page 619: Only the second reference to Long and Freese should be given, and only with the date (2006). In other words, delete the reference to Long, "J.S. and J. Freese (2006a), *Regression* …." In the next reference, change to "Long, J.S. and J. Freese (2006), *Regression* …"

ADDITIONS TO TEXT : NEW CODE; CLARIFICATION

CHAPTER 4

Page 65: expand Equation 4.14 to appear as

$$L(y=1) = \sum\{\ln(\mu/(1-\mu)) + \ln(1-\mu)\} \qquad (4.14a)$$

or

$$L(y=1) = \sum\{\ln(\mu) + (1-y)\ln(1-\mu)\} \qquad (4.14b)$$

Page 66: expand Equation 4.24 to read as:

$$\frac{\partial(L)}{\partial \mu} = \frac{y}{\mu} - (1-y)(1-\mu)^{-1} = \frac{y-\mu}{\mu(1-\mu)} \qquad (4.24)$$

CHAPTER 5

A suite of random number generators were added to Stata version 11, which only became available after the text was written. One may now use some of these new functions in place of rndbinx(Hilbe) of genbinomial (Gutierrez) to create synthetic models. The amendments below change the text from the use of **genbinomial** to functions such as **rbinomial()**. Changes are found on pages: 133, 323, 324, 326, 327, 335, and 586. A single line of genbinomial is now two lines of code.

Page 133
Change =>
CREATE BINARY LOGISTIC RESPONSE WITH DEFINED DATA
BINOMIAL DENOMINATOR
TO =>:
CREATE BINOMIAL LOGISTIC RESPONSE WITH DEFINED DATA

Page 133
DELETE =>

```
. genbinomial y, xbeta(xb) n(100)
```

REPLACE WITH (in place of the `genbinomial` command) =>

```
. gen d = 100
. gen exb = 1/(1+exp(-xb))
. gen by = rbinomial(d, exb)
```

CHAPTER 9 [update to new code]

Page 323 near top
DELETE=>

```
. genbinomial y, xbeta(xb) de(d)
```

REPLACE WITH=>

```
. gen exb = 1/(1+exp(-xb))
. gen y = rbinomial(d, exb)
```

Page 324 near bottom
DELETE=>
```
. genbinomial yi, xbeta(xbi) de(d)
```

REPLACE WITH=>
```
. gen exbi = 1/(1+exp(-xbi))
. gen yi = rbinomial(d, exbi)
```

Page 326 near top
DELETE=>
```
. genbinomial ysq, xbeta(xbsq) de(d)
```

REPLACE WITH=>
```
. gen exbq = 1/(1+exp(-xbq))
. gen ysq = rbinomial(d, exbq)
```

Page 327 near bottom
DELETE=>
```
. genbinomial yp, xbeta(xb) de(d)
```

REPLACE WITH=>
```
. gen double exbp = normprob(xb)
. replace exbp=.99999999 if exbp>.99999999  // if need 50000 obs
. gen double yp = rbinomial(1, exbp)
```

Page 335 middle of page
DELETE=>
```
. genbinomial y, xbeta(xb) de(d)
```

REPLACE WITH=>
```
. gen exb = 1/(1+exp(-xb))
. gen y = rbinomial(d, exb)
```

CHAPTER 12

Page 419, Under the top-most statistical output, and over section 12.3, change text to read: "Interpretation of the odds ratios follow the same logic as the ordered logistic model. Predicted levels may be accessed using the **ocrpred** command, as done for **ologit**. Note that the number of observations in the **ocratio** model above has been inflated to 997 from 601. The reason is based on how levels are compared: Level 1 vs Levels 2,3, and Level 2 vs Level 3.
This results in [205+(204+192)] + [204+192] = 997."

REFERENCES

Page 621
Add to the end of the reference: Shults,J., S Ratcliffe, M Leonard (2007) …
(http://www.cceb.upenn.edu/~sratclif/QLSproject.html)
Page 621
Add to the end of the reference: Shults, J., W. Sun, X. Tu, J. Amsterdam (2006)…
(http://biostats.bepress.com/upennbiostat/papers/art8/).