November 3, 2016

HILBE MCD E-BOOK2016 ERRATA 03Nov2016

Joseph M Hilbe

Modeling Count Data

Cambridge University Press: 17 Jul, 2014
e-book version: 2016 (with initial corrections)

Joseph M. Hilbe
ERRATA, CLARIFICATIONS AND ADDITIONS as of October 2016, updated 3Nov, 2016

ADDITION – NOT ERRATA
Page 50; type directly under the Stata code, " . center age, pre(c). . ."
R
```
> cage <- scale(rwm1984$age, center=TRUE, scale=FALSE)
```

CHANGE - ERRATA
p 68 top of page, delete the R code now in book. Substitute the following in its place. Graph code added since it was supposed to originally be in the book
R
```
library(COUNT); data(rwm5yr); rwm1984 <- subset(rwm5yr, year==1984);
myglm <- glm(docvis ~ outwork + age, family=poisson, data=rwm1984);
lpred <- predict(myglm, newdata=rwm1984, type="link", se.fit=TRUE);
up <- lpred$fit + (1.96 * lpred$se.fit);lo <- lpred$fit - (1.96 * lpred$se.fit);
eta <- lpred$fit; upci <- myglm$family$linkinv(up); mu <- myglm$family$linkinv(eta);
loci <- myglm$family$linkinv(lo); summary(loci); summary(mu); summary(upci);
layout(1); plot(eta, mu); lines(eta, loci, col=2, type='p');
lines(eta,upci, col=3, type='p')
```

ERRATA
Page 107 2nd paragraph on page, top line. "discusssed" should be "discussed".

NEW ERRATA
Section 5.2 (p 131)
Mid page, sentence under θ=1/α, should read, "… equations (5.1)-(5.3).."

NEW ERRATA
Table 5.4 (P 146) Middle of code, change line to read
```
xbnb <- predict(nbx); munb <-exp(xbnb)
```

ERRATA
Section 5.4.3 (p 158)
Third > on page. R code under "NEGATIVE BINOMIAL" near top of page. Add additional ")" at end of line.
```
> summary(NB <- nbinomial(cones ~ sntrees + sheight + scover, data=nut))
```

ERRATA
Section 5.5 (p 161)
Top line on page. Insert word "to" as shown below:
"predictors contribute to model extra-dispersion. Both are valuable diagnostic"

ERRATA
Page 163 5th line under Table 6.1. The world "function" should be "functions."
and ADDITION
Bottom line, change to : "… particularly with … 1 and 2, given a direct parameterization.

ADDITION-CLARIFICATION

Section 6.1 (P165) Place the following after code at the top of the page, and before text currently in book.

"where the first term in the parenthesis is the number, the second is the mean and third the value of alpha. We can place this formula …"

ADDITION-CLARIFICATION

P 167: mid page. Add the following to the line in center of page (in red):

You may compare the predicted counts for the NB2 (munb) and PIG (mupig) models as we have for other models. Note that the hat^2 statistic of $p=0.021$ indicates a violation of the linearity assumption; i.e., the model is mis-specified.

NEEDED AMENDMENT

Section 6.3 (p171)

Rewrite all of page 171 (between lines below) to appear as:

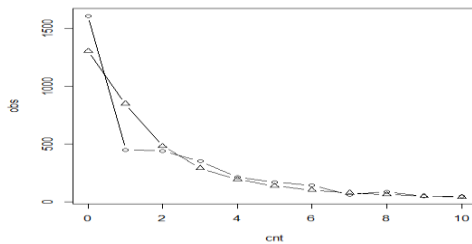==

Fig 6.1. PIG model: *docvis* observed vs predicted counts

test model for the **medpar** data is a zero-truncated model. We address both of these models in the following chapter.

Using the **rwm1984** data, we can produce a table of observed vs predicted PIG counts from 0-10 using the following R code. *pigrng.r* is located in the COUNT package and author's website.

```
library(COUNT); library(gamlss); data(rwm1984)
summary(pigmod <- gamlss(docvis ~ outwork + age + married + female +
            edlevel3 + edlevel3 + edlevel4, data=rwm1984, family=PIG))
# PIG dispersion parameter
exp(1.323)
# Predicted probabilities
yp <- pigrng(mean(pigmod$mu.fv),exp(pigmod$sigma.coefficient), 11)
ypig <- yp[ ,2]; ypig
# predicted & observed counts
pigexp <- dim(rwm1984)[1]*ypig ; pigexp
obs=table(rwm1984$docvis)[1:11] ; obs
# table of observed and predicted counts from 0-10
rbind(obs, pigexp[1:11])
chisq.test(obs, pigexp[1:11])
# Figure 6.1
pigpred <- pigexp[1:11]; cnt <- 0:10; plot(cnt, obs)
lines(cnt, obs, type="b"); lines(cnt, pigpred, type="b", pch=24)
```

Note that the line in Figure 6.1 having 0's are the observed counts; the line with triangles's at the count indicator are the predicted counts.

==

ERRATA
P. 176 mid page small caps. Change to read:
PREDICTED PROBABILITY OF 0 COUNTS FOR MEAN=9.854181

ERRATA
Section 7.1.2 (P 178)
2 lines above `ztnb` Stata code, Read as:
" … the term "1 – Pr(0)" must be …"

ADDITION – NOT ERRATA
P 187
Table 7.4: *R Components to Poisson-Logit Hurdle*
```
====================================================
visit <- ifelse(rwm1984$docvis >0, 1, 0)
table(visit)
logis <- glm(visit ~ outwork + age, data=rwm1984,
                   family=binomial(link="logit"))
summary(logis)
library(gamlss); library(gamlss.tr)
pltvis<-subset(rwm1984, rwm1984$docvis>0)
summary(ltpo <- gamlss(docvis~outwork+age,
        family=trun(0, "PO", "left"), data=pltvis))
library(pscl)
hpl2 <- hurdle(docvis ~ outwork + age, data=rwm1984,
    dist = "poisson", zero.dist="binomial", link="logit")
summary(hpl2)
====================================================
```

CLARIFICATION
Section 7.3.1 P 196 First "." Inset at bottom of page, change the last part of the second to last line on page to read:
"have a theory as to why there are a class of observations having excess zeros for both observed and expected zero counts. "

ERRATA
p. 196 Second to last line on page, amend word "supposed" to "supposedly"

CLARIFICATION
Chapter 8, p 211: Amend sentence before Eq 8.1 to read:
   "The generalized Poisson probability function is based on Consul (1989) and Famoye (1993), and this parameterization and Stata code on Harris, Yang and Hardin (2012).

ADDITION
Add a space and then the following 2 lines under Eq. 8.1

$$\mathcal{L}(\mu; y, \delta) = log(\mu) + (y-1)log(\mu + \alpha\mu) - \mu - \alpha y - log\big(gamma(y+1)\big)$$

with $\mu=\theta$.

ADDITION
Ch 8, page 216. Add sentences at end of main discussion, just above "Summary".
The *vglm* function is the VGAM package can be used to model the data using R. *vglm* output is the same as Stata's *gpoisson* output, except that the *vglm* dispersion parameter value (intercept:1) appears different from *gpoisson*. But by symbolizing the *vglm* dispersion as V, the value can be converted to Stata's *delta* value by using the formula: delta = (exp(V)-1) / (exp(V) +1). The results are the same. See Hilbe (2011) for details.
  To predict GP probabilities when delta<0, use
        . qui gen `newvar' = min(1,mu*(mu+delta*`l')^(`l'-1) * ///
            exp(-mu-delta*`l') / exp(lngamma(`l'+1))))
To predict GP probabilities when delta>0, use
        . qui replace `newvar' = mu*(mu+delta*`'l') * (`l'-1) * ///
             exp(-mu-delta*`l') / exp(lngamma(`l'+1)))) if `newvar'==.

CORRECTION
Section 9.2.1 (p 226)
3rd line following equation 9.3: 2014b should read 2015; ie, (Hardin & Hilbe, 2015)

CLARIFICATION
Section 9.2.2 (p 229)
Left censoring: "Any response in the data that is less than 3 is considered to be equal to 3."
  Right censoring: "Any response in the data that is greater than 15 is considered to be equal to 15."

CLARIFICATION
Section 9.4 (p 236) Amend line directly following equation 9.4 to read as:
"It is important to remember that the original purpose of using GAM was to determine the appropriate transformation needed by a continuous predictor in order to affect linearity, although many analysts now use GAM as a model in its own right. A GAM employs…"

CLARIFICATION
Section 9.4 (p 237) Final sentence on page. Change "…, but SAS does." to "… but SAS and R do."

CORRECTION -- ERRATA
Section 9.7 (p 246) Equation 9.6. Change to read:

$$f(y) = \frac{\Gamma(\alpha + \rho)\Gamma(k + \rho)}{\Gamma(\alpha + k + \rho)\Gamma(\rho)} \frac{\alpha_y k_y}{(\alpha + k + \rho)_y} \frac{1}{y!}$$

with $\alpha, \rho, k \geq 0$;   y = 0, 1, 2, …

GRAMMAR CORRETION
Section 9.8 (p 249) 2nd paragraph, 8th line from top. Sentence begins with a "For". Change "For": to "Given".
   "described by a gamma distribution. Given a binary variable, for instance, we"

CLARIFICATION
Third line from bottom p 249. Insert "informative" between words "no" and "priors".

CLARIFICATION- ERRATA

Section 9.9 (p 253)

Amend bottom lines of middle paragraph beginning with "Truncated and.." Read as:

"Stata's new **trncregress** command (Hardin and Hilbe (2015) provides truncated models for Poisson, NB2, PIG, generalized Poisson, NB-P, NB-W, and NB-F. Censored models for these distributions will be available in early 2016."

ERRATA - CORRECTIONS

Appendix (p271) Correct the following entries,

and J.M Hilbe (2012), *Generalized Linear Models and Extensions, third edition*, College Station, TX: Stata Press

and J.M Hilbe (2013), *Generalized Estimating Equations, Second edition*, Boca Raton, FL: Chapman and Hall/CRC.

DELETE

and J.M. Hilbe. 2014b. "Truncation Regression Models for Count Data.", *Stata Journal* 14. 2:280-291.

REPLACE WITH

and J.M. Hilbe, 2015. "Regression models for count models from truncated distributions", *Stata Journal* 15, 1:226-246.

Harris, T., J.M. Hilbe, and J.W. Hardin 2014, "Modeling Count Data with Generalized Distributions." *Stata Journal*. 14, 3:562-579.