Poisson regression with VGAM

Packages

Useful packages: *lmtest*, *MASS*, *pscl*, *sandwich*, *VGAM*

```
install.packages("lmtest", "MASS",
   "pscl", "sandwich", "VGAM")
```

The vgamadata data set

```
> head(vgamdata)
         X1         X2 Y
1 -5.120966  0.7158503 7
2 -3.925800  2.3375171 7
3 -2.115866 -2.8088281 0
4 -3.659427  6.6888561 4
5 -1.859439  2.6380000 1
6  1.758675  4.8020512 0
```

Using `vglm()` from package VGAM

```
library(VGAM)
summary(vglmFit <- vglm(Y ~ X1 + X2,
family=poissonff, data=vgamdata))
# not shown
```

Analyse event rates

offset is the exposure ln(t)

```
Nt   <- 100
Ti   <- sample(20:40, Nt, replace=TRUE)
Xt   <- rnorm(Nt, 100, 15)
Yt   <- rbinom(Nt, size=Ti, prob=0.5)
```

```
glm(Yt ~ Xt, family=poisson(link="log"), offset=log(Ti))

Call:  glm(formula = Yt ~ Xt, family = poisson(link = "log'

Coefficients:
(Intercept)           Xt
-0.5466      -0.0017

Degrees of Freedom: 99 Total (i.e. Null);  98 Residual
Null Deviance:      51.74
Residual Deviance: 50.73    AIC: 502.6
```

Overdispersion

Adjusted Poisson-regression

Same parameter estimates as in Poisson model, but different standard errors, hence different p-values

```
glmFitQP <- glm(Y ~ X1 + X2,
      family=quasipoisson(link="log"),
      data=vgamdata)
summary(glmFitQP)
```

```
Call:
glm(formula = Y ~ X1 + X2, family = quasipoisson(link = "lo
data = vgamdata)

Deviance Residuals:
Min        1Q   Median       3Q       Max
-3.3237  -1.2454  -0.2932   0.8200   2.9249
```

```
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.932e-01  1.297e-01   1.490    0.138
X1          -2.549e-01  2.839e-02  -8.979   <2e-16 ***
X2          -8.653e-05  1.443e-02  -0.006    0.995
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```

```
(Dispersion parameter for quasipoisson family taken to be 1

Null deviance: 494.5  on 199  degrees of freedom
Residual deviance: 356.2  on 197  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
```

Using vglm() from package VGAM

```
library(VGAM)
vglm(Y ~ X1 + X2, family=quasipoissonff, data=vgamdata)
# not shown
```

Heteroscedasticity consistent standard errors
Same parameter estimates as in Poisson model, but different standard errors, hence different p-values

```
library(sandwich)
hcSE <- vcovHC(glmFitP, type="HC0")

library(lmtest)
coeftest(glmFitP, vcov=hcSE)

z test of coefficients:

Estimate  Std. Error z value Pr(>|z|)
(Intercept)  0.19316888  0.13268996  1.4558   0.1455
X1          -0.25491612  0.02698458 -9.4467   <2e-16 ***
X2          -0.00008653  0.01319493 -0.0066   0.9948
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```

Negative binomial regression with VGAM

Using glm.nb() from package MASS

```
library(MASS)
glmFitNB <- glm.nb(Y ~ X1 + X2, data=vgamdata)
```

Negative binomial regression with VGAM

```
summary(glmFitNB)

Call:
glm.nb(formula = Y ~ X1 + X2, data = vgamdata, init.theta =
link = log)

Deviance Residuals:
Min        1Q    Median       3Q       Max
-2.7695  -1.0533  -0.2315   0.6472   2.4427
```

Negative binomial regression with VGAM

```
Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.166025   0.125963   1.318    0.187
X1          -0.261257   0.029119  -8.972   <2e-16 ***
X2           0.002651   0.014735   0.180    0.857
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```

Negative binomial regression with VGAM

```
(Dispersion parameter for Negative Binomial(5.1819) family

Null deviance: 345.33  on 199  degrees of freedom
Residual deviance: 256.60  on 197  degrees of freedom
AIC: 824.58
```

Negative binomial regression with VGAM

Number of Fisher Scoring iterations: 1
Theta: 5.18 Std. Err.: 1.79
2 x log-likelihood: -816.58

Negative binomial regression with VGAM

Using vglm() from package VGAM

```
library(VGAM)
vglm(Y ~ X1 + X2, family=negbinomial, data=vgamdata)
# not shown
```

Test the negative binomial model against the Poisson model

Test the negative binomial model against the Poisson model

```
library(pscl)
odTest(glmFitNB)
```

Likelihood ratio test of H0: Poisson, as restricted NB model: n.b., the distribution of the test-statistic under H0 is non-standard e.g., see help(odTest) for details/references
Critical value of test statistic at the alpha= 0.05 level: 2.7055
Chi-Square Test Statistic = 15.8533 p-value = 3.422e-05

Zero-inflated Regression models with VGAM

Zero-inflated Poisson regression

```
library(pscl)
ziFitP <- zeroinfl(Y ~ X1 + X2 | 1, dist="poisson", data=vg
summary(ziFitP)
```

Zero-inflated Poisson regression with VGAM

```
Call:
zeroinfl(formula = Y ~ X1 + X2 | 1, data = vgamdata, dist =

Pearson residuals:
Min       1Q  Median       3Q      Max
-1.6788 -0.8887 -0.1755  0.7678  3.3165
```

Zero-inflated Poisson regression with VGAM

```
Count model coefficients (poisson with log link):
Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.456827   0.124391   3.672  0.00024 ***
X1          -0.216984   0.025879  -8.385  < 2e-16 ***
X2          -0.002068   0.012180  -0.170  0.86516
```

Zero-inflated Poisson regression with VGAM

Zero-inflation model coefficients (binomial with logit link):
Estimate Std. Error z value Pr(¿—z—) (Intercept) -1.8868 0.2927 -6.446 1.15e-10 *** — Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Number of iterations in BFGS optimization: 10 Log-likelihood: -402.4 on 4 Df

Zero-inflated Poisson regression with VGAM

Using vglm() from package VGAM

```
library(VGAM)
vglm(Y ~ X1 + X2, family=zipoissonff, data=vgamdata)

# not shown
```

Vuong Test

- The Vuong test compares the zero-inflated model with an ordinary poisson regression model.
- A significant z-test indicates that the zero-inflated model is better.

Vuong-Test

Using `vuong()` from package **pscl**: Poisson model against zero-inflated Poisson model

```
library(pscl)
vuong(ziFitP, glmFitP)
Vuong Non-Nested Hypothesis Test-Statistic: -0.6038871
(test-statistic is asymptotically distributed N(0,1) under
null that the models are indistinguishible)
in this case:
model2 > model1, with p-value 0.27296
```

Zero-inflated negative binomial regression with VGAM

```
ziFitNB <- zeroinfl(Y ~ X1 + X2 | 1, dist="negbin", data=vg
summary(ziFitNB)

Call:
zeroinfl(formula = Y ~ X1 + X2 | 1, data = vgamdata, dist =
```

Zero-inflated negative binomial regression with VGAM

```
Pearson residuals:
Min       1Q  Median       3Q      Max
-1.6352 -0.8709 -0.1633  0.7424  3.3172

Count model coefficients (negbin with log link):
Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.426760   0.136718   3.121   0.0018 **
X1          -0.222478   0.028374  -7.841 4.48e-15 ***
X2          -0.001403   0.012926  -0.109   0.9135
Log(theta)   3.474664   1.443439   2.407   0.0161 *
```

Zero-inflated negative binomial regression with VGAM

```
Zero-inflation model coefficients (binomial with logit link
Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.973      0.341  -5.786  7.2e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '

Theta = 32.287
Number of iterations in BFGS optimization: 26
Log-likelihood: -402.1 on 5 Df
```

Zero-inflated negative binomial regression with VGAM

```
Using vglm() from package VGAM

library(VGAM)
vglm(Y ~ X1 + X2, family=zinegbinomial, data=vgamdata)
# not shown
```

Zero-inflated negative binomial regression with VGAM

Vuong-Test using `vuong()` from package pscl: negative binomial model against zero-inflated negative binomial model

```
library(pscl)
vuong(ziFitNB, glmFitNB)
Vuong Non-Nested Hypothesis Test-Statistic: -1.037018
(test-statistic is asymptotically distributed N(0,1) under
null that the models are indistinguishible)
in this case:
model2 > model1, with p-value 0.14986
```