pca poisson

```
##NB: Install countreg using this link in the console before running
##install.packages("countreg", repos="http://R-Forge.R-project.org")


library(pscl)
```

```
## Classes and Methods for R developed in the
## Political Science Computational Laboratory
## Department of Political Science
## Stanford University
## Simon Jackman
## hurdle and zeroinfl functions by Achim Zeileis
```

```
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 3.5.2
```

```
library(DT)
library(countreg)
```

```
## Loading required package: MASS
```

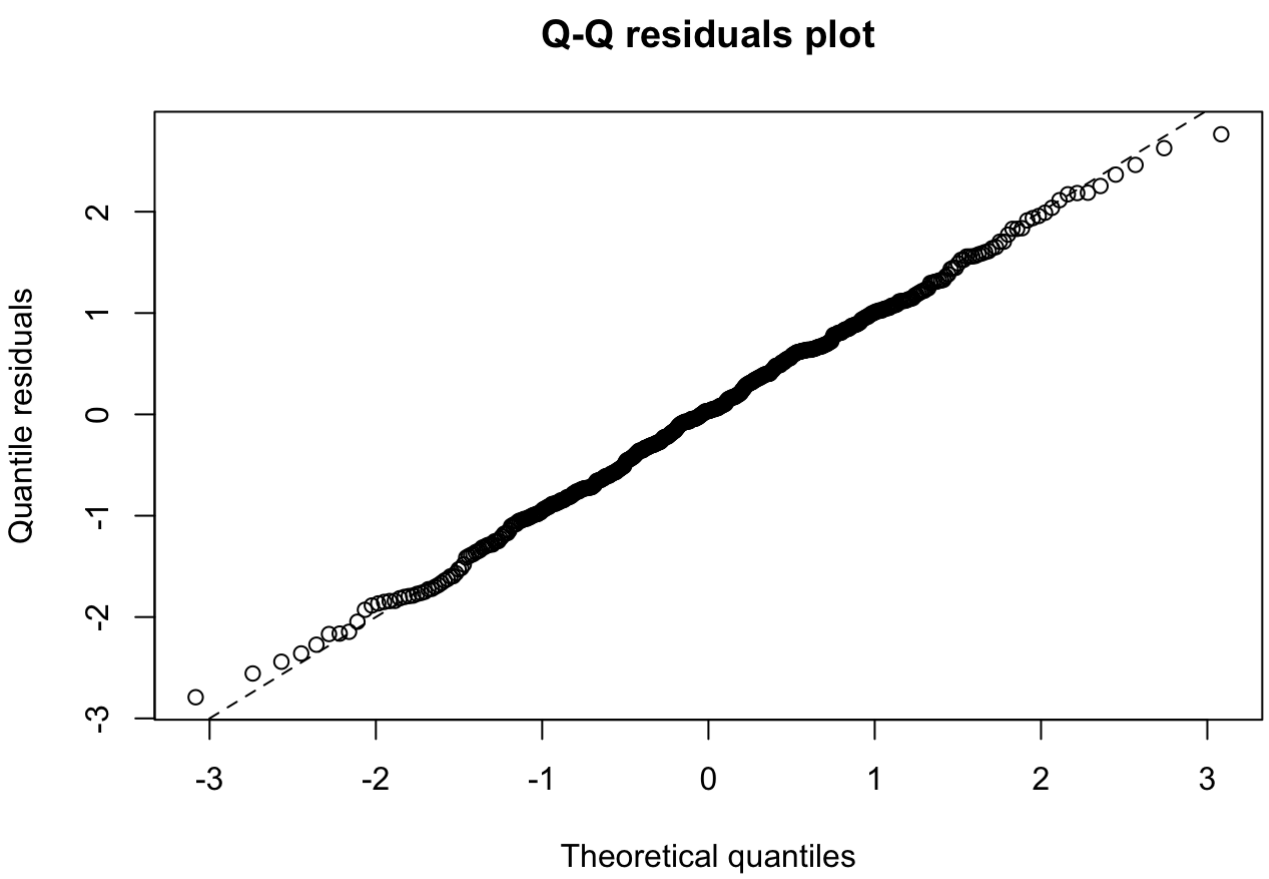
```
##
## Attaching package: 'countreg'
```

```
## The following objects are masked from 'package:pscl':
##
##     hurdle, hurdle.control, hurdletest, zeroinfl, zeroinfl.control
```

```
dat<-fread("/Users/11kolop/Downloads/chicago_by_census_with_transportation (1).csv")
dat.hom<-dat[dat$ofns_desc=="criminal homicide",]
pca <- princomp(na.omit(dat.hom)[,c(4:15)], cor = TRUE)

mod.zero.inflated.poisson<-zeroinfl(n ~ foreign_share2010 + share_black2010 + share_hisp
2010 + singleparent_share2010+mail_return_rate2010 + scale(transp/sqmi), data = na.omit
(dat.hom),dist="poisson")

mod.poisson<-glm(n ~ foreign_share2010 + share_black2010 + share_hisp2010 + singleparent
_share2010+mail_return_rate2010 + scale(transp/sqmi), data = na.omit(dat.hom),family="po
isson")

mod.pca.zero.inflated.poisson<-zeroinfl(na.omit(dat.hom)$n~scale(na.omit(dat.hom)$trans
p/na.omit(dat.hom)$sqmi)+pca$scores[,1] + pca$scores[,2]+pca$scores[,3] + pca$scores[,4
],dist="poisson")

mod.pca.poisson<-glm(na.omit(dat.hom)$n~scale(na.omit(dat.hom)$transp/na.omit(dat.hom)$s
qmi)+pca$scores[,1] + pca$scores[,2]+pca$scores[,3] + pca$scores[,4],family="poisson")

mod.zero.inflated.nb<-zeroinfl(n ~ foreign_share2010 + share_black2010 + share_hisp2010
 + singleparent_share2010+mail_return_rate2010 + scale(transp/sqmi), data = na.omit(dat.
hom),dist="negbin")

mod.pca.zero.inflated.nb<-zeroinfl(na.omit(dat.hom)$n~scale(na.omit(dat.hom)$transp/na.o
mit(dat.hom)$sqmi)+pca$scores[,1] + pca$scores[,2]+pca$scores[,3] + pca$scores[,4],dist=
"negbin")

BIC(mod.poisson,mod.zero.inflated.poisson,mod.pca.poisson,mod.pca.zero.inflated.poisson,
mod.zero.inflated.nb,mod.pca.zero.inflated.nb)
```

```
##                               df      BIC
## mod.poisson                    7 2844.371
## mod.zero.inflated.poisson     14 2741.713
## mod.pca.poisson                6 2488.937
## mod.pca.zero.inflated.poisson 12 2399.338
## mod.zero.inflated.nb          15 2252.208
## mod.pca.zero.inflated.nb      13 2169.982
```

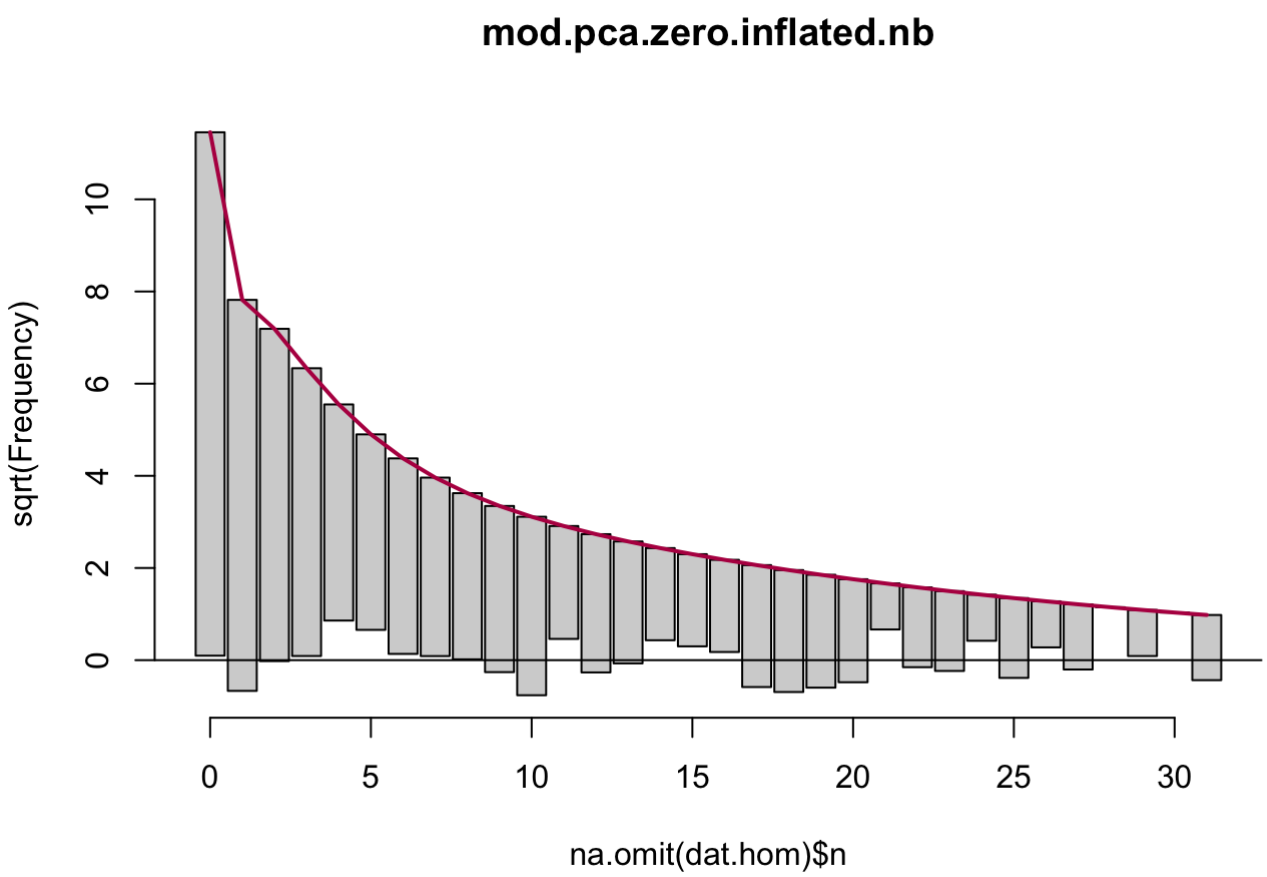
```
qqrplot(mod.pca.zero.inflated.nb)
```




```
summary(mod.pca.zero.inflated.nb)
```

```
##
## Call:
## zeroinfl(formula = na.omit(dat.hom)$n ~ scale(na.omit(dat.hom)$transp/na.omit(dat.ho
m)$sqmi) +
##     pca$scores[, 1] + pca$scores[, 2] + pca$scores[, 3] + pca$scores[,
##     4], dist = "negbin")
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.7443 -0.6702 -0.2405  0.4849  3.8149
##
## Count model coefficients (negbin with log link):
##                                                    Estimate Std. Error
## (Intercept)                                        1.234267   0.050660
## scale(na.omit(dat.hom)$transp/na.omit(dat.hom)$sqmi) -0.030214   0.052432
## pca$scores[, 1]                                    0.458097   0.022164
## pca$scores[, 2]                                   -0.019313   0.024382
## pca$scores[, 3]                                   -0.055115   0.040820
## pca$scores[, 4]                                    0.009071   0.050176
## Log(theta)                                         1.372075   0.146019
##                                                   z value Pr(>|z|)
## (Intercept)                                        24.364   <2e-16 ***
## scale(na.omit(dat.hom)$transp/na.omit(dat.hom)$sqmi)  -0.576    0.564
## pca$scores[, 1]                                    20.669   <2e-16 ***
## pca$scores[, 2]                                    -0.792    0.428
## pca$scores[, 3]                                    -1.350    0.177
## pca$scores[, 4]                                     0.181    0.857
## Log(theta)                                          9.397   <2e-16 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##                                                   Estimate Std. Error
## (Intercept)                                        -3.8318     0.8073
## scale(na.omit(dat.hom)$transp/na.omit(dat.hom)$sqmi)  -0.9916     0.4498
## pca$scores[, 1]                                    -1.1458     0.2834
## pca$scores[, 2]                                     0.9134     0.3761
## pca$scores[, 3]                                    -0.4569     0.4554
## pca$scores[, 4]                                     0.4405     0.4329
##                                                   z value Pr(>|z|)
## (Intercept)                                        -4.746 2.07e-06 ***
## scale(na.omit(dat.hom)$transp/na.omit(dat.hom)$sqmi)  -2.205   0.0275 *
## pca$scores[, 1]                                    -4.043 5.27e-05 ***
## pca$scores[, 2]                                     2.428   0.0152 *
## pca$scores[, 3]                                    -1.003   0.3158
## pca$scores[, 4]                                     1.018   0.3089
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Theta = 3.9435
## Number of iterations in BFGS optimization: 23
## Log-likelihood: -1045 on 13 Df
```

```
rootogram(mod.pca.zero.inflated.nb)
```


mod.pca.zero.inflated.nb
sqrt(Frequency)
0
2
4
6
8
10
0
5
10
15
20
25
30
na.omit(dat.hom)$n