Running Models

Erik Wang, Frankie Wunschel, Madison Volpe

4/22/2019

```
library(here)
```

```
## here() starts at /Users/madisonvolpe/Documents/Grad_School/Spring_2019/DACJS/CJA
library(gmodels)
```

Model One: Poisson Regression (Comparing Across (g) = Zipcodes)

$$s_g = Poisson(p_g * e^{\mu + \alpha c_g + \beta r_g})$$

- g is zipcode
- s_{g} is the number of 'drug arrests' in zipcode (g)
- p_{g} is the total population over age 18 in zipcode (g)
- c_{g} is # of EMS Dispatches concerning drugs made in zipcode (g) - this is our measure of 'drug activity' for this model
- r_{g} is the racial composition of zipcode (g) in this case it would be (proportion non-white)
- Some notes about this model is that we must be less granular and use zip if our measure of drug activity is ems calls bc that is the only measure that EMS data goes down too (it is pretty deidentified).

Load Data

```
mod1dat <- read.csv(here::here("data/ModelDatasets", "zipcodeModel1.csv"))
```

```
#plot(no.drug.arrests.13.18~proportionNonWhite, mod1dat, log="y")
#with(mod1dat, lines(lowess(no.drug.arrests.13.18~proportionNonWhite)))
```

Fit Poisson Model

```
mod1dat$zip <- factor(mod1dat$zip)
```

```
mod1 <- glm(formula = no.drug.arrests.13.18 ~ no.ems.calls.13.18 + proportionNonWhite,
            offset = log(Total.Population.Over18),family = poisson ,data = mod1dat)
```

```
arm::display(mod1)
```

```
## glm(formula = no.drug.arrests.13.18 ~ no.ems.calls.13.18 + proportionNonWhite,
##     family = poisson, data = mod1dat, offset = log(Total.Population.Over18))
##                    coef.est coef.se
## (Intercept)        -4.37     0.02
## no.ems.calls.13.18  0.00     0.00
## proportionNonWhite  2.07     0.04
## ---
##   n = 12, k = 3
```

```
##   residual deviance = 998.3, null deviance = 4680.1 (difference = 3681.8)
summary(mod1)
```

```
##
## Call:
## glm(formula = no.drug.arrests.13.18 ~ no.ems.calls.13.18 + proportionNonWhite,
##     family = poisson, data = mod1dat, offset = log(Total.Population.Over18))
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -16.0994   -5.1939   -0.0889    4.8409   16.9384
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -4.369e+00  2.276e-02 -191.98   <2e-16 ***
## no.ems.calls.13.18 1.353e-04  7.079e-06   19.12   <2e-16 ***
## proportionNonWhite 2.070e+00  4.233e-02   48.90   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 4680.1  on 11  degrees of freedom
## Residual deviance:  998.3  on  9  degrees of freedom
## AIC: 1109.8
##
## Number of Fisher Scoring iterations: 4
#plot(mod1)
```

- *Intercept* - no interpretation
- *Measure of drug activity* - coefficient is practically 0
- *Proportion non white* - The greater the proportion of nonwhite people in a zip, the greater the number of drug arrests. A one unit increase in the proportion of nonwhite people, the number of drug arrests will increase with a multiplicative effect of 7.92.
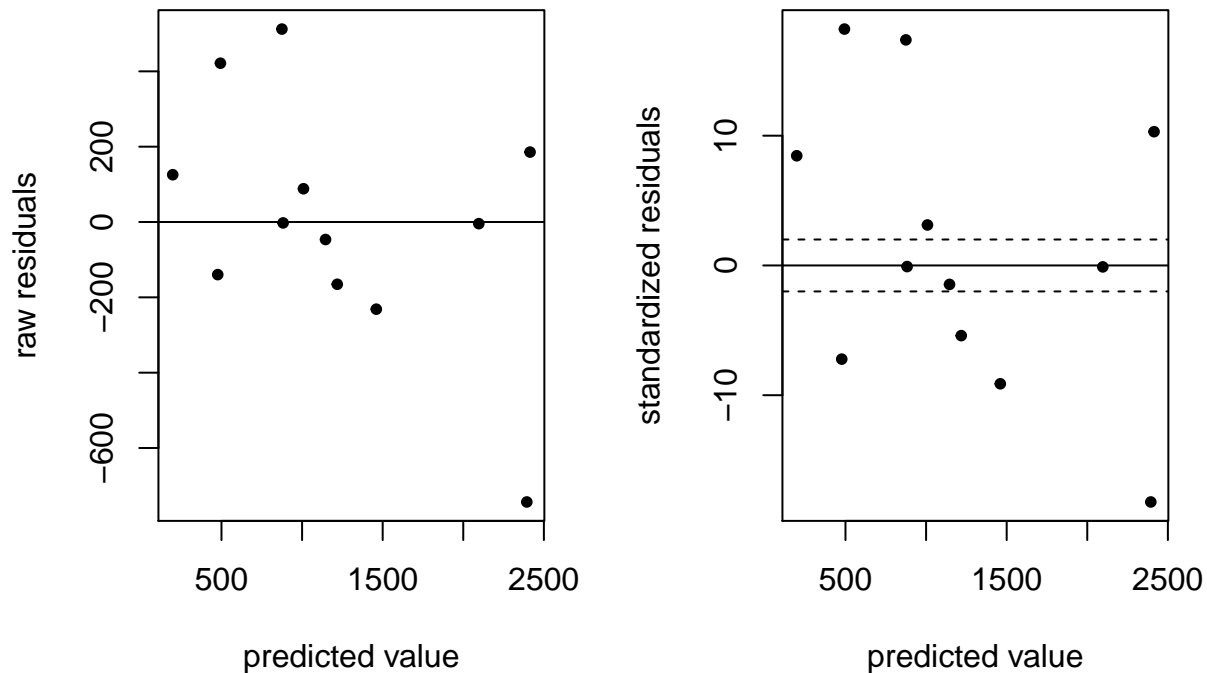
Test for overdispersion

```
yhat <- predict(mod1, type = 'response')
z <- (mod1dat$no.drug.arrests.13.18 - yhat) / sqrt(yhat)
cat('overdispersion ratio is', sum(z^2)/(12-3), "\n")
```

```
## overdispersion ratio is 121.597
cat('p-value of overdispersion test is', pchisq(sum(z^2), 12-3), "\n")
```

```
## p-value of overdispersion test is 1
par(mfrow=c(1,2))
pv <- fitted(mod1) #predicted values
r <- (mod1dat$no.drug.arrests.13.18 - fitted(mod1)) #residuals
plot(pv, r, pch=20, ylab="raw residuals", xlab="predicted value")
abline(h=0)
sr <- rstandard(mod1)#standardized residual
```

```
plot(pv, sr, pch=20, ylab="standardized residuals", xlab="predicted value")
abline(h=c(-2,0,2),lty=c(2,1,2))
```


- ALOT of overdispersion (122) and it is significant.

Attempt to Fix Overdispersion with Quasipoisson

```
mod1quasi <- glm(formula = no.drug.arrests.13.18 ~ no.ems.calls.13.18 + proportionNonWhite,
                 offset  = log(Total.Population.Over18),family = quasipoisson ,data = mod1dat)
```

```
summary(mod1quasi)
```

```
##
## Call:
## glm(formula = no.drug.arrests.13.18 ~ no.ems.calls.13.18 + proportionNonWhite,
##     family = quasipoisson, data = mod1dat, offset = log(Total.Population.Over18))
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -16.0994   -5.1939   -0.0889    4.8409   16.9384
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -4.369e+00  2.509e-01 -17.410 3.07e-08 ***
## no.ems.calls.13.18 1.353e-04  7.806e-05   1.734  0.11699
## proportionNonWhite 2.070e+00  4.668e-01   4.434  0.00164 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 121.5971)
##
```

```
##     Null deviance: 4680.1  on 11  degrees of freedom
## Residual deviance:  998.3  on  9  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```
```
#plot(mod1quasi)
#ci(mod1quasi)
```

Model Two: Poisson Regression (Comparing Across Zipcodes (g) + Time (t))

$$s_{g,t} = Poisson(p_{g,t} * e^{\mu + \alpha c_{g,t} + \beta r_{g,t}})$$

```
mod2dat <- read.csv(here("data/ModelDatasets", 'zipcodeModel2.csv'))
```

- g is zipcode
- s_{g,t} is the number of 'drug arrests' in zipcode (g)
- p_{g,t} is the total population over age 18 in zipcode (g) at time (t)
- c_{g,t} is # of EMS Dispatches concerning drugs made in zipcode (g) at time (t)- this is our measure of 'drug activity' for this model
- r_{g,t} is the racial composition of zipcode (g) at time (t) in this case it would be (proportion non-white)

```
mod2dat$zip <- factor(mod2dat$zip)
```

```
#plot(no.drug.arrests.14.17~proportionNonWhite, mod2dat, log="y")
#with(mod2dat, lines(lowess(no.drug.arrests.14.17~proportionNonWhite)))
```

Fit Poisson Model

```
mod2 <- glm(formula = no.drug.arrests.14.17 ~ no.ems.calls.14.17 + proportionNonWhite, offset = log(Tota
            family = poisson, data = mod2dat)
```

```
arm::display(mod2)
```

```
## glm(formula = no.drug.arrests.14.17 ~ no.ems.calls.14.17 + proportionNonWhite,
##     family = poisson, data = mod2dat, offset = log(Total.Population.Over18))
##                    coef.est coef.se
## (Intercept)        -6.07     0.03
## no.ems.calls.14.17  0.00     0.00
## proportionNonWhite  2.13     0.05
## ---
##   n = 48, k = 3
##   residual deviance = 1564.8, null deviance = 4203.8 (difference = 2638.9)
```

```
summary(mod2)
```

```
##
## Call:
## glm(formula = no.drug.arrests.14.17 ~ no.ems.calls.14.17 + proportionNonWhite,
##     family = poisson, data = mod2dat, offset = log(Total.Population.Over18))
##
## Deviance Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -15.0816  -3.3114   0.6571   3.9454  17.5663
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -6.073e+00  2.605e-02 -233.12   <2e-16 ***
## no.ems.calls.14.17  5.675e-04  4.282e-05   13.25   <2e-16 ***
## proportionNonWhite  2.126e+00  4.833e-02   43.99   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 4203.8  on 47  degrees of freedom
## Residual deviance: 1564.8  on 45  degrees of freedom
## AIC: 1905.7
##
## Number of Fisher Scoring iterations: 4
```

- *Intercept* - not interpretable

- *no.ems.calls.14.17* - again basically 0

- *proportionNonWhite* - so this is similar to the coefficient estimate from before, the greater the proportion of nonwhite people, the greater the number of drug arrests. A one unit increase in the proportion of nonwhite people, the number of drug arrests will increase with a multiplicative effect of 8.41.

Test for overdispersion

```
yhat <- predict(mod2, type = 'response')
z <- (mod2dat$no.drug.arrests.14.17 - yhat) / sqrt(yhat)
cat('overdispersion ratio is', sum(z^2)/(48-3), "\n") #n -k
```

```
## overdispersion ratio is 37.66266
```

```
cat('p-value of overdispersion test is', pchisq(sum(z^2), 48-3), "\n")
```

```
## p-value of overdispersion test is 1
```

- Again, overdispersion is a problem it is 37.66, and the p-value attached is 1, but adding the extra observations did seem to decrease it, which is a good thing I guess.

```
par(mfrow=c(1,2))
pv <- fitted(mod2) #predicted values
r <- (mod2dat$no.drug.arrests.14.17- fitted(mod2)) #residuals
plot(pv, r, pch=20, ylab="raw residuals", xlab="predicted value")
abline(h=0)
sr <- rstandard(mod2)#standardized residual
plot(pv, sr, pch=20, ylab="standardized residuals", xlab="predicted value")
abline(h=c(-2,0,2),lty=c(2,1,2))
```
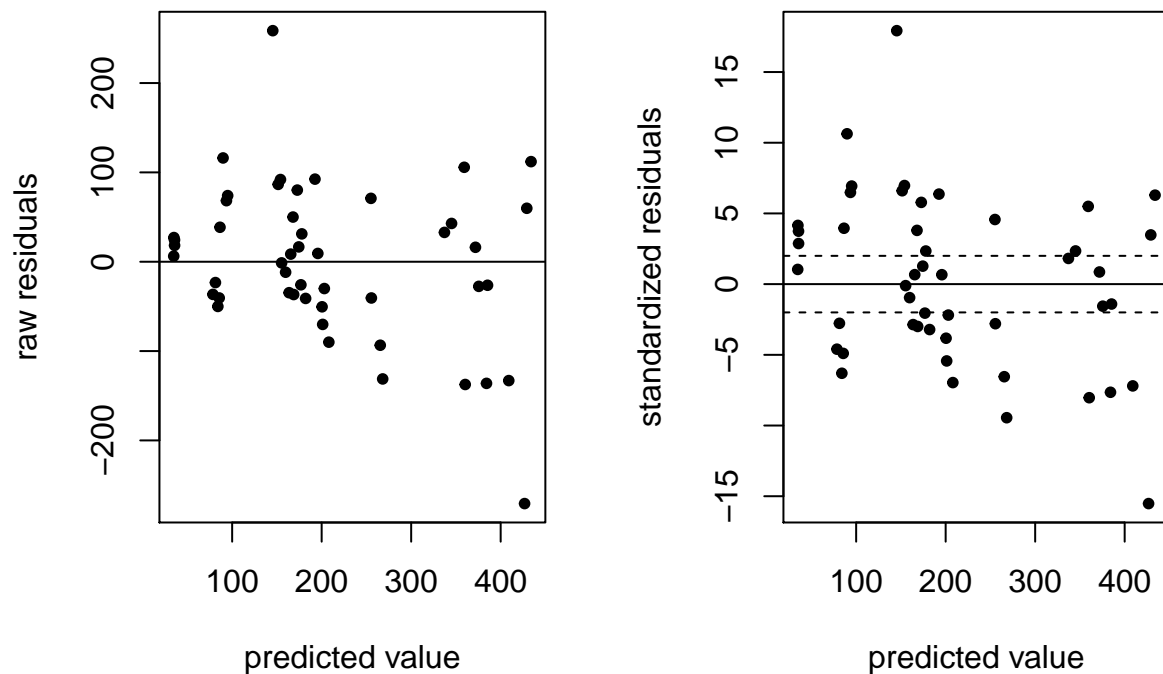
Quasipoisson

```
mod2quasi <- glm(formula = no.drug.arrests.14.17 ~ no.ems.calls.14.17 + proportionNonWhite, offset =
                 log(Total.Population.Over18),family = quasipoisson, data = mod2dat)
```

```
arm::display(mod2quasi)
```

```
## glm(formula = no.drug.arrests.14.17 ~ no.ems.calls.14.17 + proportionNonWhite,
##     family = quasipoisson, data = mod2dat, offset = log(Total.Population.Over18))
##                    coef.est coef.se
## (Intercept)          -6.07    0.16
## no.ems.calls.14.17    0.00    0.00
## proportionNonWhite    2.13    0.30
## ---
##   n = 48, k = 3
##   residual deviance = 1564.8, null deviance = 4203.8 (difference = 2638.9)
##   overdispersion parameter = 37.7
```

- The coefficient standard errors are definitely bigger than the model fit w. the poisson link, but they are not 'very big' in that I am definitely not worried about them.
- With this model, the proportionNonWhite parameter stays significant ($p<.05$), as well as the no.ems.calls.14.17 ($p<.05$).

```
#plot(mod2quasi)
#ci(mod2quasi)
```

Model Three: Same as Model One with Extra Drug Activity Measures

$$s_g = Poisson(p_g * e^{\mu+\alpha c_g+\beta r_g+\beta d_g+\beta z_g})$$

- g is zipcode

6

- s_{g} is the number of 'drug arrests' in zipcode (g)

- p_{g} is the total population in zipcode (g) - population over age 18

- c_{g} is # of EMS Dispatches concerning drugs made in zipcode (g) - this is our measure of 'drug activity' for this model

- r_{g} is the racial composition of zipcode (g) in this case it would be (proportion non-white)

- d_{g} is the total Number of Overdose deaths that occured from 2016-2018 in zipcode g

- z_{g} is the total Number of Naxolone saves that occured from 2016-2018 in zipcode g

```
mod3dat <- read.csv(here('data/ModelDatasets','zipcodeModel3.csv'))
```

```
mod3 <- glm(formula = no.drug.arrests.13.18 ~ no.ems.calls.13.18 + proportionNonWhite + Total.Naxolone.S
            Overdose.Death.Total, offset = log(Total.Population.Over18),family = poisson, data = mod3dat
```

```
arm::display(mod3)
```

```
## glm(formula = no.drug.arrests.13.18 ~ no.ems.calls.13.18 + proportionNonWhite +
##     Total.Naxolone.Saves + Overdose.Death.Total, family = poisson,
##     data = mod3dat, offset = log(Total.Population.Over18))
##                      coef.est coef.se
## (Intercept)          -3.68     0.05
## no.ems.calls.13.18    0.00     0.00
## proportionNonWhite    1.32     0.06
## Total.Naxolone.Saves -0.01     0.00
## Overdose.Death.Total  0.00     0.00
## ---
##   n = 12, k = 5
##   residual deviance = 684.2, null deviance = 4680.1 (difference = 3995.9)
```

```
summary(mod3)
```

```
##
## Call:
## glm(formula = no.drug.arrests.13.18 ~ no.ems.calls.13.18 + proportionNonWhite +
##     Total.Naxolone.Saves + Overdose.Death.Total, family = poisson,
##     data = mod3dat, offset = log(Total.Population.Over18))
##
## Deviance Residuals:
##     Min      1Q  Median      3Q     Max
## -14.866  -3.513   2.671   4.384  11.424
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -3.680e+00  4.597e-02 -80.054  < 2e-16 ***
## no.ems.calls.13.18    4.936e-04  3.046e-05  16.204  < 2e-16 ***
## proportionNonWhite    1.320e+00  5.785e-02  22.825  < 2e-16 ***
## Total.Naxolone.Saves -7.161e-03  6.002e-04 -11.932  < 2e-16 ***
## Overdose.Death.Total -4.508e-03  1.271e-03  -3.546 0.000392 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 4680.12  on 11  degrees of freedom
```

```
## Residual deviance:  684.22  on  7  degrees of freedom
## AIC: 799.7
##
## Number of Fisher Scoring iterations: 4
```

Test for overdispersion

```
yhat <- predict(mod3, type = 'response')
z <- (mod3dat$no.drug.arrests.13.18 - yhat) / sqrt(yhat)
cat('overdispersion ratio is', sum(z^2)/(12-5), "\n") #n -k
```

```
## overdispersion ratio is 94.66502
```

```
cat('p-value of overdispersion test is', pchisq(sum(z^2), 12-5), "\n")
```

```
## p-value of overdispersion test is 1
```

- There is a bit of overdispersion in this model (94.67) and it is statistically significant.

```
par(mfrow=c(1,2))
pv <- fitted(mod3) #predicted values
r <- (mod3dat$no.drug.arrests.13.18 - fitted(mod3)) #residuals
plot(pv, r, pch=20, ylab="raw residuals", xlab="predicted value")
abline(h=0)
sr <- rstandard(mod3)#standardized residual
plot(pv, sr, pch=20, ylab="standardized residuals", xlab="predicted value")
abline(h=c(-2,0,2),lty=c(2,1,2))
```
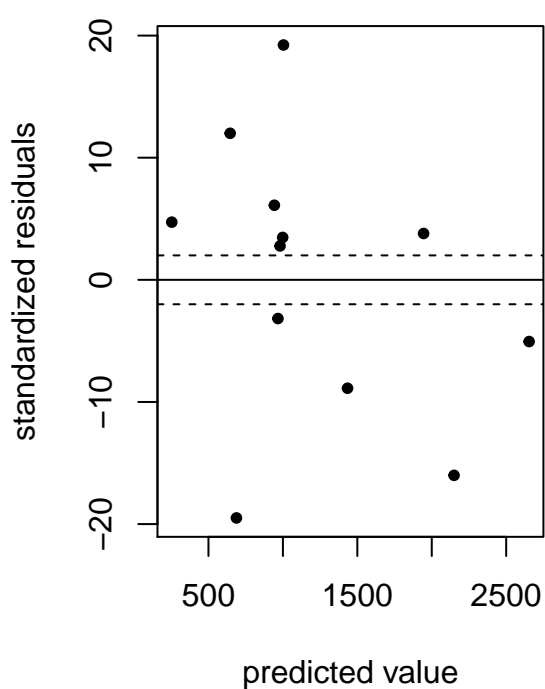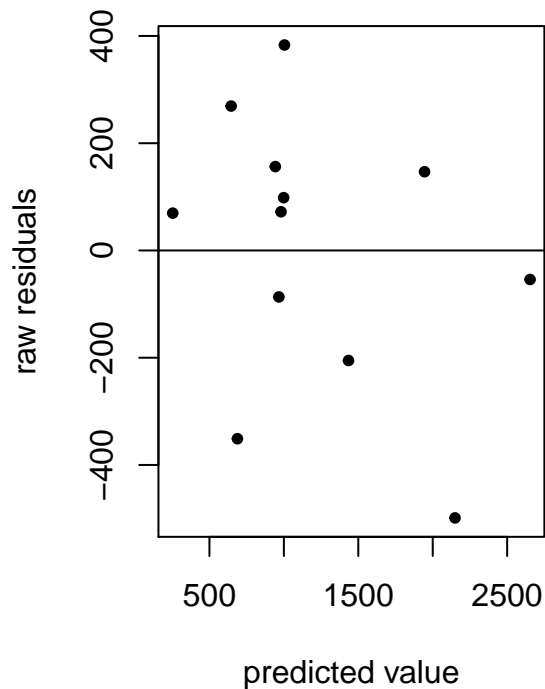

Run Quasi Poisson model

```
mod3quasi <- glm(formula = no.drug.arrests.13.18 ~ no.ems.calls.13.18 + proportionNonWhite + Total.Naxol
             Overdose.Death.Total, offset = log(Total.Population.Over18),family = quasipoisson, data = m
```
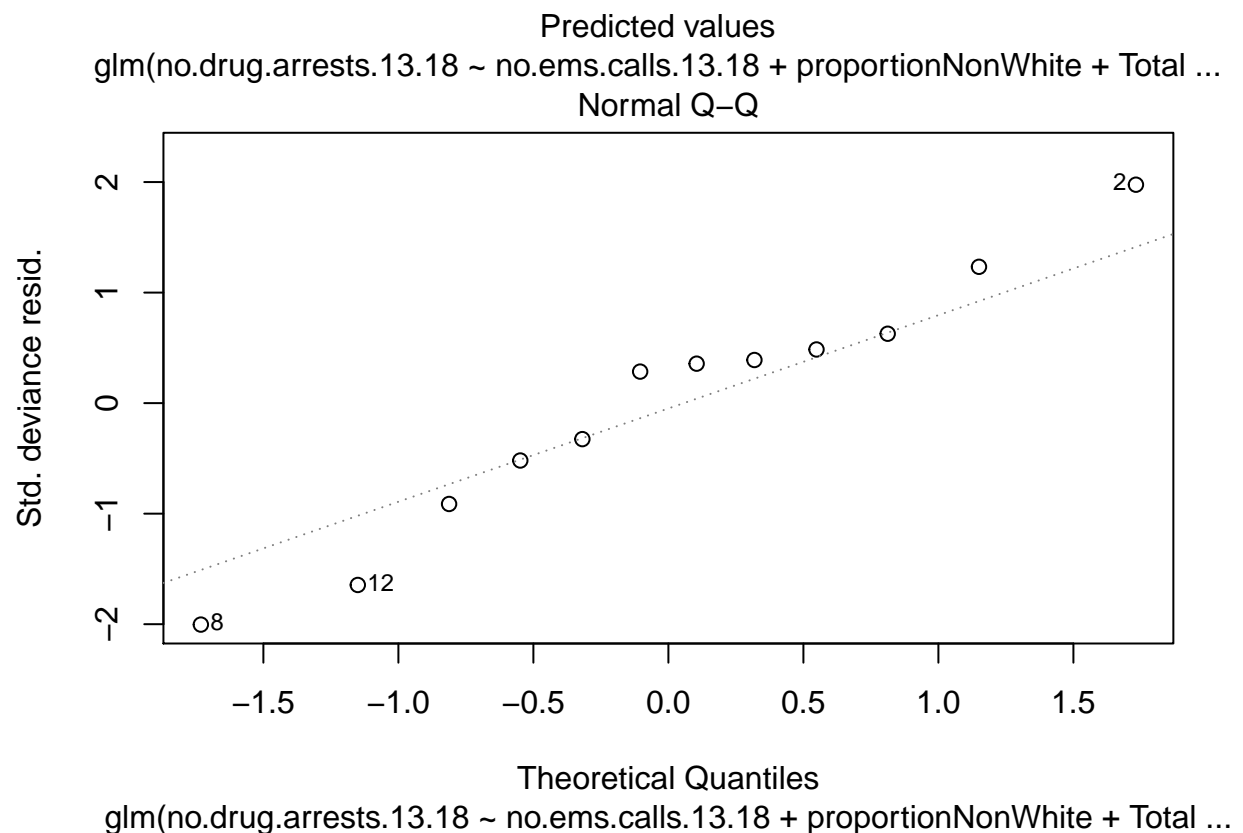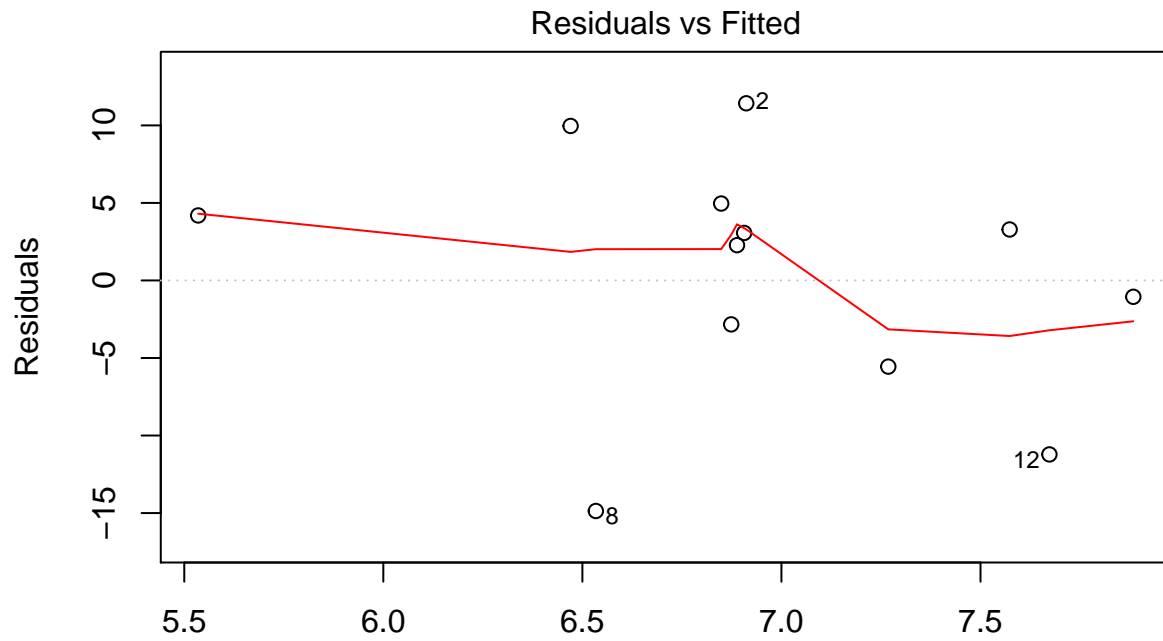
```
arm::display(mod3quasi)
```

```
## glm(formula = no.drug.arrests.13.18 ~ no.ems.calls.13.18 + proportionNonWhite +
##     Total.Naxolone.Saves + Overdose.Death.Total, family = quasipoisson,
##     data = mod3dat, offset = log(Total.Population.Over18))
##                     coef.est coef.se
## (Intercept)         -3.68     0.45
## no.ems.calls.13.18   0.00     0.00
## proportionNonWhite   1.32     0.56
## Total.Naxolone.Saves -0.01    0.01
## Overdose.Death.Total  0.00    0.01
## ---
##   n = 12, k = 5
##   residual deviance = 684.2, null deviance = 4680.1 (difference = 3995.9)
##   overdispersion parameter = 94.7
```
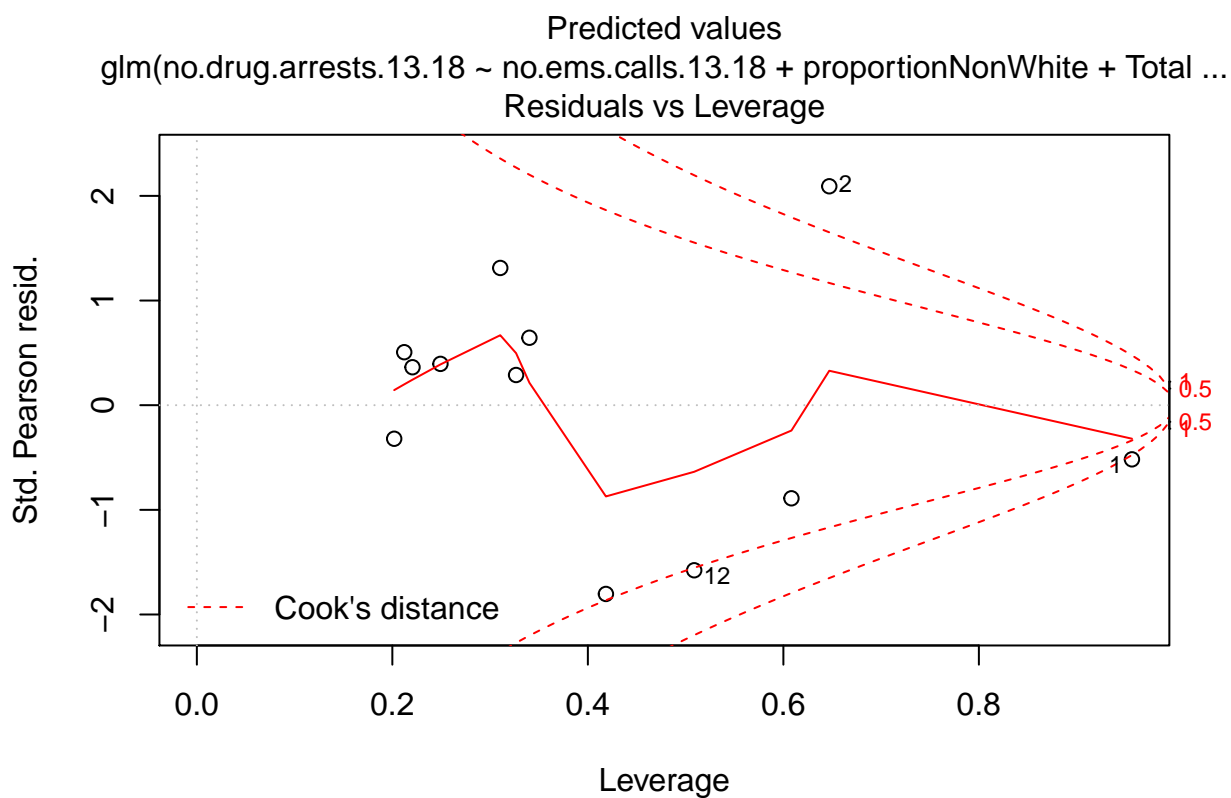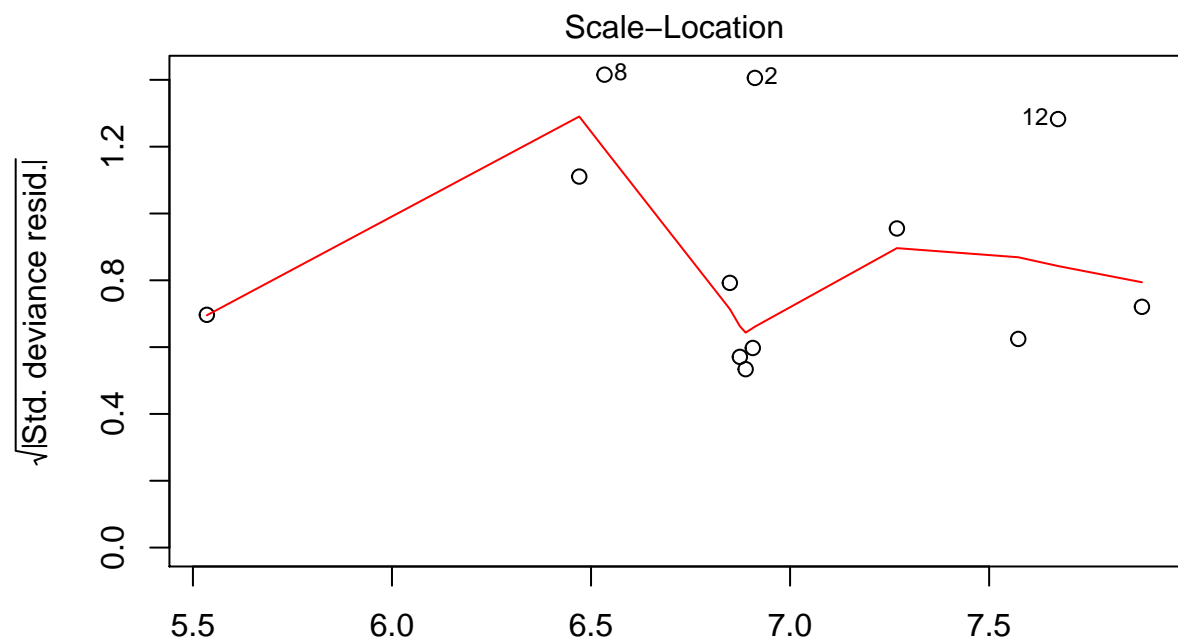
```
summary(mod3quasi)
```

```
##
## Call:
## glm(formula = no.drug.arrests.13.18 ~ no.ems.calls.13.18 + proportionNonWhite +
##     Total.Naxolone.Saves + Overdose.Death.Total, family = quasipoisson,
##     data = mod3dat, offset = log(Total.Population.Over18))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -14.866   -3.513    2.671    4.384   11.424
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -3.6801465  0.4472790  -8.228 7.62e-05 ***
## no.ems.calls.13.18    0.0004936  0.0002964   1.665   0.1398
## proportionNonWhite    1.3203276  0.5628233   2.346   0.0514 .
## Total.Naxolone.Saves -0.0071610  0.0058394  -1.226   0.2597
## Overdose.Death.Total -0.0045079  0.0123699  -0.364   0.7263
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 94.66543)
##
##     Null deviance: 4680.12  on 11  degrees of freedom
## Residual deviance:  684.22  on 7  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

```
plot(mod3quasi)
```

Residuals vs Fitted

Predicted values
glm(no.drug.arrests.13.18 ~ no.ems.calls.13.18 + proportionNonWhite + Total ...

Normal Q–Q

Theoretical Quantiles
glm(no.drug.arrests.13.18 ~ no.ems.calls.13.18 + proportionNonWhite + Total ...

Scale−Location

Predicted values
glm(no.drug.arrests.13.18 ~ no.ems.calls.13.18 + proportionNonWhite + Total ...

Residuals vs Leverage

Cook's distance

Leverage
glm(no.drug.arrests.13.18 ~ no.ems.calls.13.18 + proportionNonWhite + Total ...

```
ci(mod3quasi)
```

```
##                       Estimate     CI lower     CI upper    Std. Error
## (Intercept)         -3.6801465165 -4.7377933479 -2.622499685 0.4472790257
## no.ems.calls.13.18   0.0004936447 -0.0002072411  0.001194530 0.0002964047
## proportionNonWhite   1.3203275524 -0.0105380657  2.651193171 0.5628232973
```

```
## Total.Naxolone.Saves -0.0071610320 -0.0209690146  0.006646951 0.0058393982
## Overdose.Death.Total -0.0045078812 -0.0337580700  0.024742308 0.0123699098
##                            p-value
## (Intercept)          7.615307e-05
## no.ems.calls.13.18   1.397676e-01
## proportionNonWhite   5.139663e-02
## Total.Naxolone.Saves 2.597217e-01
## Overdose.Death.Total 7.263018e-01
```

- When you add the extra variables (naxolone saves, overdose deaths) and then account for overdispersion with the quaispoisson model, the poroportionNonWhite variable is still positive, but remains significant only at the .1 signficance level. The other variables (no.ems.calls.13.18, Total.Naxolone.Saves, Overdose.Death.Total - do not remain significant at all).

```
# stargazer::stargazer(mod1quasi, mod2quasi, mod3quasi, title = "Poisson Regression Results", no.space
#              dep.var.labels = c("Drug Arrests (2013-2018)", "Drug Arrests (2014-2017)",
#                                 "Drug Arrests (2013-2018)"),
#              column.labels = c("Aggregated", "", "Aggregated"),
#              covariate.labels = c("EMS Calls 2013-2018", "EMS Calls 2014-2017",
#                                   "Proportion Non-White", "Naloxone Saves",
#                                   "Drug Overdose Deaths",
#                                   "Constant"))
```

Table 1: Poisson Regression Results

	Dependent variable:		
	Drug Arrests (2013-2018) Aggregated	Drug Arrests (2014-2017)	Drug Arrests (2013-2018) Aggregated
	(1)	(2)	(3)
EMS Calls 2013-2018	0.0001		0.0005
	(0.0001)		(0.0003)
EMS Calls 2014-2017		0.001**	
		(0.0003)	
Proportion Non-White	2.070***	2.126***	1.320*
	(0.467)	(0.297)	(0.563)
Naloxone Saves			−0.007
			(0.006)
Drug Overdose Deaths			−0.005
			(0.012)
Constant	−4.369***	−6.073***	−3.680***
	(0.251)	(0.160)	(0.447)
Observations	12	48	12

Note: *p<0.1; **p<0.05; ***p<0.01