regressionmodelsforthesis

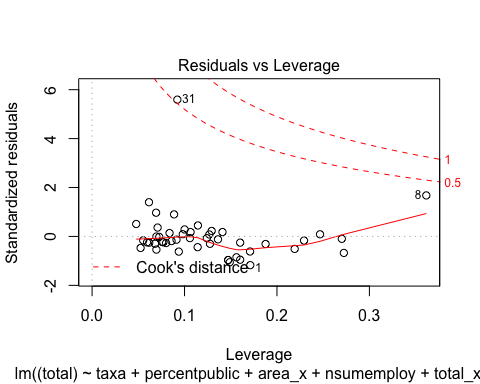
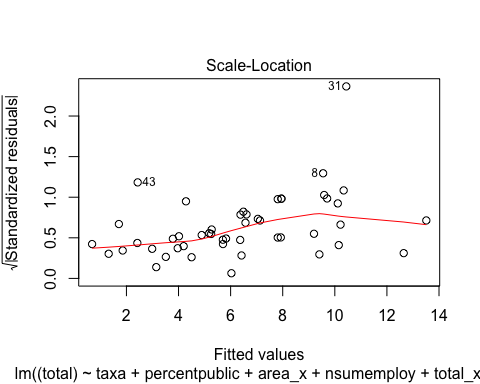
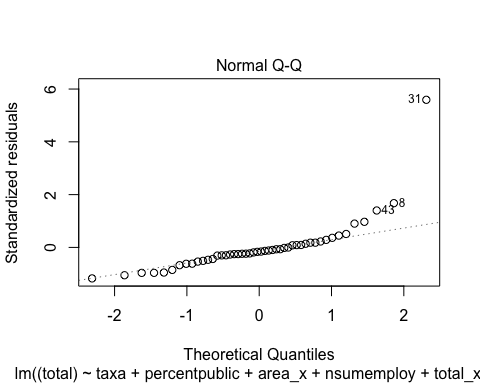
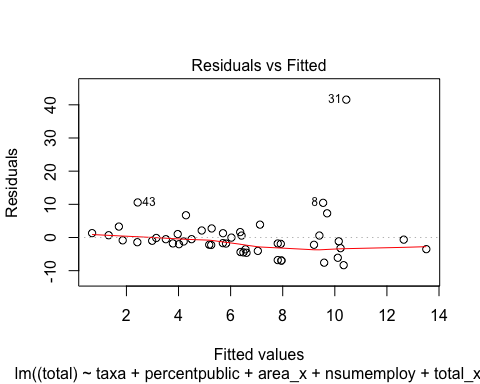
Annabelle

April 22, 2020

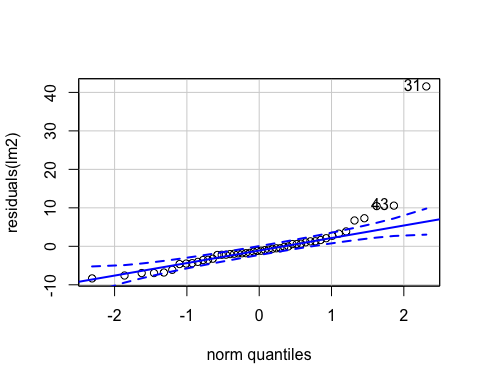
lm2 <- lm(log(total) ~ taxa + percentpublic + area\_x + nsumemploy + total\_x\_x, data=RegData)  
#withou log   
lm2 <- lm((total) ~ taxa + percentpublic + area\_x + nsumemploy + total\_x\_x, data=RegData)  
  
summary(lm2)

##   
## Call:  
## lm(formula = (total) ~ taxa + percentpublic + area\_x + nsumemploy +   
## total\_x\_x, data = RegData)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.340 -3.293 -1.174 1.097 41.560   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.591e+00 3.468e+00 1.900 0.0643 .  
## taxa -2.472e+00 2.550e+00 -0.969 0.3380   
## percentpublic -5.350e+00 4.307e+00 -1.242 0.2210   
## area\_x 2.754e-11 3.475e-11 0.793 0.4325   
## nsumemploy 1.570e-05 2.971e-05 0.528 0.6001   
## total\_x\_x 1.316e+00 9.135e-01 1.440 0.1572   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.799 on 42 degrees of freedom  
## (3 observations deleted due to missingness)  
## Multiple R-squared: 0.143, Adjusted R-squared: 0.04101   
## F-statistic: 1.402 on 5 and 42 DF, p-value: 0.2433

#kable(table(lm2$coefficients))  
  
par(mfrow = c(1, 1))  
plot(lm2)



qqPlot(residuals(lm2))



## 31 43   
## 30 41

par(mfrow = c(1, 1))  
  
#plot(predicted(lm2), residuals(lm2))  
#hist(residuals(lm2))  
# how add CI equivalent around q-q plot?   
AIC(lm2)

## [1] 340.9962

#check\_overdispersion(lm2)

Changed area units

Conclusion: same model if modify area units - literally no changes

Now, keeping area units changed - move into glmmTMB

gaus <- modRegData  
gaussian <- glmmTMB(log(total) ~ taxa + percentpublic + total\_x\_x + narea + nsumemploy, data=gaus, family=gaussian())  
summary(gaussian)  
  
par(mfrow = c(2, 2))  
#plot(predicted(MODELNAME), residuals(MODEL))  
par(mfrow = c(1, 1))  
  
qqplot(residuals(gaussian))  
  
qqnorm(residuals(gaussian), ylab="Residuals")  
qqline(residuals(gaussian)) #qqline won't print  
hist(residuals(gaussian))  
par(mfrow = c(1, 1))  
  
AIC(gaussian)  
#Dispersion estimate for gaussian family (sigma^2): 0.574

Conclusion - I also don’t think this model is doing anything differently (beside package used to run)

Now, try possion  
Have issue with convergence as soon as add + nsumemploy when leave this variable out … Or if change area from modRegData units

pois <- modRegData  
#if change to RegData (no area conversation.. )  
fit\_genpoisson <- glmmTMB(log(total) ~ taxa + percentpublic + total\_x\_x + narea, data=pois, family=genpois(link = "log"))

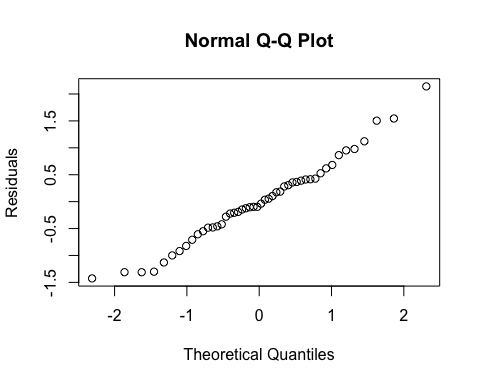
## Warning in glmmTMB(log(total) ~ taxa + percentpublic + total\_x\_x + narea, :  
## non-integer counts in a genpois model

## Warning in (function (start, objective, gradient = NULL, hessian = NULL, :  
## NA/NaN function evaluation

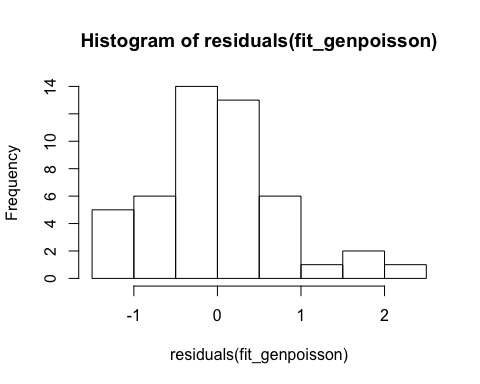
# issue : non-integer counts in a genpois modelNA/NaN function evaluation   
  
summary(fit\_genpoisson)

## Family: genpois ( log )  
## Formula: log(total) ~ taxa + percentpublic + total\_x\_x + narea  
## Data: pois  
##   
## AIC BIC logLik deviance df.resid   
## 123.3 134.5 -55.6 111.3 42   
##   
##   
## Overdispersion parameter for genpois family (): 0.469   
##   
## Conditional model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.113883 0.265412 0.429 0.6679   
## taxa -0.150748 0.163536 -0.922 0.3566   
## percentpublic -0.217155 0.293102 -0.741 0.4588   
## total\_x\_x 0.156199 0.065736 2.376 0.0175 \*  
## narea 0.002497 0.002424 1.030 0.3028   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

par(mfrow = c(2, 2))  
#plot(predicted(MODELNAME), residuals(MODEL))  
par(mfrow = c(1, 1))  
  
#qqplot(residuals(gaussian))  
  
qqnorm(residuals(fit\_genpoisson), ylab="Residuals")



#qqline(residuals(fit\_genpoisson)) #qqline won't print  
hist(residuals(fit\_genpoisson))



par(mfrow = c(1, 1))  
  
AIC(fit\_genpoisson) #much much higher before took the log - 123.25

## [1] 123.25

#Dispersion parameter for genpois family (): 0.469   
r2(fit\_genpoisson)

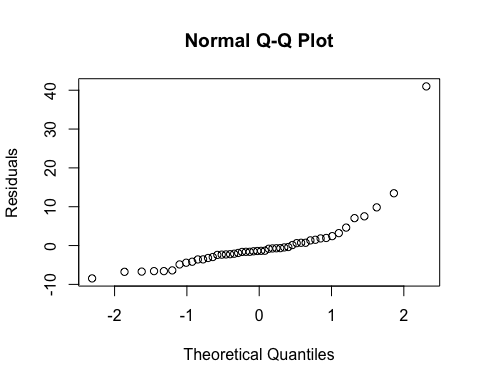
## [1] NA

t-poisson unable to take the log of total because “Error in glmmTMB(log(total) ~ taxa + percentpublic + total\_x\_x + narea, : ‘log(total)’ contains zeros (or values close to zero). Zeros are compatible with a truncated distribution only when zero-inflation is added”

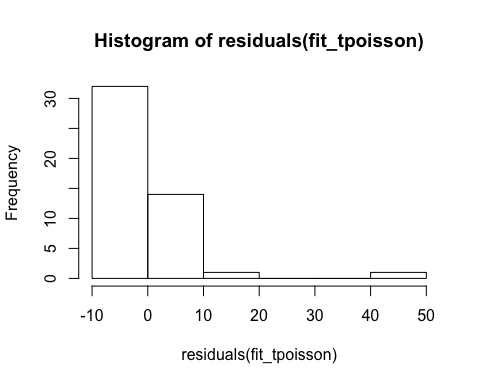
pois <- modRegData  
  
fit\_tpoisson <- glmmTMB(total ~ taxa + percentpublic + total\_x\_x +narea, data=pois, family=truncated\_poisson(link = "log"))  
  
summary(fit\_tpoisson)

## Family: truncated\_poisson ( log )  
## Formula: total ~ taxa + percentpublic + total\_x\_x + narea  
## Data: pois  
##   
## AIC BIC logLik deviance df.resid   
## 381.7 391.1 -185.9 371.7 43   
##   
##   
## Conditional model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.815624 0.169456 10.714 < 2e-16 \*\*\*  
## taxa -0.315542 0.122328 -2.579 0.0099 \*\*   
## percentpublic -0.854538 0.206259 -4.143 3.43e-05 \*\*\*  
## total\_x\_x 0.213358 0.043675 4.885 1.03e-06 \*\*\*  
## narea 0.003404 0.001564 2.176 0.0295 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

par(mfrow = c(2, 2))  
#plot(predicted(MODELNAME), residuals(MODEL))  
par(mfrow = c(1, 1))  
  
#qqplot(residuals(fit\_tpoisson))  
  
qqnorm(residuals(fit\_tpoisson), ylab="Residuals")



#qqline(residuals(fit\_tpoisson)) #qqline won't print  
hist(residuals(fit\_tpoisson))



par(mfrow = c(1, 1))  
  
AIC(fit\_tpoisson) #much much higher because can't take the log = 381.7

## [1] 381.7343

check\_overdispersion(fit\_tpoisson)

## # Overdispersion test  
##   
## dispersion ratio = 6.987  
## Pearson's Chi-Squared = 300.462  
## p-value = < 0.001

## Overdispersion detected.

r2(fit\_tpoisson)

## [1] NA