STATS 330 Mid-semester Test

Your name and ID here

Due Date: 7:15pm, Thursday 16th September

Question 1:

The histogram clearly shows the variable is right-skewed. A log transformation would make this variable approximately normal and therefore better for linear modelling.

The GAM plots reaffirm this, the plot has less 'waves'.

Question 2:

We can not do a deviance check because we are modelling very sparse data.

However, we can compare deviance, the deviance in the neg-bin model is much smaller and thus the model is more appropriate.

The negative binomial model has a smaller AIC with a difference of much greater than 10, suggesting the second model is more appropriate.

Deviance and AIC agree is the negative binomial model is better.

Question 3:

Banding suggests the we are modelling very sparse data. Our response variable is "the number of articles submitted in this time period", it is likely that this data contains many values below 5. Some PHD graduates could submit very few journals. This leads to sparse data and banding.

The quantile residuals suggest the negative binomial is a better fit. This is because if the model is appropriate quantile residuals are approximately standard normal distribution with zero mean and constant variance 1.

This means it is very unlikely we will have values outside the interval [-2,2] and even less likely outside the interval [-3,3]. The poisson model has values above 4 and and less than -4, the negative binomial does not, suggestion the negative binomial model is more appropriate.

Question 4:

We didn't study t-values, ignore this.

Question 5:

We can use the formula and plug in the relevant values

$log(\mu_1) = \beta_0 \times fem + \beta_1 \times log(ment + 1) + \beta_2 \times phd + \beta_3 \times mar + \beta_4 \times kid5$

```
#Plugging in the values
fem <- 0
ment <- 8
phd <- 1.38
mar <- 1
kid5 <- 2
art <- 3
solution <- -0.1138 + -0.2254*fem + 0.3616*(log(ment+1)) + -0.0109*phd + 0.1768*mar + -0.1731*kid5

#We get our solution which is consistent with the question

solution
```

```
## [1] 0.4962744
```

Question 6 :

Remembering the formula and variance calclator for negative binomoal models My calcs are below

```
fittedVal <- 3
expectedVal <- exp(solution)
theta <- 2.4012

#In poisson variance is equal to mean, so

variance <- expectedVal + (expectedVal^2)/theta

resid <- (fittedVal - expectedVal)/(sqrt(variance))

resid
```

```
## [1] 0.8161429
```

Question 7:

Add an interaction term with the variables fem and kid5, and examine p values of the term fem:kid5. If it is significant this suggests having a kid under 5 may have a more adverse effect on one of the genders.

Code : art ~ .. fem*kid5.. (.. refers to the other variables.) # Question 8:

Our Goal was to examine and explain what determines the number of articles submitted by a cohort of recent PHD graduates.

Firstly the prestige of the PHD programme has no effect on the number of articles submitted.

Being female reduces the expected number of articles published by between 8.1% and 30.7%/

A 1 unit increase in the logged number of articles submitted by a mentor plus 1 will increase the expected number of articles by between 33.3% and 54.8%

If an individual is married the expected number of articles increases by between 1.8% and 40%

If an individual has children unde rthe age of 5, the expected number of articles submitted decreases by etween 6.8% and 24.2%

insert text/calculations here