STATS 330 Mid-semester Test

Your name and ID here

Due Date: 5pm, Thursday 7th May 2020

## Question 1

For the 24 hour duration of this test, I confirm that I will not discuss the content of the test with anyone else. I will not give any assistance to another student taking this test. I will not receive any assistance from any person or tutoring service.

Hasnain Cheena

## Question 2

Where of the ith doctor replied Yes he/she smokes and if the ith doctor replied No he/she does not smoke. Furthermore, is the mid-point of a given age group in years and is the log transformed total number of person-years of observation for a given group. Moreover, follows a Poisson distribution with mean of and represents the number of deaths in the ith age group.

1. Where of the ith doctor replied Yes he/she smokes and if the ith doctor replied No he/she does not smoke. Furthermore, is the mid-point of a given age group in years. Moreover, follows a Binomial distribution with number of trials and probability of success of .
2. In the variable is being used as an offset. This is because we know the effect of years on the response variable deaths and as such do not need to estimate it. Generally as someone gets older, their chance of dying increases. This effect is displayed as years decreases and the number of deaths increases.The offset is incorporated into the model by fitting the log of years.

## Question 3

where if the insect is a male and if the insect is a female. Moreover, is the age in years of the ith insect and is the weight in grams of the ith insect. Furthermore, follows a Binomial distribution with number of trials and probability of success of .

1. We have evidence () to suggest that the sex of an insect had an effect on whether or not it became infected. We estimate that the odds of insects that are male becoming infected is 21.6% to 94.9% less than insects that are female, for the same age and weight.
2. Looking purely at the GAM plots I agree with her decision to fit age with a quadratic effect but not weight. This is because looking at the GAM plot for weight a straight line can easily be fit in between the dashed lines. Therefore it was sensible not to fit a quadratic term for weight at this stage. In comparision, looking at the GAM plot for age while a straight line can be fit there is reasonable amount of curvature hinting at a non-linear (quadratic) relationship between age and the response.
3. I would recommend over . This is because even though both models are appropriate (Pearson, deviance and randomised quantile residual plots are patternless bands around 0) and good fits (from the deviance we evidence suggesting the models are good fits to the dataset) the quadratic term in is significant (at a 5% level) and cannot be ignored. This implies a non-linear relationship between the response and age which is captured by . Furthermore, the deviance of is lower than implying it is a better fit.
4. The null hypothesis being tested is that there is no interaction between the any of the explanatory variables (age, weight or sex).
5. We have no evidence (p-value 0.39) to suggest that there any interactions between age, weight and sex.

## Question 4

$

1. Residuals for

ii) Residuals for

C) I would plot and examine the pearson, deviance and randomised quantile plots of to assess if the model is appropriate. If the residual plots are fine, I would then use the deviance to check the goodness of fit. If was a good fit () I would then use information-theory criteria (AIC, AICc and/or BIC) to compare both models.If showed lower AIC, AICc and/or BIC scores (and all other conditions above were met) I wold then consider it a suitable replacement for .