**Solutions: ST2304 Exercises Week 12 - Log-linear Models**

**All answers are marked in red**

**Question 1**



**Fig. 1.** Count data of rare species of Class II where the shaded grey area corresponds to era A (era under suspicion) for the three locations Sussex (red), Hastings (blue) and Kent (brown).

1. For Class II (the intermediate rarities), fit the same model as I did for Class I. Is Hastings any difference between eras A and B? If so, what is the difference?

We should fit the same linear model as for class I and look at the summary table and parameter estimates:



By assessing the parameter estimates from the linear regression we could see that there were significant differences between the locations but the interaction describing whether there was a difference between areas at the hastings location was not signficant. For counts class II at Hastings during Era A the linear predictor for the log of the Poisson expectation $λ$ is given by

ln($λ$) = α + βhastings + βeraA + βareahastings:eraA.

And if we put in the estimates, an estimate of $λ$ = EY, since Y~pois($λ$), is

$\hat{λ}$= exp(-0.83 + 1.37 + 0.21 + 0.54) = 3.53 individuals

1. Is there any evidence for overdispersion? If there is, what effect does correcting for it have on the

model?

Overdispersion is where there is greater variation in the data than is expected by the statistical model that we use. In a Poisson model, the variance is the response should be equal to the mean and we can test for overdispersion by asking whether the model deviance is greater than would be expected base don the residual degrees of freedom. By this approach we get an overdispersion value of 1.42, which is not too much but we could still adjust the model accordingly but this only slightly increases the standard errors.

1. Plot the residuals against year. Does there seem to be any extra effect of year?

There seems to be some trend over time in the residuals and similar to class I indicates and increase towards the end of the time period. Therefore the maybe be a temporal effect that influences all classes of species that we have not accounted for.

*Now for Class III:*

1. For Class III (the more common rarities), fit the same model as I did for Class I. Is Hastings any

different between eras A and B? If so, what is the difference?

By re-fitting the Poisson model to the subsetted data for Class III only we can see counts are generally larger in Hastings than the intercept which corresponds to Sussex (βhastings = exp(1.26) = 3.53), however, the interaction describing the addition effect of Era A in Hastings is highly non-significant (p = 0.88). Also the difference between Eras A and B, not account for location, is also not significant for this class (p=0.14).

1. Is there any evidence for overdispersion? If there is, what effect does correcting for it have on the

model?

The estimate of overdispersion is 2.1, which is still not too high however, it does noticeably increase the standard errors when we account for it.

1. Plot the residuals against year. Does there seem to be any extra effect of year?

Again there still appears to be an increasing trend towards the end of the time series which could be account for by adding a year effect.

1. Is there any difference between the estimated coefficients for the AreaHastings:EraA interactions between the three different classes?

For class I the interaction was 1.7405, for class II it was 0.5427 and for class II the interaction parameter was 0.03763, although all three did not have significant p-values the differences between eras seems to have been greater in Hastings for rare classes than for more common ones.

**Exam-type Questions**

1. From the output of the analysis of Class II, does the Analysis of Deviance table suggest there is variation between areas and eras?

Yes both area and era, as well as their interaction effects, are considered to explain significant variance in the counts.

1. What statistics would you use to investigate if there is over-dispersion, and how strong it is? What additional information would you need?

We can assess whether the variance in the response is equal to the mean, by dividing the residual deviance by the residual degrees of freedom, which is an assumption in Poisson statistical model.

1. From the summary, for Class II, what are the main differences between eras and areas?

Counts are generally higher in Hastings (β=1.36) and also in Kent but to a lesser degree (β=0.76) compared to sussex which was the intercept in this model (intercept = -0.83). There was only a significant difference between eras in Kent, where era A had lower counts (interaction = -1.14).

1. The data were collected to see if there was evidence of fraud, which we would see if there were more rarities in Era A in the Hastings area. Does the analysis from Class II suggest that there could be fraud?

There is a generally tendency for high counts in Hastings for the rare classes, which would indicate to this effect, however, since this interaction term is not significant we cannot conclude this for sure.

**Question 2: Himmicanes**

1. How would you write the model? (if it’s easier, write the model as the interaction of Gender with Minpressure, plus the interaction of Gender and NDAM).

Y ~ Pois(α + Gender\*(Minpressure + NDAM))

1. Fit the model assuming a Poisson distribution. Does it suggest an effect of gender?

Gender as an additive is close to significant (p=0.07), however the interaction between gender and normalised damage is highly significant. Indicating that gender appears to influence how much the hurricane cost.

1. Is there any evidence for over-dispersion?

This model is highly overdispersed, where the overdispersion parameter is 28.21.

1. Correct the overdispersion using a negative binomial distribution (this is what was done in the paper), using the glm.nb() function in the MASS package. Now look at the residuals: plot them against the predicted values and the covariates. Does the model look OK, or can it be improved?

We compare the residuals from (a) the normal poisson glm model and (b) the model correct for overdispersion using glm.nb.

1. Not correcting for overdispersion:



1. Correcting for overdispersion:



The inclusion of a parameter to estimate overdispersion has not hugely improved the model fit, particularly at high fitted values the residuals are far below the mean.

1. Based on this, try to find a model that fits better. Does this model show an effect of gender?

Quasipoisson models can also used to estimate a scale parameter so that the variance is proportion to the mean rather than equal to the mean as in a standard poisson model. By fitting a quasipoisson model like this:

glm(alldeaths ~ Gender\*(NDAM+Minpressure), family=quasipoisson, data=Data)

our residuals then look like this, which is still not great, however gender and its interaction with NDAM are no longer significant.

****