

test_sl_2019

Question 1)

a)

$$\log(\text{Snapper}_i) = \beta_0 + \beta_1 \times \text{Hahei}_i + \beta_2 \times \text{TAWH}_i + \beta_3 \times \text{Season}_i + \beta_4 \times \text{Year}_i$$

Where Snapper_i refers to the number of snapper for the i th observation. TAWH refers to the if the i th observation marine reserve was located in Tawharanui, 1 if yes, 0 if not. Hahei_i refers to if the i th observation marine reserve was located in Hahei, 1 if yes, 0 if not Season_i refers to the i th observations season, 1 if Spring, 0 if Autumn Year_i refers to the i th observations year, 1 if 98-99, 0 if 97-98.

b)

```
exp(-0.928897)
```

```
## [1] 0.3949891
```

```
exp(-0.66690)
```

```
## [1] 0.5132973
```

If season is Spring the expected number of Snapper is between 0.3949891 and 0.5132973 times the expected number of Snapper in the same reserve and year but in Autumn.

Question 2)

a)

H0: The sub model is appropriate, the interaction terms can be removed from the model

b)

We can reject the null hypothesis; the sub model is NOT appropriate and it is very likely there is at least one interaction between variables. ## Question 3)

a)

We have very sparse data, there exist μ_i that are smaller than 5. Therefore we should not rely on chi-squared tests. But we can compare deviances.

The deviance in `pois2.fit` is significantly reduced relative to `pois1.fit`. This suggests `pois1.fit` is inappropriate and we should use `pois2.fit` as our model.

b)

It is unclear if `pois2fit` is appropriate because we do not know if fitted values exceed the value 5. This especially given there exist observations less than 5.

Question 4)

I prefer the negative binomial fit. The quasi poisson deviance residuals firstly show a non-constant variance. The first band's variance is smaller than the second and third bands. Secondly the deviance residual should be approximately standard normal $N(0,1)$ and values should be inside the interval $[-3,3]$. This is false for quasi poisson, we have values above 4 and below -4. All this suggests the model is not appropriate.

Negative binomial quantile residuals are approximately standard normal distributed with mean zero and constant variance, all values are also inside the interval $[-3,3]$. All suggesting the model is appropriate.

Question 5)

a)

i) Poisson Fit

$$Var_2 = \mu_2$$

This indicates $Var_2 = 1.8527$

ii) Quasi-Poisson

$$Var_2 = k \times \mu_2$$

Our K values = 3.325656

We can calculate Var_2 below

```
k <- 3.245656
mu2Quasi <- 1.8527
var2Quasi <- k*mu2Quasi
var2Quasi
```

```
## [1] 6.013227
```

$$Var_2 = 6.013227$$

iii) Negative_Binomial Fit

$$Var_2 = \mu_2 + \frac{\mu_2^2}{\theta}$$

We do the calculations below

```
mu2NegBin <- 1.5321
theta <- 1.5522

var2NegBin <- mu2NegBin + (mu2NegBin^2)/theta

var2NegBin
```

```
## [1] 3.04436
```

We have $Var_2 = 3.04436$

b) Pearson Residuals

$$r_i = \frac{y_i - \hat{E}(Y_i)}{\sqrt{\hat{Var}(Y_i)}}$$

i) Poisson Pearson Residual

```
y2 <- 4
mu2 <- 1.8527
var2 <- mu2

r2 <- (y2-mu2)/(sqrt(var2))

r2
```

```
## [1] 1.577576
```

Pearson Residual for Poisson is $r_2 = 1.414424$

ii) Negative Binomial Pearson Residual

```
y2 <- 4
mu2 <- 1.5321
var2 <- var2NegBin

r2 <- (y2-mu2)/(sqrt(mu2))

r2
```

```
## [1] 1.993811
```

Negative Binomial Pearson Residual is $r_2 = 1.993811$

Question 6)

We could use dredge complex model with the formula *Snapper Reserve * Season * Year* We can do this in R by first fitting a complex model `fit <- glm(Snapper ~ Reserve * Season * Year, ...)` and putting this model inside the dredge function in R

This should work given there are not too many options for R to fit. This process would select the model with the best AICc.

In the unlikely case this fails we could use backward step-wise selection. we can do this in R with the same complex model we fit before (`fit <- glm()`). Then plug in the code “`step(fit, direction = "backward")`“. Then R will go through each model, removing a variable each time till it thinks it has minimised AIC. R will select the best model for us.

Question 7)

Firstly We should fit an interaction with Reserves and Spring. In R formula this would be:

*Snapper ~ Reserve * Spring*

Then we would also include the year variable, which would result in the model

*Snapper ~ Reserve * Spring + Year*