

## test\_s1\_2021

### Question 1)

For the 75 minute 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.

### Question 2)

a)

$$\log(\text{death}_i) = \beta_0 + \beta_1 \times \text{Gender}_i + \beta_2 \times \text{Day}_i + \beta_3 \times \text{Gender}_i \times \text{Day}_i + \log(\text{person.days}_i)$$

where  $\text{Gender}_i$  refers to the gender of the victim, 1 if Male, 0 if female AND  $\text{Day}_i$  is the Friday date for the  $i$ th day, 1 if the day is NOT Friday the 13th, 0 if Friday the 13th  $\text{person.days}_i$  is the total number of person-days over which the deaths occurred for the  $i$ th group

b)

Person.days is an offset. It is essentially transforming the response variable into

$$\log\left(\frac{\text{death}_i}{\text{person.days}_i}\right)$$

So we have logged expected number of deaths per person.days as our response.

c)

i)

In the model our baseline is Female and Friday

i) let  $\text{dayOtherFri} = 0$ ,  $\text{genderMale} = 1$

```
exp(0.7725)
```

```
## [1] 2.165172
```

On a Friday 13th the death rate for males is about 2.165 times that for females

ii) Let Male = 0, Other Fri = 1

```
(100*(exp(-0.4789)-1))
```

```
## [1] -38.05356
```

For females, the death rate for other Fridays is about 62% percent of the death rate for Friday 13th.

d)

The deviance of the Fri13.fit is 0 because it is the saturated model. We know this because this data set has 4 observations and the model has 4 parameters

#e)

Day does not have the same effect for male and female. For females we see the death rate for other Fridays is about 62% percent of the death rate for Friday the 13th

In males we see the rate for other Fridays is about 95% of the death rate for Friday the 13th (calculations below). The effect is the addition of the  $\beta_1 + \beta_2$

This difference is evident in the use of the interaction term: GenderMale:DayOtherFri

```
100+(100*(exp(-0.4789+0.4282)-1))
```

```
## [1] 95.05638
```

### Question 3)

a)

i) Variance for Poisson.fit using formula  $Var(Y_4) = 30.32639 = \mu_4$

ii) For quasi-poisson  $Var(Y_4) = k \times \mu_4 = 20.63818 \times 30.32639$

Variance is equal to 625.888 (Calc Below)

```
20.63818*30.32639
```

```
## [1] 625.8815
```

iii) for Negative Binomial

$$Var(Y_4) = \mu_4 + \frac{\mu_4^2}{\theta}$$

```
mu4 <- 25.69835  
theta <- (5.7626)
```

```
varianceNegBin <- mu4 + ((mu4^2)/theta)  
varianceNegBin
```

```
## [1] 140.3003
```

variance is equal to 140.3

b)

i)

Poisson first

```
observed <- 19
expected <- 30.32639
variance <- 30.32639

#Raw residual
rawPoisson <- observed - expected
rawPoisson
```

```
## [1] -11.32639
```

```
#pearson
pearsonPoisson <- (observed - expected)/sqrt(variance)

pearsonPoisson
```

```
## [1] -2.056748
```

For poisson regression we have raw residual = -11.32639

AND the pearson residual = -2.056748

Now Negative Binomial

```
observed <- 19
expected <- 25.69835
variance <- varianceNegBin

rawNegBin <- observed - expected
rawNegBin
```

```
## [1] -6.69835
```

```
pearsonNegBin <- (observed - expected)/sqrt(variance)
pearsonNegBin
```

```
## [1] -0.5655077
```

For negative binomial regression we have a raw residual of -6.69835 AND a pearson residual of -0.5655077.

c)

I prefer the negative binomial model

AIC is much lower for the negative binomial fit, there is a difference of

```
## [1] 521.71
```

There is nothing to suggest we have particularly sparse data. Therefore we can run chi-squared tests on the deviance residuals. Chi-Sq test suggest the poisson regression is not suitable.

They also suggest that the negative binomial regression is suitable.

Finally the pearson and quantile residuals are not approximately standard normal distribution, there are values outside the interval  $[-5,5]$ . This reaffirms the poisson regression is not suitable.

The negative binomial pearson and quantile residual plots are approximately standard normal distribution, suggesting the model is appropriate.

Deviance, AIC and residual plots all agree that the negative binomial model is better.