Math 445 HW 3

Erick Castillo

2/23/2021

Question 1: Fit a Negative Binomial Regression model with Sa as a response variable and all other variables as predictors:

```
nb.mod1 = glm.nb(Sa~factor(C)+factor(S)+W+Wt)
summary(nb.mod1)
```

```
##
## Call:
## glm.nb(formula = Sa ~ factor(C) + factor(S) + W + Wt, init.theta = 0.9650380207,
##     link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8788  -1.3685  -0.3267   0.4224   2.2288
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.274784   1.950675  -0.141   0.8880
## factor(C)2  -0.320766   0.372716  -0.861   0.3894
## factor(C)3  -0.596232   0.417342  -1.429   0.1531
## factor(C)4  -0.579357   0.466470  -1.242   0.2142
## factor(S)2  -0.242827   0.398357  -0.610   0.5421
## factor(S)3   0.042811   0.248427   0.172   0.8632
## W           -0.002522   0.099678  -0.025   0.9798
## Wt           0.700752   0.356375   1.966   0.0493 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(0.965) family taken to be 1)
##
##     Null deviance: 220.68  on 172  degrees of freedom
## Residual deviance: 196.52  on 165  degrees of freedom
## AIC: 763.32
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  0.965
##           Std. Err.:  0.176
##
##  2 x log-likelihood:  -745.319
```

Question 2: Find the AIC and the BIC of this model, compare with the Poisson model from the prior assignment. Which is the better model?

```
AIC(nb.mod1)
```

```
## [1] 763.3188
```

```
BIC(nb.mod1)
```

```
## [1] 791.6984
```

```
poiss.mod2 = glm(Sa~factor(C)+factor(S)+W+Wt, family = 'poisson')
AIC(poiss.mod2)
```

```
## [1] 920.8618
```

```
BIC(poiss.mod2)
```

```
## [1] 946.0881
```

Comparing both the AIC and BIC of each model shows me that the better model to fit the data would be the negative binomial fit.

Question 3: Using the full NB model, test whether S is significant. Write down the null and alternative hypothesis, the test statistic, and the p-value.

```
nb.mod2 = glm.nb(Sa~factor(C)+W+Wt)
anova(nb.mod1, nb.mod2)
```

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: Sa
##                             Model    theta Resid. df    2 x log-lik.   Test
## 1             factor(C) + W + Wt 0.960323       167       -745.9076
## 2 factor(C) + factor(S) + W + Wt 0.965038       165       -745.3188 1 vs 2
##      df  LR stat.   Pr(Chi)
## 1
## 2     2 0.5887885 0.7449827
```

The null hypothesis is $H_0 : \beta_{factor(C)} = 0$ and the alternative hypothesis is $H_A : \beta_{factor(C)} \neq 0$. The test statistic is given to be ≈ 0.5888, while the p-value is 0.74498. In this case, because the p-value is very high, therefore I will fail to reject H_0.

Question 4: Test the hypothesis that the variable C and S are not significant at the same time using the LRT. Once again write down the null and alternative hypotheses, the test statistic, and the p-value.

```
nb.mod3 = glm.nb(Sa~W+Wt)
anova(nb.mod1, nb.mod3)
```

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: Sa
##                            Model     theta Resid. df    2 x log-lik.   Test
## 1                         W + Wt 0.9323858       170       -748.5723
## 2 factor(C) + factor(S) + W + Wt 0.9650380       165       -745.3188 1 vs 2
##      df LR stat.   Pr(Chi)
## 1
## 2     5  3.25357 0.6609573
```

The null hypothesis is $H_0 : \beta_{factor(C)} = \beta_{factor(S)} = 0$ while the alternative hypothesis is H_A : either $\beta_{factor(C)} \neq 0$ or $\beta_{factor(S)} \neq 0$. The test statistic is given by ≈ 3.253 while the p-value is ≈ 0.661. Based on the high p-value, I can conclude that I must fail to reject H_0., that is, both of the set of dummies are not significant in the model.

Question 5: Compute confidence intervals for the coefficients of each predictor.

```
confint(nb.mod1)
```

```
## Waiting for profiling to be done...
```

```
##                  2.5 %    97.5 %
## (Intercept) -4.4667257 3.8836393
## factor(C)2  -1.1041867 0.4043830
## factor(C)3  -1.4458285 0.2092686
## factor(C)4  -1.5341597 0.3489912
## factor(S)2  -0.9960406 0.5377573
## factor(S)3  -0.4507821 0.5196211
## W           -0.2198627 0.2158015
## Wt          -0.0849274 1.5014320
```