Math 445 HW 2

Erick Castillo

2/12/2021

A. Fit the intercept only model. What does this model measure?

```r
null.mod = lm(Sa~1, data = df)
summary(null.mod)
```

```
##
## Call:
## lm(formula = Sa ~ 1, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9191 -2.9191 -0.9191  2.0809 12.0809
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.9191     0.2394    12.2   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.148 on 172 degrees of freedom
```

Response: The intercept only model has $\beta_0 = 2.9191$. This is the mean of the values in the Sa column.

B. Fit a Poisson regression on W. Comment on the results and interpret the parameters.

```r
poiss.mod = glm(Sa~W, family = 'poisson', data = df)
summary(poiss.mod)
```

```
##
## Call:
## glm(formula = Sa ~ W, family = "poisson", data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.8526  -1.9884  -0.4933   1.0970   4.9221
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.30476    0.54224  -6.095  1.1e-09 ***
## W            0.16405    0.01997   8.216  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 632.79  on 172  degrees of freedom
## Residual deviance: 567.88  on 171  degrees of freedom
## AIC: 927.18
##
## Number of Fisher Scoring iterations: 6
```

Response: W has a very small p-value meaning that its presence is significant in the model. Fixing $x = 0$ grants a λ value of $e^{-3.305}$. For every one unit increase in x, the predictor variable has a multiplicative effect of $e^{0.164}$.

C. Compute the predicted count $\hat{y}$ for each entry W of the data. Compute the sum of the residuals squared.

```
yhat =exp(predict(poiss.mod))
sum((yhat-df$Sa)^2)
```

```
## [1] 1537.331
```

Response: The residual sum of squares is given by the 1537.33 up above.

D. Use the given function to predict the new dataset.

```
newdt = data.frame(W=26.3)
predict.glm(poiss.mod, type="response", newdata = newdt)
```

```
##        1
## 2.744581
```

Response: The above code fixes $x_W = 26.3$. When doing so, the output is basically saying, "With a carapace width of 26.3 units, there will be approximately 3 satellites".

E. Fit a Poisson regression model with Sa as response variable and include all other variables as predictors. Make sure to create dummy variables for C and S variables.

```r
full.poiss = glm(Sa~factor(C)+factor(S)+W+Wt, family = 'poisson', data = df)
summary(full.poiss)
```

```
##
## Call:
## glm(formula = Sa ~ factor(C) + factor(S) + W + Wt, family = "poisson",
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.0291  -1.8632  -0.5991   0.9331   4.9449
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.35722    0.96700  -0.369  0.71182
## factor(C)2  -0.26491    0.16811  -1.576  0.11507
## factor(C)3  -0.51374    0.19536  -2.630  0.00855 **
## factor(C)4  -0.53126    0.22692  -2.341  0.01922 *
## factor(S)2  -0.15044    0.21358  -0.704  0.48119
## factor(S)3   0.08742    0.11993   0.729  0.46604
## W            0.01651    0.04894   0.337  0.73582
## Wt           0.49712    0.16628   2.990  0.00279 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 632.79  on 172  degrees of freedom
## Residual deviance: 549.56  on 165  degrees of freedom
## AIC: 920.86
##
## Number of Fisher Scoring iterations: 6
```

Response: Notice that the AIC of this model is 7 points shorter than the model with W as the only predictor. Furthermore, the only significant predictors, based on the significance, are the factor(C)3, factor(C)4, and Wt. Because the AIC is smaller than the only W model, this model seems to be better. Maybe a negative binomial regression might be necessary for a better fit.

F. For your model in e), test the hypothesis that the parameters for C and S are equal to 0. Comment on the results.

```r
red.mod = glm(Sa~W+Wt, family="poisson", data = df)
anova(full.poiss, red.mod, test = 'LRT')
```

```
## Analysis of Deviance Table
##
## Model 1: Sa ~ factor(C) + factor(S) + W + Wt
## Model 2: Sa ~ W + Wt
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       165     549.56
## 2       170     559.89 -5  -10.321  0.06663 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Response: The above code creates a reduced model where the dummies are dropped. Utilizing the anova function on both the full and reduced models with an LRT test, the function produces a p-value of 0.0666. This is significant at 10%, thus in my case I would reject the null hypothesis. This implies neither of the dummies equal to zero.