**Assignment 4 – Poisson Regression**

We are interested in species of birds, plants, and mammals thrive on one of the islands. So no offset will be needed in the Poisson Regression Model.

(a) Using the Galapagos islands data (**Gala**), create plots of the logarithm of the observed counts of total plant species, log(species) versus the potential covariates: area and elevation.

      Then repeat the previous step using the logarithm of each of the covariates in place of the original covariate value.

      Does the Poisson regression model seem a good choice for these data? Why or why not?

(b) Fit a Poisson regression model to the total species count with covariates log (area) and log (elevation).   
     Conduct the Wald's test to determine if the covariates are significant.   
     Examine the deviance and Pearson statistics for this model, and comment on the model's goodness of fit.

(c) Perform the LRT to determine if the addition of the covariates nearest, scruz, and adjacent significantly improves the model's predictive ability.  
     Examine the deviance and Pearson statistics for this model, and comment on the model's goodness of fit.

(d) Compare model (b) and (c), which model is better?

1. Chart, scatter chart

   Description automatically generated

**log(species) Vs area and elevation**

Graph has a slight exponential curve when log(species) is plotted against area and elevation

**log(species) Vs log(area) and log(elevation)**

Graph shows a positive linear correlation, which indicates Poisson regression model seems to be a good choice for the data.

**b. Poisson regression model**

**λ hat (species) = exp (3.81680 + 0.37224 log (area) - 0.11298 log (elevation))**

**lambda hat is the poisson response variable**

**Wald’s Test – Area**

Ho: β1 log(area) = 0

Ha = β1 log(area)! = 0

Zcalc: 22.682

P value: 2e-16

P value < 0.05

Decision: Reject Ho

Conclusion: There is sufficient evidence to conclude that there exists a significant relation between area and total species count at 5% significance level.

**Wald’s Test - Elevation**

Ho: β1 log( elevation) = 0

Ha = β1 log(elevation)! = 0

Zcalc: -2.355

P value: 0.0185

P value < 0.05

Decision: Reject Ho

Conclusion: There is sufficient evidence to conclude that there exists a significant relation between elevation and area at 5% significance level.

**Deviance Statistic Test**

H0: the Poisson regression model fits the data well

Ha: the model does not fit the data well

p-value = 0

p- value < 0.05

Degrees of freedom =27

Decision: Reject Ho

Test statistic = 646.2065\*\*\*\*\*\*(deviance)

Conclusion: We can conclude that the Poisson regression model does not fit the data well at 5% level of significance.

**Pearson Chi-Square Test**

H0: the Poisson regression model fits the data well

Ha: the model does not fit the data well

p-value = 0

p- value < 0.05

Test statistic = 645.1482\*\*\*\*\*\*(deviance)

Decision: Reject Ho

Conclusion: We can conclude that the Poisson regression model does not fit the data well at 5% level of significance.

**c. Likelihood Ratio Test (can test any num of coeff)**

Ho: β nearest = β scruz = β adjacent = 0

Ha: At least one of the coefficients in H0 is different from other.

**Reduced Model**

model. species <- glm (species ~ log(area) + log(elevation), family = poisson, data = data. gala)

**Full Model**

model. species1 <- glm (species ~ log(area) + log(elevation) + nearest + scruz + adjacent, family = poisson, data = data. gala)

Gcalc = 218.7233

P value = 0

Text

Description automatically generated with low confidence

**Decision**: Reject Ho

**Conclusion:** at least one variable provides significant effect on the response variable.

At least one beta is diff from 0

**Deviance Statistic test**

H0: the Poisson regression model fits the data well

Ha: the model does not fit the data well

p-value = 0

p- value < 0.05

Decision: Reject Ho

Test stat = 427.48

Conclusion: We can conclude that the Poisson regression model does not fit the data well at 5% level of significance.

**Pearson Chi-Square Test**

H0: the Poisson regression model fits the data well

Ha: the model does not fit the data well

p-value = 0

p- value < 0.05

test stat – 434.04

Decision: Reject Ho

Conclusion: We can conclude that the Poisson regression model does not fit the data well .at 5% level of significance.

**Compare 2 models**

**Reduced Model**

model. species <- glm (species ~ log(area) + log(elevation), family = poisson, data = data. gala)

**Full Model**

model. species1 <- glm (species ~ log(area) + log(elevation) + nearest + scruz + adjacent, family = poisson, data = data. gala)

**Text

Description automatically generated**

model.species1 has the lowest AIC & BIC. So, model. species1 is the better model.

As per the question, C is a better model than B