Homework 4

Yuanyou Yao

December 9, 2019

Problem 1

(a) Firstly, we fit the model, and check *extra - Psoion variation*.

```
> library(MASS)
> rawdata=read.csv("Galapagos.csv",header = T)
> attach(rawdata)
> lnarea= log(Area)
> lnelev=log(Elev)
> lndistnear = log(DistNear)
> lnareanear = log(AreaNear)
> detach(rawdata)
> glm_1 = glm(Native~lnarea+lnelev+lndistnear+
    lnareanear, data = rawdata ,family= poisson("log"))
> summary(glm_1)

Call:
glm(formula = Native ~ lnarea + lnelev + lndistnear +
    lnareanear, family = poisson("log"), data = rawdata)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-3.4515  -1.6623    0.2330    0.7056    3.6582
```

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.22136    0.45948   4.835 1.33e-06 ***
lnarea       0.24788    0.02965   8.360  < 2e-16 ***
lnelev       0.07663    0.09321   0.822  0.41101
lndistnear  -0.06046    0.02111  -2.864  0.00418 **
lnareanear  -0.05163    0.01083  -4.767 1.87e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 700.717  on 29  degrees of freedom
Residual deviance:  95.764  on 25  degrees of freedom
AIC: 243.05

Number of Fisher Scoring iterations: 5

> glm_2 = glm(Native~lnarea+lnelev+lndistnear+lnareanear,
      data = rawdata, family = quasipoisson)
> summary(glm_2)

Call:
glm(formula = Native ~ lnarea + lnelev + lndistnear +
    lnareanear, family = quasipoisson, data = rawdata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.4515  -1.6623   0.2330   0.7056   3.6582

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)  2.22136    0.88621   2.507 0.019059 *
lnarea       0.24788    0.05718   4.335 0.000209 ***
lnelev       0.07663    0.17979   0.426 0.673576
lndistnear  -0.06046    0.04071  -1.485 0.150036
lnareanear  -0.05163    0.02089  -2.471 0.020623 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to
    be 3.720004)

    Null deviance: 700.717  on 29  degrees of freedom
Residual deviance:  95.764  on 25  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
```

The Dispersion parameter for quasipoisoon family is 3.720004. It is interpreted that the model is overdisperion.

For Poisson, the most upgrade is negative binomial.

```
> glm_3 = glm.nb(Native~lnarea+lnelev+lndistnear+
    lnareanear, data = rawdata)
> summary(glm_3)

Call:
glm.nb(formula = Native ~ lnarea + lnelev + lndistnear +
    lnareanear, data = rawdata, init.theta = 7.651836043,
    link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8669  -0.9043   0.0536   0.6058   1.7595
```

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.572408   0.883340   2.912  0.00359 **
lnarea       0.272043   0.057445   4.736 2.18e-06 ***
lnelev      -0.003908   0.179064  -0.022  0.98259
lndistnear  -0.068951   0.049617  -1.390  0.16463
lnareanear  -0.023301   0.025106  -0.928  0.35336
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(7.6518) family
    taken to be 1)

    Null deviance: 170.314  on 29  degrees of freedom
Residual deviance:  35.361  on 25  degrees of freedom
AIC: 222.12

Number of Fisher Scoring iterations: 1


              Theta:  7.65
          Std. Err.:  3.37

 2 x log-likelihood:  -210.116
```

(b) Using backward elimination,the chosen variable is *lnarea* according to AIC

```
> step(glm_3)
Start:  AIC=220.12
Native ~ lnarea + lnelev + lndistnear + lnareanear
```

```
              Df Deviance     AIC
- lnelev       1   35.361  218.12
- lnareanear   1   36.233  218.99
- lndistnear   1   37.299  220.06
<none>                    35.361  220.12
- lnarea       1   58.214  240.97

Step:  AIC=218.12
Native ~ lnarea + lndistnear + lnareanear

              Df Deviance     AIC
- lnareanear   1   36.242  216.99
- lndistnear   1   37.315  218.06
<none>                    35.369  218.12
- lnarea       1  165.205  345.95

Step:  AIC=216.94
Native ~ lnarea + lndistnear

              Df Deviance     AIC
- lndistnear   1   35.950  216.54
<none>                    34.345  216.94
- lnarea       1  154.747  335.34

Step:  AIC=216.49
Native ~ lnarea

          Df Deviance     AIC
<none>                 33.953  216.49
- lnarea   1  148.851  329.39

Call:  glm.nb(formula = Native ~ lnarea, data = rawdata,
```

```
      init.theta = 6.343853188, link = log)

Coefficients:
(Intercept)       lnarea
     2.4332       0.2694

Degrees of Freedom: 29 Total (i.e. Null);  28 Residual
Null Deviance:      148.9
Residual Deviance: 33.95        AIC: 218.5
```

(c) The remaining explanatory variable is *lnarea*, and it is positive proportional to the native species.

```
> glm_4 = glm.nb(Native~lnarea,data = rawdata)
> summary(glm_4)

Call:
glm.nb(formula = Native ~ lnarea, data = rawdata,
      init.theta = 6.343852721, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8609  -0.7346  -0.0325   0.6672   1.5328

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.43322    0.10751   22.63   <2e-16 ***
lnarea       0.26939    0.02648   10.17   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(6.3439) family
      taken to be 1)
```

```
    Null deviance: 148.851  on 29  degrees of freedom
Residual deviance:  33.953  on 28  degrees of freedom
AIC: 218.49

Number of Fisher Scoring iterations: 1


              Theta:  6.34
          Std. Err.:  2.52

 2 x log-likelihood:  -212.487
```

(d) We set nonnative species as the response variable. Also we check for extra-Poisson variation: Dispersion parameter for quasipoisson family taken to be 16.07358. So this model is also obviously overdispersion. Like the previous question, we use the negative binomial distribution.

```
> new_1 = glm(Total-Native~lnarea+lnelev+lndistnear+lnareanear,
      data = rawdata, family = poisson("log"))
> summary(new_1)

Call:
glm(formula = Total - Native ~ lnarea + lnelev + lndistnear +
    lnareanear, family = poisson("log"), data = rawdata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.9427  -2.9385  -0.6448   2.1254   8.7721

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.521868   0.363671   6.934 4.08e-12 ***
```

```
lnarea        0.410968   0.023631  17.391  < 2e-16 ***
lnelev        0.043994   0.073296   0.600 0.548353
lndistnear   -0.054287   0.014139  -3.839 0.000123 ***
lnareanear   -0.125707   0.007707 -16.310  < 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: 2950.5  on 29  degrees of freedom
Residual deviance:  328.7  on 25  degrees of freedom
AIC: 474.54

Number of Fisher Scoring iterations: 5

> new_2 = glm(Total-Native~lnarea+lnelev+lndistnear+lnareanear,
      data = rawdata, family = quasipoisson)
> new_2$coefficients
(Intercept)      lnarea      lnelev   lndistnear   lnareanear
 2.52186804  0.41096837  0.04399438 -0.05428670 -0.12570695
> new_3 = glm.nb(Total-Native~lnarea+lnelev+lndistnear+lnareanear ,
      data = rawdata)
> summary(new_3)

Call:
glm.nb(formula = Total - Native ~ lnarea + lnelev + lndistnear +
    lnareanear, data = rawdata, init.theta = 1.548034338, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3402  -0.8733  -0.3772   0.5433   1.7164
```

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.17068    1.69189   2.465   0.0137 *
lnarea       0.48794    0.11026   4.425 9.63e-06 ***
lnelev      -0.28192    0.34314  -0.822   0.4113
lndistnear  -0.12403    0.09790  -1.267   0.2052
lnareanear  -0.05808    0.04917  -1.181   0.2375
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(1.548) family
    taken to be 1)

    Null deviance: 112.361  on 29  degrees of freedom
Residual deviance:  36.606  on 25  degrees of freedom
AIC: 265.13

Number of Fisher Scoring iterations: 1


              Theta:  1.548
          Std. Err.:  0.483

 2 x log-likelihood:  -253.131
```

We then use the backward elimination to get the result:

```
> step(new_3)
Start:  AIC=263.13
Total - Native ~ lnarea + lnelev + lndistnear + lnareanear

             Df Deviance    AIC
- lnelev      1   37.328 261.85
- lndistnear  1   37.924 262.45
```

```
- lnareanear  1    38.084 262.61
<none>              36.606 263.13
- lnarea      1    56.894 281.42

Step:  AIC=261.84
Total - Native ~ lnarea + lndistnear + lnareanear

             Df Deviance     AIC
- lnareanear  1    37.986 261.25
- lndistnear  1    38.438 261.71
<none>              36.576 261.84
- lnarea      1   106.019 329.29

Step:  AIC=261.21
Total - Native ~ lnarea + lndistnear

             Df Deviance     AIC
- lndistnear  1    37.554 260.60
<none>              36.164 261.21
- lnarea      1   100.839 323.88

Step:  AIC=260.56
Total - Native ~ lnarea

         Df Deviance     AIC
<none>         36.033 260.56
- lnarea  1    98.474 321.00

Call:  glm.nb(formula = Total - Native ~ lnarea, data = rawdata,
    init.theta = 1.331670759, link = log)

Coefficients:
```

```
(Intercept)       lnarea
     2.6017       0.3936

Degrees of Freedom: 29 Total (i.e. Null);  28 Residual
Null Deviance:      98.47
Residual Deviance: 36.03        AIC: 262.6
> new_4 = glm.nb(Total-Native~lnarea, data = rawdata)
> summary(new_4)

Call:
glm.nb(formula = Total - Native ~ lnarea, data = rawdata,
    init.theta = 1.33167056, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2308  -0.9959  -0.3464   0.5097   1.8919

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.60166    0.19221  13.536  < 2e-16 ***
lnarea       0.39362    0.05035   7.818 5.35e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(1.3317) family
    taken to be 1)

    Null deviance: 98.474  on 29  degrees of freedom
Residual deviance: 36.033  on 28  degrees of freedom
AIC: 262.56

Number of Fisher Scoring iterations: 1
```

```
          Theta:  1.332
      Std. Err.:  0.396

 2 x log-likelihood:  -256.561
```

It is clear that we only take the variable lnarea. The effects of each remaining variable: lnarea is positive proportional to the non-native species.

Problem 2

(a) First of all, we check the mean and variance of *Storms* and we get:

$$Mean = 9.395833$$

$$Variance = 10.371897$$

Since the mean of variable Storms is close its variance, Poisson log-linear model makes sense. First we fit Poisson log-linear model without interaction of Temperature and WestAfrica.

```
Call:
glm(formula = Storms ~ Temperature + WestAfrica,
    family = poisson(link = "log"), data = mydata)

Deviance Residuals:
     Min       1Q   Median       3Q      Max
-1.58885  -0.55182  -0.01495  0.42804  1.93586

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.30891    0.10107  22.846  < 2e-16 ***
Temperature0 -0.06399    0.11377  -0.562  0.57380
Temperature1 -0.37645    0.12848  -2.930  0.00339 **
WestAfrica1   0.15596    0.10260   1.520  0.12849
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 50.875  on 47  degrees of freedom
Residual deviance: 33.693  on 44  degrees of freedom
AIC: 235.61

Number of Fisher Scoring iterations: 4
```

Then fit the model with interaction terms.

```
Call:
glm(formula = Storms ~ Temperature * WestAfrica,
    family = poisson(link = "log"), data = mydata)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.67320  -0.59282   0.01746   0.39963   1.95558

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                2.31911    0.12804  18.113   <2e-16 ***
Temperature0              -0.04418    0.16043  -0.275   0.7830
Temperature1              -0.42972    0.16740  -2.567   0.0103 *
WestAfrica1                0.14047    0.15793   0.889   0.3737
Temperature0:WestAfrica1  -0.07360    0.23134  -0.318   0.7504
Temperature1:WestAfrica1   0.20372    0.26885   0.758   0.4486
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)
```

```
    Null deviance: 50.875  on 47  degrees of freedom
Residual deviance: 32.678  on 42  degrees of freedom
AIC: 238.59

Number of Fisher Scoring iterations: 4

> deviance(glm_2)-deviance(glm_1)
[1] -1.014462
```

The deviance of model with interaction is smaller. Thus, the model with interaction is a better fit.

Then, we check extra-Poisson variation with quasi-Poisson model.

```
> glm_3 = glm(Storms~Temperature*WestAfrica, data=mydata,
    family = quasipoisson(link="log"))
> summary(glm_3)

Call:
glm(formula = Storms ~ Temperature * WestAfrica,
    family = quasipoisson(link = "log"), data = mydata)

Deviance Residuals:
     Min       1Q   Median       3Q      Max
-1.67320  -0.59282  0.01746  0.39963  1.95558

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                 2.31911    0.11365  20.405  < 2e-16 ***
Temperature0               -0.04418    0.14241  -0.310  0.75792
Temperature1               -0.42972    0.14859  -2.892  0.00604 **
WestAfrica1                 0.14047    0.14018   1.002  0.32204
Temperature0:WestAfrica1   -0.07360    0.20535  -0.358  0.72182
Temperature1:WestAfrica1    0.20372    0.23865   0.854  0.39815
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family
    taken to be 0.787926)

    Null deviance: 50.875  on 47  degrees of freedom
Residual deviance: 32.678  on 42  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4
```

From the above we can see that the dispersion parameter for quasipoisson family taken to be $0.787926 \leq 1$. So we don't consider it as an overdispersion.

Thus, we get the best model:

$$ln(\mu_i) = 2.319 - 0.044Temperature_{1,i} - 0.430Temperature_{2,i} + 0.140WestAfrica_i -$$

$$0.074Temperature_{1,i} * WestAfrica_i + 0.204Temperature_{2,i} * WestAfrica_i$$

It is same to conclude that more storms tend to occur in a cold El Nino year or when west Africa is relatively dry.

(b) We check the mean and variance of the number of hurricanes.

$$Mean = 5.750000$$

$$Variance = 5.595745$$

The mean of Hurricanes is close to its variance, so we consider fitting a Poisson log-linear model again.

Fitting Poisson log-linear model with no interaction terms.

```
> glm_4 = glm(Hurricanes~Temperature+WestAfrica,
      data=mydata, family = poisson(link="log"))
```

```
> summary(glm_4)

Call:
glm(formula = Hurricanes ~ Temperature + WestAfrica,
    family = poisson(link = "log"), data = mydata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.312  -0.500  -0.274   0.480   1.859

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.80344    0.12915  13.964  < 2e-16 ***
Temperature0 -0.04463    0.14353  -0.311  0.75585
Temperature1 -0.46206    0.16741  -2.760  0.00578 **
WestAfrica1   0.21913    0.13046   1.680  0.09303 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 44.414  on 47  degrees of freedom
Residual deviance: 27.322  on 44  degrees of freedom
AIC: 205.15

Number of Fisher Scoring iterations: 4
```

Then fit another Poisson regression model with interaction terms.

```
> glm_5 = glm(Hurricanes~Temperature*WestAfrica,
    data=mydata,family = poisson(link="log"))
> summary(glm_5)

Call:
```

```
glm(formula = Hurricanes ~ Temperature * WestAfrica,
    family = poisson(link = "log"), data = mydata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3030  -0.5242  -0.2092   0.4746   1.8020

Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                 1.763589   0.169031  10.434   <2e-16 ***
Temperature0               -0.002601   0.210229  -0.012   0.9901
Temperature1               -0.396712   0.219498  -1.807   0.0707 .
WestAfrica1                 0.277632   0.203859   1.362   0.1732
Temperature0:WestAfrica1   -0.064539   0.291481  -0.221   0.8248
Temperature1:WestAfrica1   -0.178171   0.371604  -0.479   0.6316
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 44.414  on 47  degrees of freedom
Residual deviance: 27.086  on 42  degrees of freedom
AIC: 208.92

Number of Fisher Scoring iterations: 4

> deviance(glm_5)-deviance(glm_4)
[1] -0.2356582
```

Compared with the deviance of no-interaction model, the model with interaction terms has a smaller deviance which means it is a better fit in this case. Then, we check extra-Poisson variation by quasi-Poisson regression model.

```
Call:
```

```
glm(formula = Hurricanes ~ Temperature * WestAfrica,
    family = quasipoisson(link = "log"), data = mydata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3030  -0.5242  -0.2092   0.4746   1.8020

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                1.763589   0.138883  12.698 5.69e-16 ***
Temperature0              -0.002601   0.172733  -0.015   0.9881
Temperature1              -0.396712   0.180348  -2.200   0.0334 *
WestAfrica1                0.277632   0.167499   1.658   0.1049
Temperature0:WestAfrica1  -0.064539   0.239493  -0.269   0.7889
Temperature1:WestAfrica1  -0.178171   0.305325  -0.584   0.5626
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family
    taken to be 0.6750954)

    Null deviance: 44.414  on 47  degrees of freedom
Residual deviance: 27.086  on 42  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4
```

Dispersion parameter for quasipoisson family taken to be 0.675. There is no overdispersion problem.

Therefore, we get the best model:

$$ln(\mu_i) = 1.764 - 0.003Temperature_{1,i} - 0.397Temperature_{2,i} + 0.278WestAfrica_i -$$

$$0.065Temperature_{1,i} * WestAfrica_i - 0.178Temperature_{2,i} * WestAfrica_i$$