Gala

Heejung Shim

There are 30 Galapagos islands and 7 variables in the dataset. The relationship between the number of plant species and several geographic variables is of interest.

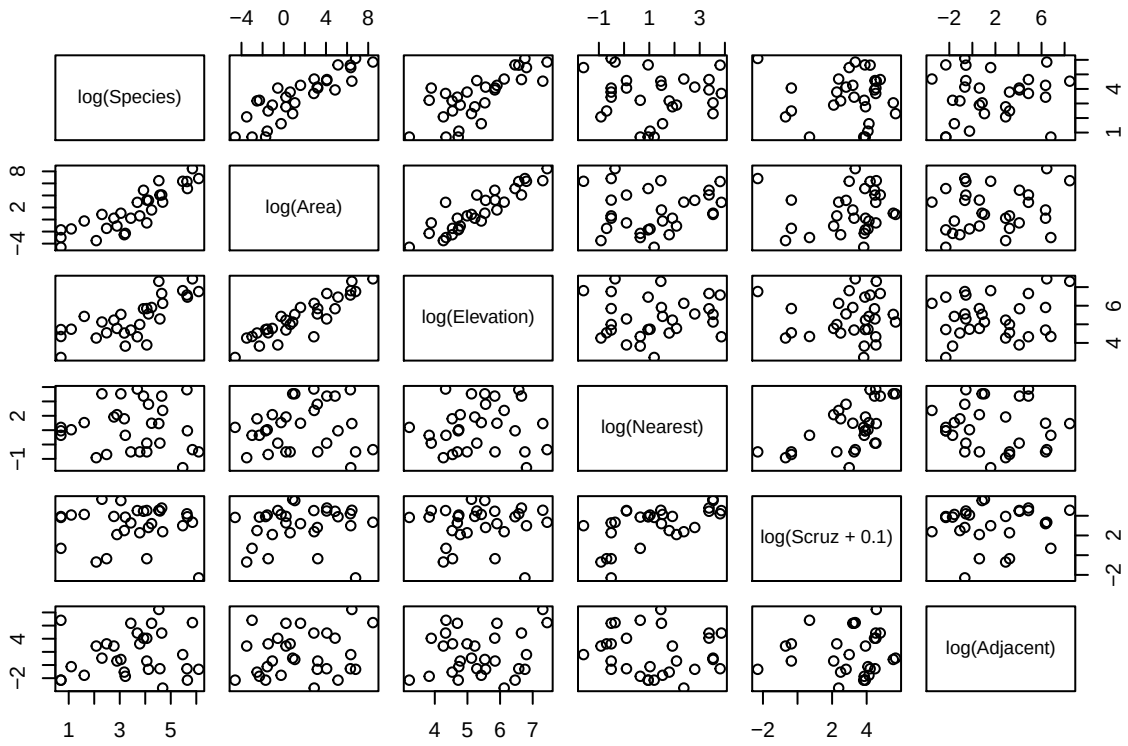
```
library(faraway)
data(gala)
str(gala)
```

```
## 'data.frame':    30 obs. of  7 variables:
##  $ Species  : num  58 31 3 25 2 18 24 10 8 2 ...
##  $ Endemics : num  23 21 3 9 1 11 0 7 4 2 ...
##  $ Area     : num  25.09 1.24 0.21 0.1 0.05 ...
##  $ Elevation: num  346 109 114 46 77 119 93 168 71 112 ...
##  $ Nearest  : num  0.6 0.6 2.8 1.9 1.9 8 6 34.1 0.4 2.6 ...
##  $ Scruz    : num  0.6 26.3 58.7 47.4 1.9 ...
##  $ Adjacent : num  1.84 572.33 0.78 0.18 903.82 ...
```

```
?gala
```

Check pairwise relationship and fit the Poisson regression.

```
with(gala, pairs(~ log(Species) + log(Area) + log(Elevation) +
                   log(Nearest) + log(Scruz+0.1) + log(Adjacent)))
```



```
mod <- glm(Species ~ log(Area) + log(Elevation) +
           log(Nearest) + log(Scruz+0.1) + log(Adjacent),
           family=poisson, gala)
summary(mod)
```

```
##
## Call:
## glm(formula = Species ~ log(Area) + log(Elevation) + log(Nearest) +
##     log(Scruz + 0.1) + log(Adjacent), family = poisson, data = gala)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.4479  -2.6717  -0.4547   2.5613   8.2970
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       3.287941   0.284661  11.550  < 2e-16 ***
## log(Area)         0.348445   0.018029  19.327  < 2e-16 ***
## log(Elevation)    0.036421   0.056983   0.639  0.52272
## log(Nearest)     -0.040644   0.013781  -2.949  0.00318 **
## log(Scruz + 0.1) -0.030045   0.010492  -2.864  0.00419 **
## log(Adjacent)    -0.089014   0.006948 -12.812  < 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: 3510.73  on 29  degrees of freedom
## Residual deviance:  359.12  on 24  degrees of freedom
## AIC: 531.96
##
## Number of Fisher Scoring iterations: 5
```

```
modp = step(mod)
```

```
## Start:  AIC=531.96
## Species ~ log(Area) + log(Elevation) + log(Nearest) + log(Scruz +
##     0.1) + log(Adjacent)
##
##                    Df Deviance    AIC
## - log(Elevation)    1   359.54 530.37
## <none>                  359.12 531.96
## - log(Scruz + 0.1)  1   367.27 538.10
## - log(Nearest)      1   367.79 538.62
## - log(Adjacent)     1   525.13 695.96
## - log(Area)         1   714.98 885.81
##
## Step:  AIC=530.37
## Species ~ log(Area) + log(Nearest) + log(Scruz + 0.1) + log(Adjacent)
##
##                    Df Deviance    AIC
## <none>                   359.5  530.4
## - log(Scruz + 0.1)  1    367.7  536.6
## - log(Nearest)      1    368.5  537.3
```

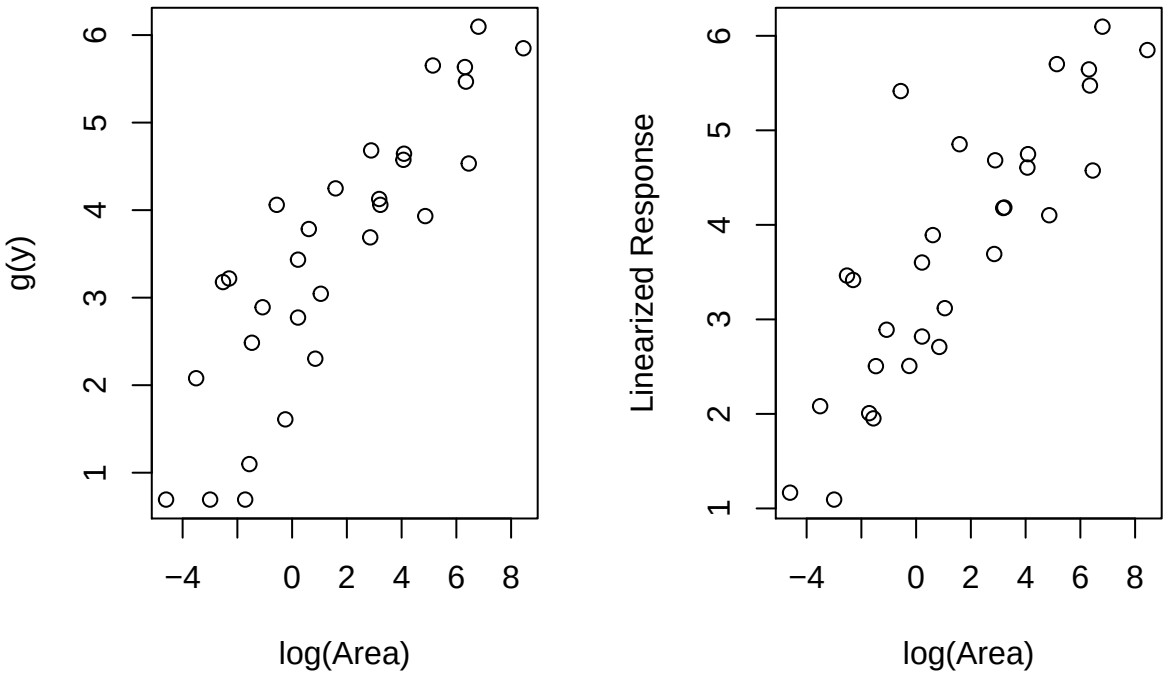
```
## - log(Adjacent)     1    528.6  697.4
## - log(Area)         1   3266.1 3434.9
```

```
summary(modp)
```

```
##
## Call:
## glm(formula = Species ~ log(Area) + log(Nearest) + log(Scruz +
##     0.1) + log(Adjacent), family = poisson, data = gala)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.3457  -2.7891  -0.6233   2.5129   8.1217
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       3.466484   0.053704  64.549  < 2e-16 ***
## log(Area)         0.358711   0.008254  43.460  < 2e-16 ***
## log(Nearest)     -0.041117   0.013733  -2.994  0.00275 **
## log(Scruz + 0.1) -0.030098   0.010478  -2.873  0.00407 **
## log(Adjacent)    -0.088224   0.006842 -12.895  < 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: 3510.73  on 29  degrees of freedom
## Residual deviance:  359.54  on 25  degrees of freedom
## AIC: 530.37
##
## Number of Fisher Scoring iterations: 5
```

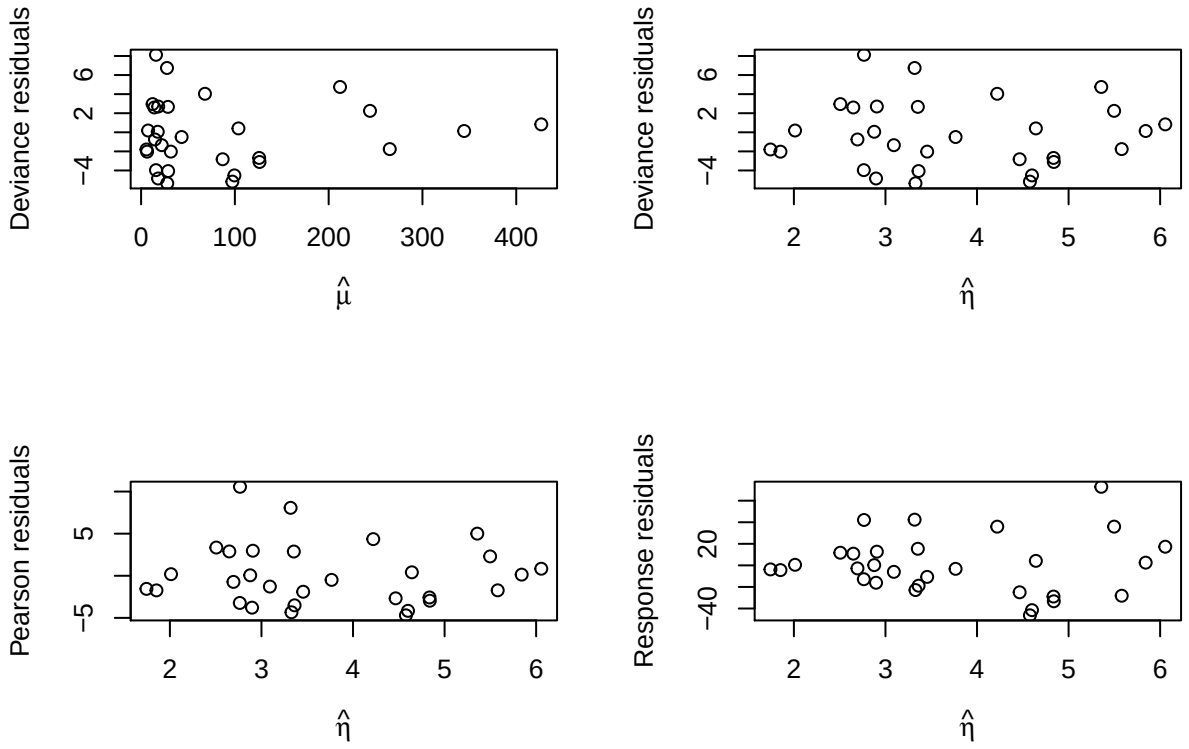
Checking linearity.

```
par(mfrow=c(1,2))
# g(y) vs log(Area)
plot(log(Species) ~ log(Area), gala, ylab="g(y)")
# linearised response vs log(Area)
mu <- predict(modp, type="response")
z <- predict(modp) + (gala$Species - mu)/mu
plot(z ~ log(Area), gala, ylab="Linearized Response")
```

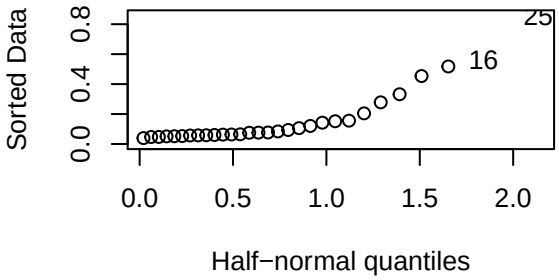


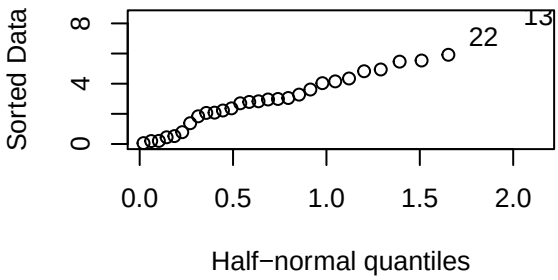
Check outliers/influential points.

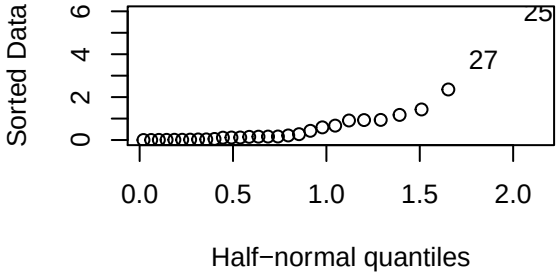
```
par(mfrow=c(2,2))
# deviance residuals vs fitted values - cramped
plot(residuals(modp) ~ predict(modp,type="response"),
     xlab=expression(hat(mu)), ylab="Deviance residuals")
# deviance residuals vs linear fitted values - look OK
plot(residuals(modp) ~ predict(modp,type="link"),
     xlab=expression(hat(eta)), ylab="Deviance residuals")
# pearson residuals vs linear fitted values - look OK
plot(residuals(modp,type="pearson") ~ predict(modp, type="link"),
      xlab=expression(hat(eta)), ylab="Pearson residuals")
# response residuals vs linear fitted values - heteroskedastic
plot(residuals(modp,type="response") ~ predict(modp, type="link"),
      xlab=expression(hat(eta)), ylab="Response residuals")
```




```
par(mfrow=c(2,2))
# leverages - obs 25 has moderately high leverage
halfnorm(influence(modp)$hat)
# jackknife residuals - no evidence of outliers
halfnorm(rstudent(modp))
# Cook's distance - obs 25 looking influential
halfnorm(cooks.distance(modp))
# observation 25 has Scruz=0, which gives -infty when taking log
# our artificial adjustment by 0.1 could be the problem
```







Effect of removing obs 25 on model

```
mod2 <- glm(Species ~ log(Area) + log(Elevation) +
            log(Nearest) + log(Scruz+0.1) + log(Adjacent),
            family=poisson, gala, subset=-25)
summary(mod2)
```

```
##
## Call:
## glm(formula = Species ~ log(Area) + log(Elevation) + log(Nearest) +
##     log(Scruz + 0.1) + log(Adjacent), family = poisson, data = gala,
##     subset = -25)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.7237  -2.7539  -0.3181   2.6401   7.9333
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)        3.05070    0.30033  10.158  < 2e-16 ***
## log(Area)          0.33453    0.01883  17.770  < 2e-16 ***
## log(Elevation)     0.05960    0.05743   1.038 0.299325
## log(Nearest)      -0.05255    0.01469  -3.578 0.000347 ***
## log(Scruz + 0.1)   0.01592    0.02218   0.718 0.472998
## log(Adjacent)     -0.08852    0.00696 -12.717  < 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: 2707.88  on 28  degrees of freedom
## Residual deviance:  353.42  on 23  degrees of freedom
## AIC: 518.32
##
## Number of Fisher Scoring iterations: 5
```

```
modp2 = step(mod2)
```

```
## Start:  AIC=518.32
## Species ~ log(Area) + log(Elevation) + log(Nearest) + log(Scruz +
##     0.1) + log(Adjacent)
##
##                     Df Deviance    AIC
## - log(Scruz + 0.1)  1   353.94 516.84
## - log(Elevation)    1   354.51 517.41
## <none>                  353.42 518.32
## - log(Nearest)      1   366.21 529.11
## - log(Adjacent)     1   516.83 679.73
## - log(Area)         1   663.37 826.27
##
## Step:  AIC=516.84
## Species ~ log(Area) + log(Elevation) + log(Nearest) + log(Adjacent)
##
##                   Df Deviance    AIC
## - log(Elevation)  1   354.83 515.72
## <none>                353.94 516.84
## - log(Nearest)    1   368.20 529.09
## - log(Adjacent)   1   519.96 680.86
## - log(Area)       1   679.00 839.90
##
## Step:  AIC=515.72
## Species ~ log(Area) + log(Nearest) + log(Adjacent)
##
##                 Df Deviance     AIC
## <none>               354.83  515.72
## - log(Nearest)   1   369.86  528.76
## - log(Adjacent)  1   521.71  680.60
## - log(Area)      1  2679.93 2838.82
```

```
summary(modp2)
```

```
##
## Call:
## glm(formula = Species ~ log(Area) + log(Nearest) + log(Adjacent),
##     family = poisson, data = gala, subset = -25)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.4799  -2.9871  -0.6226   2.5441   7.7881
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    3.384649   0.048760  69.415  < 2e-16 ***
## log(Area)      0.352919   0.008598  41.048  < 2e-16 ***
```

```
## log(Nearest)  -0.047878   0.012378  -3.868  0.00011 ***
## log(Adjacent) -0.086616   0.006776 -12.783  < 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: 2707.88  on 28  degrees of freedom
## Residual deviance:  354.83  on 25  degrees of freedom
## AIC: 515.72
##
## Number of Fisher Scoring iterations: 5
```

```
# without obs 25 log(Scruz+0.1) and log(Elevation) not significant
```