

The Problem

Paralleling what we did with the Galapagos Islands data of Johnson and Raven (1973) fit the following five models to the variable species richness S as a function of island area A .

In each model β_0 and β_1 are parameters to be estimated. Observe that models 1 and 3 can be fit as ordinary regression models (general linear models), model 2 is a nonlinear model (fit using nonlinear least squares), and models 4 and 5 are generalized linear models proper.

For each model,

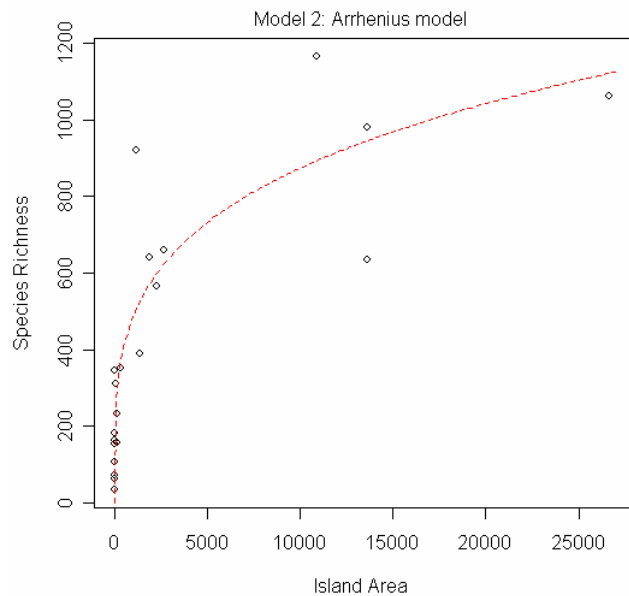
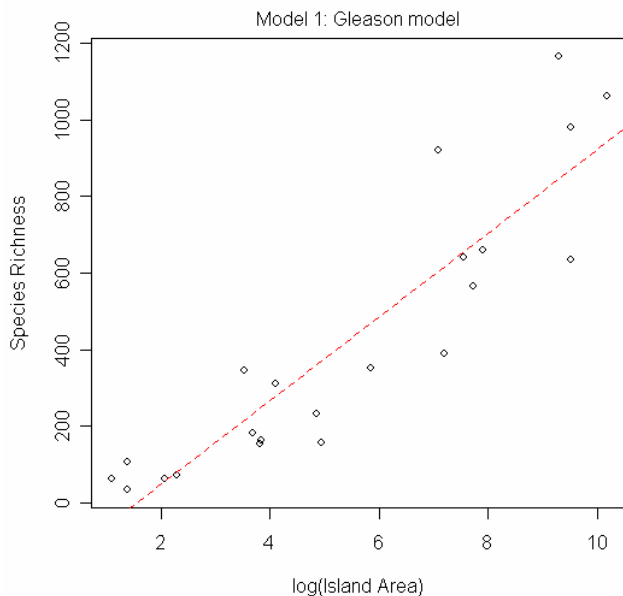
1. Report the fitted equation you obtain.
2. Graph the fitted equation superimposed on a scatter plot of the data. You may include more than one model on the same scatter plot when it's appropriate.
3. Finally compare all five models and determine which is (are) best for these data.

1. Gleason model: $S \sim \text{Normal}(\mu, \sigma^2)$
 with identity link such that $\mu = \beta_0 + \beta_1 \log A$

The fitted equation: $\mu = -171.0272 + 109.4431 \log A$.

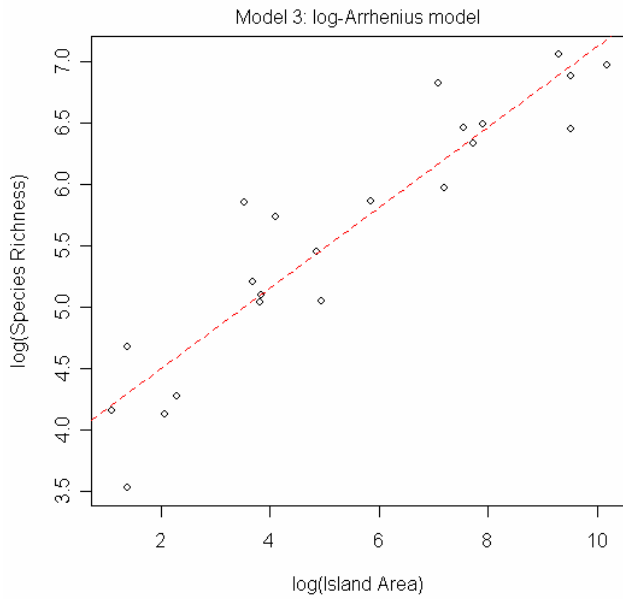
2. Arrhenius model: $S \sim \text{Normal}(\mu, \sigma^2)$
 with identity link such that $\mu = \beta_0 A^{\beta_1}$

The fitted equation: $\mu = 83.4782041 * A^{0.2549087}$.



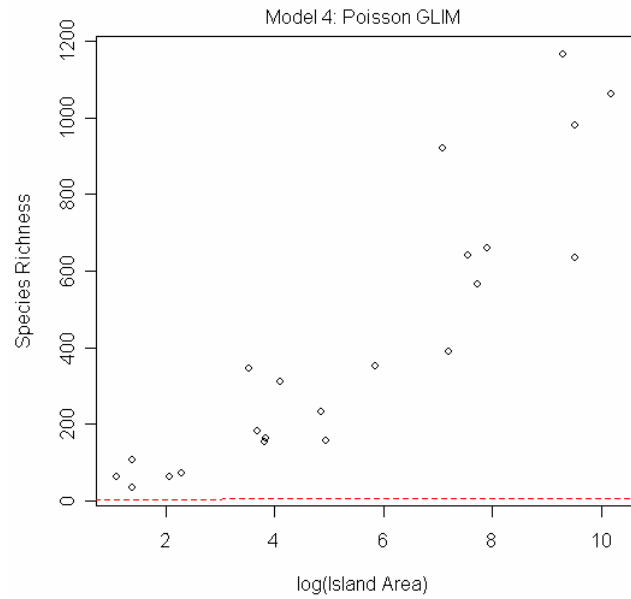
3. Log-Arrhenius model: $\log S \sim \text{Normal}(\mu, \sigma^2)$
 with identity link such that $\mu = \beta_0 + \beta_1 \log A$

The fitted equation: $\mu = 3.8390 + 0.3283 \log A$.



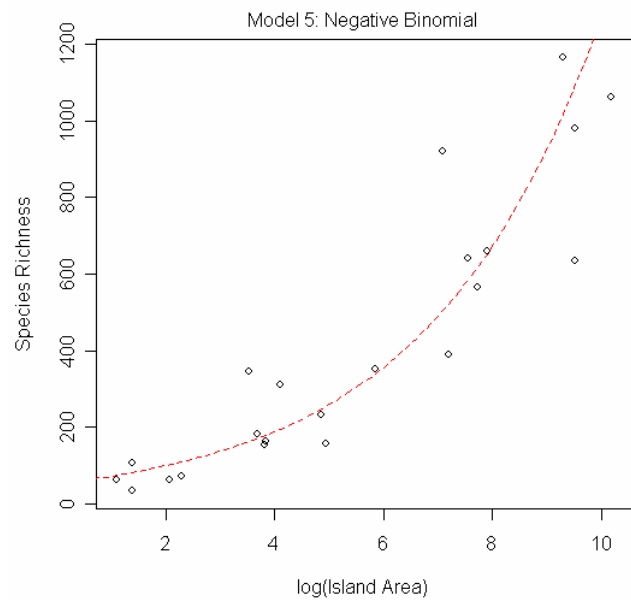
4. Poisson GLIM: $S \sim \text{Poisson}(\lambda)$
 with log link such that $\log \lambda = \beta_0 + \beta_1 \log A$

The fitted equation: $\log \lambda = 4.1925 + 0.2836 \log A$.



5. Negative binomial GLIM: $S \sim \text{Negative binomial}(\mu, \theta)$
 with log link such that $\log \mu = \beta_0 + \beta_1 \log A$

The fitted equation: $\log \mu = 3.9730 + 0.3172 \log A$.



Comparing the 5 Models

```
> models<-list(model1,model2,model3,model4,model5)
> model.names<-c('Gleason','Arrhenius','log-Arrhenius','Poisson GLIM','NB
GLIM')
> loglike<-c(logLik(model1), logLik(model2), model3.vals[[1]],
logLik(model4), logLik(model5))
> numparms<-c(3,3,3,2,3) #the 3 normal-based models have 3 parameters:  $\beta_0$ ,
 $\beta_1$ , and  $\sigma_2$ ; the Poisson has 2:  $\beta_0$  and  $\beta_1$ , and the NB has 3:  $\beta_0$ ,  $\beta_1$ , and  $\theta$ .

> AIC.func<-function(LL,K,n,modelnames)
{
#LL is loglikelihood, #K is number of estimated params., #n is sample size.
AIC<- -2*LL + 2*K
AICc<-AIC + 2*K*(K+1)/(n-K+1)
output<-cbind(LL,K,AIC,AICc)
colnames(output)<-c('LogL','K','AIC','AICc')
minAICc<-min(output[,"AICc"])
deltai<-output[,"AICc"]-minAICc
rel.like<-exp(-deltai/2)
wi<-round(rel.like/sum(rel.like),3)
out<-data.frame(modelnames,output,deltai,wi)
out
}
> dim(islands) #22 is the number of islands.
[1] 22 14

> AIC.func(loglike,numparms,dim(islands)[1],model.names)
  modelnames      LogL K      AIC      AICc      deltai      wi
1     Gleason -141.4406 3  288.8813  290.0813  16.586909 0.000
2   Arrhenius -140.4282 3  286.8563  288.0563  14.561930 0.000
3 log-Arrhenius -133.1472 3  272.2944  273.4944   0.000000 0.632
4   Poisson GLIM -542.1834 2 1088.3668 1088.9382 815.443834 0.000
5         NB GLIM -133.6894 3  273.3788  274.5788   1.084434 0.367
```

This analysis suggests that the best model for the species richness-island area relationship is the log-Arrhenius model, because it has the lowest AICc (273.4944) and the highest Akaike weight (0.632). However, the negative binomial-based GLIM has the next lowest AICc (274.5788). Its Akaike weight of 0.367 indicates it is also a model that should be considered, and a Δ_i between 0 and 2 (1.084434) suggests there is excellent evidence in the data for this model. Thus, although the log-Arrhenius model is the best model, the negative binomial GLIM is also a considerable model to describe the species-area relationship in this group of islands.

R Code

1. Gleason model: $S \sim \text{Normal}(\mu, \sigma^2)$ with identity link such that $\mu = \beta_0 + \beta_1 \log A$

```
> islands<-
read.table('http://www.unc.edu/courses/2006spring/ecol/145/001/data/lab6/islands.txt', header=TRUE, sep=',')

> table(islands$sp.richness)
 34  62  64  72 107 155 156 163 182 232 311 347 353
  1   1   1   1   1   1   1   1   1   1   1   1   1
390 564 633 641 661 920 979 1060 1166
  1   1   1   1   1   1   1   1   1

> table(islands$island.area)
  3   4   8  10  34  40  46  47  61 128 140
  1   2   1   1   1   1   1   1   1   1   1
350 1190 1350 1900 2300 2707 10900 13600 26668
  1   1   1   1   1   1   1   1   2   1

> model1<-lm(islands$sp.richness~log(islands$island.area))
> model1
Call:
lm(formula = islands$sp.richness ~ log(islands$island.area))
Coefficients:
      (Intercept)  log(islands$island.area)
      -171.0272      109.4431

> logLik(model1)
'log Lik.' -141.4406 (df=3)
> AIC(model1)
[1] 288.8813

> coef(model1)
      (Intercept)  log(islands$island.area)
      -171.0272      109.4431
```

The fitted equation: $\mu = -171.0272 + 109.4431 \log A$.

```
> plot(log(islands$island.area), islands$sp.richness, xlab='log(Island Area)',
ylab='Species Richness')
> abline(model1, col=2, lty=2)
> mtext('Model 1: Gleason model', side=3, line=.5)
```

2. Arrhenius model: $S \sim \text{Normal}(\mu, \sigma^2)$ with identity link such that $\mu = \beta_0 A^{\beta_1}$

```
> model2<-nls(islands$sp.richness~b0*islands$island.area^b1, data=islands,
start=list(b0=1,b1=.5))
> model2
Nonlinear regression model
model: islands$sp.richness ~ b0 * islands$island.area^b1
```

```

data: islands
      b0      b1
83.4782041 0.2549087
residual sum-of-squares: 451062.6

```

The fitted equation: $\mu = 83.4782041 * A^{0.2549087}$.

```

> logLik(model2)
'log Lik.' -140.4282 (df=2)
> -2*logLik(model2)+2*(length(coef(model2))+1) #correct AIC that accounts for sigma^2
[1] 286.8563 #the AIC term

> arrhenius.func<-function(x) coef(model2)[1]*x^(coef(model2)[2])
> plot(islands$island.area,islands$sp.richness, xlab='Island Area',
ylab='Species Richness')
> lines(seq(0,27000,100),arrhenius.func(seq(0,27000,100)),lty=2,col=2)
> mtext('Model 2: Arrhenius model',side=3,line=.5)

```

3. Log-Arrhenius model: $\log S \sim \text{Normal}(\mu, \sigma^2)$ with identity link such that

$\mu = \beta_0 + \beta_1 \log A$

```

> model3<-lm(log(sp.richness)~log(island.area), data=islands)
> model3
Call:
lm(formula = log(sp.richness) ~ log(island.area), data = islands)

Coefficients:
(Intercept)  log(island.area)
      3.8390           0.3283

```

The fitted equation: $\mu = 3.8390 + 0.3283 \log A$.

```

> norm.loglike<-function(data,model)
{
t.y<-log(data$sp.richness)
sigma2<-(sum(residuals(model)^2)/dim(data)[1])
loglike<-sum(log(dnorm(t.y, mean=predict(model),
sd=sqrt(sigma2))*1/(data$sp.richness)))
out<-list(loglike, c(coef(model), sigma2))
out
}

> norm.loglike(islands,model3)
[[1]]
[1] -133.1472

[[2]]
(Intercept) log(island.area)
      3.8389781      0.3283347      0.1407124

> norm.loglike(islands,model3)->model3.vals

> -2*model3.vals[[1]]+2*length(model3.vals[[2]]) #correct AIC

```

[1] 272.2944

```
> plot(log(islands$island.area),log(islands$sp.richness), xlab='log(Island Area)', ylab='log(Species Richness)')
> abline(model3,col=2,lty=2)
> mtext('Model 3: log-Arrhenius model', side=3, line=.5)
```

4. Poisson GLIM: $S \sim \text{Poisson}(\lambda)$ with log link such that $\log \lambda = \beta_0 + \beta_1 \log A$

```
> model4<-glm(sp.richness~log(island.area),data=islands,family=poisson)
> model4
Call: glm(formula = sp.richness ~ log(island.area), family = poisson,
data = islands)
```

Coefficients:

```
(Intercept) log(island.area)
4.1925      0.2836
```

```
Degrees of Freedom: 21 Total (i.e. Null); 20 Residual
Null Deviance: 6211
Residual Deviance: 920.4 AIC: 1088
```

The fitted equation: $\log \lambda = 4.1925 + 0.2836 \log A$.

```
> logLik(model4)
'log Lik.' -542.1834 (df=2)
> AIC(model4)
[1] 1088.367 (incredibly bad fit)

> coef(model4)
(Intercept) log(island.area)
4.1925024    0.2835591
> pois.func<-function(x) coef(model4)[1]+ coef(model4)[2]*x
> plot(log(islands$island.area), islands$sp.richness, xlab='log(Island Area)', ylab='Species Richness')
> lines(seq(0,11,.1), pois.func(seq(0,11,.1)), lty=2, col=2)
> mtext('Model 4: Poisson',side=3,line=.5)
```

5. Negative binomial GLIM: $S \sim \text{Negative binomial}(\mu, \theta)$ with log link such that $\log \mu = \beta_0 + \beta_1 \log A$

```
> library(MASS)
> model5<-glm.nb(sp.richness~log(island.area),data=islands)
```

```
Call: glm.nb(formula = sp.richness ~ log(island.area), data = islands,
init.theta = 7.44217589357906, link = log)
```

Coefficients:

```
(Intercept) log(island.area)
3.9730      0.3172
```

Degrees of Freedom: 21 Total (i.e. Null); 20 Residual
Null Deviance: 135.7
Residual Deviance: 22.65 AIC: 273.4

The fitted equation: $\log \mu = 3.9730 + 0.3172 \log A$.

```
> logLik(model5)
'log Lik.' -133.6894 (df=3)
> AIC(model5)
[1] 273.3788

> coef(model5)
      (Intercept) log(island.area)
      3.9729777      0.3172363
> NB.func<-function(x) exp(coef(model5)[1]+ coef(model5)[2]*x)
> plot(log(islands$island.area), islands$sp.richness, xlab='log(Island
Area)', ylab='Species Richness')
> lines(seq(0,11,.1), NB.func(seq(0,11,.1)), lty=2, col=2)
> mtext('Model 5: Negative Binomial',side=3,line=.5)
```