Follow-up course regression, GLM & GAM

Highland Statistics Ltd   www.highstat.comn

Exercise 4: Responses of bald Eagles to human activity during the summer in interior Alaska

Description

Steidl and Anthony (1996) studied the impacts of human activities on bald eagle populations in Alaska. They measured the flush distance response of breeding bald eagles (*Haliaeetus leucocephalus*) to recreational boating activities along the Gulkana River in interior Alaska from 1989 to 1992. Eagles are attracted to the area because salmon spawn throughout the basin during the summer, and river areas within the basin are impacted differentially by recreational activities. On this system, therefore, management strategies are necessary if eagles and humans are to coexist successfully.

To understand how the disturbance context influenced flash distance the effect of various covariates was assessed.

The data are in the file Eagles3.xls and Table 1 gives the variables.

Table 1. Variables in the file Eagle.xls.

Variable	Description	Type	Nature
Flush	Distance at which eagle flushed (meters)	Response	Continuous
Date	Date of observation	Covariate	Continuous
Age	Age of eagle: A = Adult, N = Near adult, B & C = Subadult, J = Juvenile	Covariate	Factor
Activity	Activity: P = Perched; N = On Nest	Covariate	Factor
NumEagles	Number of eagles in immediate vicinity	Covariate	Continuous
Time	Time of day	Covariate	Continuous
DisRiv	Distance eagle perched inland from river's edge (meters)	Covariate	Continuous
Breeding	Yes or No	Covariate	Factor
Height	Height eagle perched above ground (meters)	Covariate	Continuous
Fvis	Distance at which we were first visible to eagle (meters)	Covariate	Continuous
Temp	Ambient temperature (degrees F)	Covariate	Continuous
Cloud	Cloud cover (%)	Covariate	Continuous
fWind	Windspeed: C = calm, L = low, M = medium	Covariate	Factor
fRain	Raining: Yes or No	Covariate	Factor
fLocation	Location: U = Upper Gulkana, L = Lower Gulkana, W = West Fork, M = Middle Fork	Covariate	Factor

References:

- Steidl,R.J., and R.G. Anthony (1996). Responses of bald Eagles to human activity during the summer in interior Alaska. Ecological Applications. 6 (2). Pp 482-491.

Underlying questions and task

The response variable is Flush distance:

- A scientist walks towards an eagle (who sites in a tree) and measures at which distance the eagle flies away. This is called the **flush distance** (in meters).

The underlying question is simple; which covariates drive the Flush distance?

We have already done one exercise using linear regression and we applied classical model selection using AIC. In this exercise we will use an Information Theoretic approach. For this we need to create about 10 – 15 hypotheses what could drive flush distances.

We are not experts on eagles so view the following hypotheses as home-cooked ideas that we will use for the purpose of explaining IT. Here are 13 potential hypotheses:

1. Flush distance is a function of human disturbance, where 'human disturbance' refers to the distance at which an eagle sees the scientist for the first time (**Fvis**). This theory would lead to a model of the form: **Flush = Fvis + noise**. Note that Flush is the distance between the eagle and the observer (in meters) at which the eagle flies away. Fvis is the distance at which the eagle sees the scientist for the first time.

2. Flush distance is weather related. Perhaps absence of wind makes it more difficult for eagles to fly away from the nest, and it will therefore wait longer (lower flush values) before flying away when the observer approaches. Interactions between some of the weather variables may make sense. This theory would lead to a model of the form: **Flush = Temp + Cloud + fWind + fRain + noise**, perhaps with all 2-way interactions.

3. An eagle that sits in a high tree may wait longer before flying away after spotting the scientist. This would mean a model of form **Flush = Height + Fvis + Height x Fvis + noise**. Note that it contains the interaction between Height and Fvis.

4. We can extend the model in 3 by adding location, and an interaction between Height and location, and Fvis and Location. This leads to a model of the form: **Flush = Height + Fvis + Location + Height x Fvis + Height x Location + Fvis x Location + noise**. So this is a human disturbance effect, depending on location and height.

5. Age effect. Flush distance is mainly determined by Age (class) of the birds. This leads to a model of the form: **Flush = Age + noise**.

6. Age effect and human disturbance. Perhaps adult eagles respond in a different way to human disturbance than juveniles. This would lead to a model of the form: **Flush = Age + Fvis + Age x Fvis + noise**.

7. Flush distance is determined by location (perhaps the habitat is different). This leads to a model of the form: Flush = Location + noise, or perhaps better: **Flush = Location x Fvis + noise** if we hypothesize that the **Fvis** effect differs per study area.

8. If an eagle sits with lots of other eagles in a tree it may decide to be braver and wait longer before flying away. On the other hand, if one eagle flies away, then all others may follow. This may lead to models of the form: **Flush = NumEagles + noise**.

9. Distance to the river may be important, leading to a model of the form: **Flush = DisRiv + noise**. Perhaps there also an interaction with human disturbance leading to **Flush = DisRiv + Fvis + DisRiv x Fvis + noise**

Let us now increase the complexity of the models.

10. Flush distance is a function of Fvis, wheather, location and Age. This leads to a model of the form: **Flush = Fvis + Temp + Cloud + fWind + fRain + Location + Age + noise.** It makes sense to add the 2-way interactions between the weather variables, and also between Location and Fvis.

11. There is a human disturbance effect, location effect and age effect. And there is also a wind effect. This model takes the form: **Flush = Fvis + Location + Age + Wind + noise**.

12. Flush distance may depend on time (Year and time of day), number of eagles and distance to river, height and Fvis, age class and whether the eagle is breeding, location and weather variables. This leads to a model of the form: **Flush = Year + Time + NumEagles + DisRiv + Height + Fvis + Age + Breeding + fLocation + fWind + fRain + Temp + Cloud + noise**. There are no interactions between the variables.

13. And a simple model: Flush distance is not driven by any of these covariates. This is a model without covariates.

Surely various other hypotheses (and therefore models) can be formulated, but for the sake of showing an IT example, let us stop here. Before fitting any of these models, first apply a data exploration. You still need to address questions like whether there are outliers, collinearity and perhaps we need to transformation some variables. And we also need to verify whether all these variables can be used (are the variables balanced?) and the type of model that should be applied (e.g. regression).

After the data exploration….

After doing the data exploration, we would like you to apply the following 13 models. These models match the 12 hypotheses.

```
#Data exploration results:
Eagles2 <- na.exclude(Eagles)
Eagles3 <- Eagles2[Eagles2$Height < 900,]       #Forgot to recode NA
Eagles3$LogDisRiv <- log(Eagles3$DisRiv + 1)    #Outliers
Eagles4 <- Eagles3[Eagles3$fLocation != "W",]  #Unbalanced

M1 <- lm(Flush ~ Fvis, data = Eagles4)
M2 <- lm(Flush ~ (fWind + fRain + Temp + Cloud)^2, data = Eagles4)
```

```
M3 <- lm(Flush ~ Height + Fvis + Height : Fvis, data = Eagles4)
M4 <- lm(Flush ~ Height + Fvis + fLocation +
                 Height : Fvis +
                 Height : fLocation +
                 Fvis : fLocation, data = Eagles4)

M5 <- lm(Flush ~ Age, data = Eagles4)
M6 <- lm(Flush ~ Age + Fvis + Age : Fvis, data = Eagles4)
M7 <- lm(Flush ~ fLocation * Fvis, data = Eagles4)
M8 <- lm(Flush ~ NumEagles, data = Eagles4)
M9 <- lm(Flush ~ LogDisRiv * Fvis, data = Eagles4)
M10 <- lm(Flush ~ Fvis * fLocation +
                (Temp + Cloud + fWind + fRain)^2,
          data = Eagles4)
M11 <- lm(Flush ~ Fvis + Age + fLocation + fWind, data = Eagles4)
M12 <- lm(Flush ~ factor(Year) + Time + NumEagles +
                  LogDisRiv + Height + Fvis + Age +
                  Breeding + fLocation + fWind +
                  fRain + Temp + Cloud,
          data = Eagles4)
M13 <- lm(Flush ~ 1, data = Eagles4)
```

From each model extract the AIC, determine the model with the smallest AIC, and calculate the difference between the AIC of each model and the smallest AIC. We then convert these differences into Akaike weights w_i and they tell us the following: If we repeat the sampling process a large number of times then in w_i % of the cases model *i* is selected as the best model.

```
#The code to extract AIC and calculate Akaike weights is as follows
AICs <- AIC(M1, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12, M13)

MyDf <- AICs[,1]
AICsNum <- AICs[,2]
minAW <- min(AICsNum)
Delta <- AICsNum-minAW
RL <- exp(-0.5 * Delta)
wi <- RL / sum(RL)
Z <- data.frame(MyDf, AICsNum, Delta, wi)
Z <- round(Z, digits = 3)
colnames(Z)<- c("Df", "AIC", "AIC differences", "Akaike weights")
Z
```

Now present and discuss all models that have a reasonable high Akaike weight. Hence, there is no single best model but a set of potential (and competing) models, and each should be discussed. Apply a (quick) model validation and interpretation on each selected model.