Université Grenoble Alpes

TP : Simple and multiple regression, analysis of variance

1 Presentation

The goal of this TD is to implement the estimation and classical tests presentes in class for simple and multiple regression models and analysis of variance. We are going to work on the file bats.csv which presents data from a study on the brain size of bats. The study is presented in [Hutcheon et al., 2002]. The abbreviations given in the file are described in the article like Diet categories (1 = phytophage; 2 = gleaner; 3 = aerial insectivore; 4 = vampire), BOW = body mass, BRW = brain mass, AUD = auditory nuclei volume, MOB = main olfactory bulb volume ; HIP = hippocampus volume.

Open RStudio and read the data file bats.csv

```
myData <- read.table(file="bats.csv", sep=";", skip=3, header=T)
summary(myData)
```

2 Study of the relationship between brain weight and body mass

- Brain mass is expected to depend on the size of the bat under consideration and thus on its weight. For this purpose, we propose to set up an analysis focusing only on plant-eating bats. Create a data table, named phyto, containing the different variables measured only for the phytophagous.

  ```
  phyto=myData[(myData$Diet==1),]
  ```

  Then create a drawing that represents total brain weight as a function of body mass.
- We propose to fit a simple regression model to explain total brain weight as a function of body mass. For that R has the function lm, to use as in the following line:

  ```
  reg1 = lm(BRW ~ BOW, data=phyto)
  ```

- Write in mathematical form the model estimated by R. When this instruction has been used, R has adjusted the model and has stored all the elements necessary for the implementation of the tests seen in class. We obtain, for example, the values of the estimated parameters by the instruction :

  ```
  summary(reg1)
  ```

- What is the estimate of the intercept? What is the value of the test statistics for the model test? What is the H0 hypothesis of this test? What do the degrees of freedom 1 and 27 correspond to? What can be said about the relationship between brain weight and body mass? How much is the coefficient of determination?
- We obtain the analysis of variance table with the command :

  ```
  anova(reg1)
  ```

  What additional information is presented in this table? What is the sum of the residual squares? The variable reg1 is an object that contains, among other information, the values of the residuals and the values predicted by the model. We can therefore draw the graph of the residuals according to the predicted values, like a diagnostic graph.

```
plot(reg1$fitted.values, reg1$residuals, xlab="Predicted", ylab="Residuals")
```

- Discuss this graph. Identify the point that appears atypical and corresponds to a predicted brain weight approaching 10000 mg. Redo the analysis without this individual and record the results of the model in reg2. Compare the results obtained.

```
which(phyto$BRW>8000)
phytobis=phyto[which(phyto$BRW<8000),]

reg2 = lm(BRW ~ BOW, data=phytobis)
summary(reg2)
```

- R proposes by default four graphs to help in the diagnosis of the validity of the linear model approach. They are obtained with the commands :

```
par(mfcol=c(2,2))
plot(reg1)
```

Compare diagnostic graphs 2 and 3 obtained for reg1 and reg2 and discuss.

3 Study of the contribution to the total weight of each part of the brain

In this part, we try to understand the contribution of each part of the brain to the total weight. The variable to explain is the total weight of the brain (variable BRW). The potentially explanatory variables are the volume of the auditory part of the brain (variable AUD), the volume of the olfactory zone (MOB), and the volume of the hippocampus (HIP). Again, we work only with phytophagous species.

- We are initially interested in the correlations of the variables two by two. The `cor` function of R can be applied to an object of type data.frame if all the variables are of numerical type, and here we have factors. We will therefore create a new table containing only the variables of interest, calculate the correlation matrix and visualise it (if not loaded, use `library(corrplot)`

```
phytoNum=phyto[, c(4:8)]
mat.cor=cor(phytoNum)
corrplot(mat.cor, type="upper")
```

- Run the following tests :

```
cor.test(phyto$BRW,phyto$HIP)
cor.test(phyto$BRW,phyto$MOB)
cor.test(phyto$BRW,phyto$AUD)
```

What can we conclude ?

- Run the following multiple regression model

```
reg3=lm(BRW~AUD+MOB+HIP,data=phytobis)
summary(reg3)
```

Write the multiple regression model corresponding to this analysis. Discuss the validity of the linear model approach. What can be said about the relationship between total brain mass and these three variables. What is the coefficient associated with each variable? Discuss the relevance of the coefficient associated with the variable MOB.

- Test the following command

```
step(regm)
```

What is its purpose ? what is the conclusion ?

4 Link between volume of the auditory part and diet

The most important sense for an insectivorous bat is echo localisation, which is an auditory ability. On the contrary, a nectar-feeding bat must rather privilege capacities related to spatial memory, capacities governed by the hippocampus. We therefore propose now to study whether diet and sensory capacities are linked. To do so, we ask ourselves if the volume of the auditory lobe and the diet are related.

- Graph the auditory volume as a function of diet. To have a meaningful plot, we have to transform the diet data in factor. Comapre the two graphs.

```
myData$Diet_F = as.factor(myData$Diet)
with(myData, plot(AUD~Diet))
with(myData, plot(AUD~Diet_F))
```

- Do the regression analysis. It is important, again, to note that the type of the DIET variable is critical and R decides to do a regression or analysis of variance only on this type.

```
lm1 = lm(AUD~Diet, data=myData)
anova(lm1)
lm2 = lm(AUD~Diet_F, data=myData)
anova(lm2)
```

What can we conclude about the relationship linking auditory brain volume and diet? Is it surprising?

5 References

[Hutcheon et al., 2002] Hutcheon, J. M., Kirsch, J. A. W., and Garland, T. (2002). A Compara- tive Analysis of Brain Size in Relation to Foraging Ecology and Phylogeny in the Chiroptera. Brain, Behavior and Evolution, 60(3) :165-180.