Practical Assignment Regression 2019

Contents

This assignment is the practical assignment given in the course *Regression*. The assignment consists of an exploratory part, and an analysis part. The solution is to be made in groups of 2-4 persons, and must be in the form of an html or pdf report including R code, tables and figures, which is uploaded to Absalon.

The assignment will be made available on Thursday, February 21. The deadline for uploading the report is Sunday, March 3 at 23:59. Only one person from each group should upload the report. The names of all members in the group should be clearly stated on the front page. On Monday, March 4, the reports will be distributed among the groups, such that each group will have a report to read. You will then give feedback to the group and receive feedback on your own report no later than Thursday, March 7. All groups then have until Tuesday, March 12 to correct their report and upload a new version on absalon, which will be corrected by the instructors. If the report is not approved you will be notified by Thursday, March 14 and an improved version is to be handed in no later than Tuesday, March 19. The groups then present their reports during some of the exercise sessions in course weeks 6 and 7. Each report for group presentation will be assigned to another group, that has to read and prepare questions for the presentation.

There is no single "correct" solution. The evaluation of it will be based equally on three criteria; the formal correctness of the analyses carried out, how suitable and adequate the models and methods are, and the quality of the report.

The assignment is open ended. The data considered is biological data, and the objective is to fit a model to elucidate certain biological questions of interest. Your task is to analyze the data and come up with a suitable model. Remember to write in clear text what you do and the conclusions you draw along the way – though they may seem "obvious" from a plot or a table. Remember to report or visualize final models, to report parameter estimates or predictions including confidence intervals, and to give final conclusions. Remember also to discuss assumptions on which the analysis is based and justify to the extent it is possible those assumptions.

Narwhal diving data

The narwhal is the unicorn of the sea. Narwhals inhabiting high Arctic areas are seasonally exposed to extensive ice coverage with darkness prevailing in mid-winter and limited daylight for half of the year. Furthermore, narwhals dive extensively to depths much below the photic zone: in some areas they frequently dive to 1000 m and may occasionally dive below 1500 m. Narwhals therefore mostly depend on acoustics for sensing their environment, navigating the underwater icescape, capturing prey at depth, and communicating with conspecifics.

Changes in climate are rapidly modifying the Arctic environment, and human activities, such as noisy seismic exploration, are predicted to increase in remote areas of Greenland, including those inhabited by the narwhals of East Greenland. It is therefore of interest to explore these whales' diving and acoustic behavior and their reactions to anthropogenic sounds.

The sensitivity of narwhals to seismic exploration has been studied in an experiment with a seismic vessel (Paamiut) in East Greenland. The study was conducted in august 2017 and includes shooting with a small sized airgun at the summering ground of narwhals in Scoresby Sound, and instrumentation of 2 narwhals. The instruments are GPS-based satellite transmitters that provide detailed information about the movements

and positions of the whales, acoustic recorders that record the vocalizations of the whales, and activity recorders that quantifies the detailed movements of the whales.

The various measurements provide information about the positions of the whales every second, their diving activity, any vocalizations in the categories of calls, echolocation clicks and intensive foraging buzzes, as well as stroking rates and swimming activity. A series of trials were conducted, where the ship sailed into the proximity of the two tagged whales, and there were periods of continuous shooting of the airgun. In the data set it is registered in each second whether there is seismic activity (shooting of airgun), the distances of the whales to the seismic vessel, and whether the seismic vessel is in line of sight of the whale.

For the assignment we consider this data set on diving behaviour and sound production of narwhals in Scoresby Sund in Greenland. The data set can be downloaded as follows.

```
narwhal <- read.table(
  "http://www.math.ku.dk/~susanne/narwhal.txt",
  header = TRUE,
  colClasses = c("character", "character", "numeric", "factor", "factor",
                 "factor", "factor", "factor", "numeric", "factor",
                 "numeric", "factor", "factor", "numeric", "numeric",
                 "numeric", "numeric", "numeric", "factor")
  )
```

```
summary(narwhal)
```

```
##      Date               Hour               Depth           Seismik
##  Length:1058507     Length:1058507     Min.   :-741.5370   0:877819
##  Class :character   Class :character   1st Qu.:-114.9177   1:180688
##  Mode  :character   Mode  :character   Median : -14.7699
##                                        Mean   : -91.1954
##                                        3rd Qu.:  -0.8795
##                                        Max.   :   2.5829
##
##      Phase        Area        Call        Acou.qua    dist.to.Paamiut
##  B      :538654   F : 67430   0:1054122   B:   5330   Min.   :   -1
##  I2     :106278   G :144945   1:   4385   G:1053177   1st Qu.:   -1
##  I3     : 83027   IF: 30283                           Median :   -1
##  T1     : 78824   IG:772943                           Mean   : 4859
##  I4     : 54119   OG: 42906                           3rd Qu.: 5718
##  I5     : 47478                                       Max.   :50366
##  (Other):150127
##     Ind         dist.to.shore     Click      Buzz
##  Helge:689254   Min.   :   -1.0   0:833027   0:1043704
##  Thor :369253   1st Qu.:  630.7   1:225480   1:  14803
##                 Median : 1515.2
##                 Mean   : 1758.5
##                 3rd Qu.: 2474.0
##                 Max.   :12015.5
##
##       ODBA              VeDBA               Lat             Long
##  Min.   :-1.00000   Min.   :-1.00000   Min.   :-1.00   Min.   :-28.41
##  1st Qu.: 0.09301   1st Qu.: 0.06349   1st Qu.:69.99   1st Qu.:-27.35
##  Median : 0.14747   Median : 0.10109   Median :70.04   Median :-27.28
##  Mean   : 0.16917   Mean   : 0.11583   Mean   :69.38   Mean   :-26.89
##  3rd Qu.: 0.21490   3rd Qu.: 0.14637   3rd Qu.:70.20   3rd Qu.:-27.15
##  Max.   : 3.04904   Max.   : 1.91827   Max.   :70.52   Max.   : -1.00
##
```

```
##    Strokerate    los
##  Min.   :-1.00   0:847778
##  1st Qu.:21.00   1:210729
##  Median :26.00
##  Mean   :26.36
##  3rd Qu.:31.00
##  Max.   :78.00
##
```

The variables in the data set are:

- **Date**: Date of measurement.
- **Hour**: Hour of measurement.
- **Depth**: Depth of whale in meters below sea level.
- **Seismik**: Whether Paamiut produces seismic activity (shooting of airgun) or not (0 = no seismic activity, 1 = seismic activity).
- **Phase**: Code for the different phases of the experiment (B = before Paamiut arrives, I0 = Paamiut present before trial 0, T0 = trial 0 (Seismik), I1 = Paamiut present before trial 1, T1 = trial 1, I2 = Paamiut present before trial 2, T2 = trial 2, I3 = Paamiut present before trial 3, T3 = trial 3, I4 = Paamiut present before trial 4, T4 = trial 4, I5 = Paamiut present before trial 5, T5 = trial 5, I6 = Paamiut present, A = after the experiment is finished and Paamiut is not present).
- **Area**: The area where the whale is at time of measurement (IF= Inner Fønfjord, F = Fønfjord, OG = Outer Gåsefjord, G = Gåsefjord, IG = Inner Gåsefjord).
- **Call**: Whether the whale makes a call. A call is communication with other whales (0 = no calling, 1 = calling).
- **Acou.qua**: The quality of the acoustic recording (G = good, B = bad).
- **dist.to.Paamiut**: Distance of the whale to the seismic vessel in metres, NA = -1.
- **Ind**: Name of individual (Helge or Thor)
- **dist.to.shore**: Distance of the whale to the coast in metres, NA = -1.
- **Click**: Whether the whale makes a click. A click is echolocation (0 = no clicking, 1 = clicking).
- **Buzz**: Whether the whale makes a buzz. A buzz is for intensive foraging (0 = no buzzing, 1 = buzzing).
- **ODBA**: Overall dynamic body acceleration, NA = -1.
- **VeDBA**: Vector dynamic body acceleration, NA = -1.
- **Lat**: Latitude position of the whale, NA = -1.
- **long**: Longitude position of the whale, NA = -1.
- **Strokerate**: Strokerate of tail of the whale, NA = -1.
- **los**: An indicator of whether the whale is in line of sight of the ship (1 = the whale and Paamiut are within line of sight, 0 = the whale and Paamiut are not).

Exploratory analysis

The first task is an exploratory data analysis of the variables in the entire data set. Remember to be aware of any problems with the data like extreme or unrealistic observations for some variables, missing data, skewness and spread of marginal distributions, collinearity etc and take appropriate actions for such cases.

Note that the depth in Scoresby Sund is large, and except for the whale being very close to the shore, you can assume that the whale has "infinite" depth at its disposal, such that diving depth does not depend on the depth of the sea, but is rather decided by location of prey or other behavioral factors. Some measurements of depth are slightly positive, this is probably due to a wrong calibration of the measurement devices. You can choose to ignore it, or just set them to 0 or transfer all depth measurements such that max is 0.

Analysis

There are different biological questions that are of interest. You are not supposed to answer them all, but to construct a model and analyze the data that might help to elucidate some of these questions. These questions are:

- How can we characterize diving patterns of a whale?
- How can we characterize sound production of a whale?
- Are the whales affected by the presence of the seismic ship?
- Are the whales affected by the airgun shooting of the seismic ship?

You might decide to only analyze data from one of the two whales. You have to

- decide which variables you want to include in a model,
- decide which model you want to use,
- fit a model,
- explore possible inclusion of nonlinear effects or interaction effects,
- consider model diagnostics,
- report a final model,
- and make an interpretation of the results.