

BIOL343 – Assignment #5

Predicting bird abundance in fragmented habitats using multiple regression

Assigned Sunday 9 February

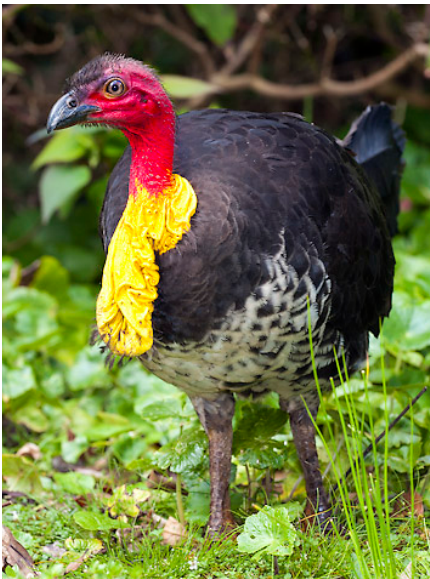
Due Saturday 15 February 1159pm

A major goal of ecology is to determine how human activities influence the diversity and abundance of plants and other animals. In this assignment, you will analyze data from a study by Australian ornithologist R.H. Loyn who evaluated the influence of landscape features on the abundance of birds in forest fragments in the state of Victoria, Australia.

Loyn quantified abundance of forest birds in 56 woodland fragment that varied widely in size as well as other variables that could affect bird abundance.



The data are in: **BirdsInForestPatches.csv**, which includes the following columns:



The response variable:

abundance = abundance of forest birds in each fragment

The potential predictors:

area_ha = fragment area (ha)

yr_isol = the year that the fragment was isolated from other forest by human habitat alteration

dist_nearest_km = distance to nearest other fragment (km)

dist_large_km = distance to nearest large fragment (km)

grazing = grazing intensity in fragment (1 to 5 indicating light to heavy)

altitude_masl = altitude (m above sea level) of the fragment

Use **dplyr**, **ggplot2**, **cowplot** and **ggfortify** functions whenever possible.

Your goal is to use simple linear regression and multiple regression to determine which of these potential predictors accounts for variation in forest bird abundance. In doing so, you will also evaluate whether the assumptions of regression analysis are met by these data.

We will leave model selection to the next assignment. In this assignment, you will evaluate the univariate regressions and then build a multiple regression, which you will visualize with added-variable plots. You will evaluate assumptions for both the univariate and multiple regressions.



Here's what Hana and I are expecting you do do.

1. The first step in any analysis is to get to know the data by running the `summary()` function.
2. It is probably a good idea to deal with the `yr_isol` variable which is the year each fragment was isolated. The study was conducted in 1985, so you can convert `yr_isol` to the number of years the fragment was isolated before the data were collected (`nyrs_isol`).
3. Next, plot the distribution of all the variables using histograms. Because you will make the same type of plot over and over again, you can avoid code-bloat by only writing down the code for the analysis once and switching up the variable under examination. For instance, you can define a new variable as one of the variables under examination: `birds$varb <- birds$area_ha` and simply run the code that refers to `varb` and then switch `varb` to the next variable in line. As a result, we only want to see one chunk of multi-purpose code a sample histogram for one of the variables. This figure can be utilitarian and does not need formal axis labels or a figure caption.
4. Although the predictor variables do not have to be normally distributed. It's a good idea to transform variables that are obviously log-normal. In the text part of your notebook describe any transformations that you think are good to do.
5. Follow this by a code chunk transforming the appropriate variables and the selecting only those columns from the dataframe that will go on to the next stage of analysis.
6. Next comes the univariate analyses. Regress abundance over each of the 6 potential predictor variables. Put these 6 univariate scatterplots in a single composite plot, with axes labeled appropriately and all axes aligned properly. Only put regression lines on plots when the relation with abundance is significant. There's a bunch of different ways of doing this in R, but I recommend checking out `plot_grid()` in the **cowplot** package. Make this graph as professional and thoughtfully laid out as possible. All axes of the individual graph panels must be tweaked and properly labelled (or de-labelled) using the `scale_...` functions, and the style of the graph should be modified with the `theme()` function. Include a suitable-for-publication figure caption.



7. Evaluate the significance of each regression using the `lm()` function. Again this requires repetitive analyses so use the code-reduction strategy suggested above. We only need to see one chunk of multi-purpose code for one of the predictors.
8. For each predictor use graphical analysis (a single function) to verify that the data fit the assumptions of regression. Again, use the code reduction strategy. We only need to see one chunk of multi-purpose code.
9. In a section of text, summarize the results of all the analyses that you did in #6, 7 & 8.
10. Now you will take all the predictors that have a significant relation with abundance and, using `lm()`, put them in a multiple regression model to explain abundance. This is not really how we build a multiple regression model but it will have to do until we learn model selection.
11. Use `pairs()` and `cor()` to evaluate the extent to which the predictors in #10 are intercorrelated (i.e. colinear).
12. In a section of text describe how the r^2 of the multiple regression model compares to the r^2 values of the univariate regressions involving the same predictors.
13. Again, use a one-function graphical analysis to evaluate whether the assumptions of multiple regression are met.
14. Use `ggplot` to make an added-variable (AV) plot for each predictor and put these graphs in a composite plot as above (#6). The first step is to calculate the appropriate x-axis and y-axis residuals for each predictor.
15. The second step is to make the AV plots and put them together in a composite plot using `plot_grid()`. This figure should also be beautiful and include a figure caption.
16. All the formatting guidelines introduced in previous assignments apply here.

Save the .html file produced by your R notebook as a PDF file and submit that to the OnQ dropbox as "StudentNumber_A5.pdf", where the file name starts with your student number, then an underscore, then the letter A (assignment) then 2, which is the assignment number, then the .pdf file extension. You should upload your file to the Assignment #5 OnQ dropbox by Saturday 15 February at 1159pm.

Before uploading the final PDF file, check it carefully to make sure that all components are included. There will be no regrading of incomplete assignment files that were mistakenly uploaded.

