Dear students,

In this project you will experience how a statistical investigation of a data set might be. The journey you will take is, this time, guided. This will take away some of the "fun" (struggles and eureka-moments) of statistical work. Nevertheless, I hope that I have shaped the project such that there is still place for some fun, as well as you will get a taste for statistical investigations.

The theme of the project is "factors affecting extinction". I hope you find the theme relevant and motivating. However, when you consider the data set and following exercises, you might deem that the data set and exercises does not match the theme. This is not the case, but you might need a bit more understanding of how research can consider and work with such a big theme. Each research study published on this theme is only a part, one piece, of the puzzle. One publication does never stand alone. This means that we will need a lot of research publications to see the full picture of the factors affecting extinction among birds. In research, this network of publications is important to acknowledge and also why most research publications begin with a section mapping previous research on the subject. In this project, you will not be tasked to make a literature background section, we will focus on the statistics and the statistical journey.

To consider the theme of factors affecting extinction we will consider a data set that measures breeding pairs of birds. The data in Table 1 are measurements on breeding pairs of land-bird species collected from 16 islands around Britain over the course of several decades. For each species, the data set contains an average time of extinction on those islands where it appeared (this is actually the reciprocal of the average of 1/T, where T is the length of time species remained on the island, and 1/T is taken to be zero if the species did not become extinct on the islands); the average number of nesting pairs (the average, over all islands where the birds appeared, of the number of nesting pairs per year); the size of the species (categorized as large or small); and the migratory status of the species (migrant or resident). (Data from S. L. Pimm, H. L. Jones, and J. Diamond, "On the Risk of Extinction," American Naturalist 132 (1988): 757-85).

It is expected that species with large numbers of nesting pairs will tend to remain longer before becoming extinct. Of interest is whether, after accounting for number of nesting pairs, size or migratory status has any effect. If any species have unusually small or large extinction times compared to other species with smaller values of the explanatory variables, it would be useful to point them out.

The following list of items to go through in your project is meant to (mimic and) guide you on a statistical investigation of the data set, to investigate possible factors influencing extinction in bird populations.

As this is a project to be handed in and orally presented, each step of your investigation should include an explanatory text. 4 combos, i.e. (L,R),(L,M),(S,R),(S,M)
filter data based on combo criteria, plot extinction time vs. numbers of pairs

1. Begin with describing and fitting a full model in which the intercepts and slopes of the extinction times versus numbers of pairs may be different in all four combinations of size and migratory status.

2. Examine the data set for possible transformations and outliers: Make a residual plot from the fit of the model. What can you tell based on the residual plot? (i.e. fit of the model/ possible transformations needed + outliers)?

3. Try the following transformations log("time"), sqrt("time") and 1/("time"). Which seems to be the best fit (examine residual plots)?

4. How does the outlier in each end of the scale affect the results? Argue for why to keep them in the data set.

5. Should the "pairs" variable be transformed: Make an informal assessment of whether or not there are linear relationships between log("time") versus "pairs" in all four combinations of "size" and "migratory status".

6. Make a test of the hypothesis that the slopes for all four combinations of "size" and "migratory status" are equal.

7. Make a reduced model based on your findings for previous item.

8. What will your conclusion be (look at the introductory text)?

Spieces	Ave. ext. time (years)	Ave. nr. of nesting pairs	Size (L or S)	Migratory status (resident or migrant)
Sparrowhawk	3.030	1.000	L	R
Buzzard	5.464	2.000	L	R
Kestrel	4.098	1.210	L	R
Peregrine	1.681	1.125	L	R
Grey partridge	8.850	5.167	L	R
Quail	1.493	1.000	L	M
Red-legged partridge	7.692	2.750	L	R
Pheasant	3.846	5.630	L	R
Water rail	16.667	3.000	L	R
Corncrake	4.219	4.670	L	M
Moorhen	8.1300	4.056	L	R
Coot	5.000	1.000	L	R
Lapwing	7.299	6.960	L	M
Golden plover	1.000	1.670	L	M
Ringed plover	27.027	5.560	L	R
Curlew	3.106	2.830	L	M
Redshank	4.00	4.375	L	M
Snipe	16.129	4.129	L	M
Stock dove	3.484	3.670	L	R
Rock dove	37.037	8.330	L	R
Wood pigeon	7.299	2.750	L	R
Cuckoo	2.525	1.430	L	M
Short-eared owl	4.132	2.000	L	R
Little owl	2.000	2.750	L	R
Magpie	10.000	4.500	L	R
Jackdaw	2.667	7.120	L	R
Carrion crow	4.587	4.580	L	R
Raven	58.824	2.350	L	R
Skylark	32.258	6.870	S	R
Swallow	2.571	3.830	S	M
House martin	2.160	5.000	S	M
Yellow wagtail	1.000	1.250	S	M
Pied wagtail	2.967	2.270	S	R
Meadow pipit	9.524	5.350	S	R
Wren	11.1111	8.700	S	R
Dunnock	7.299	6.100	S	R
Robin	4.000	3.330	S	R
Stonechat	2.381	3.640	S	R
Wheatear	2.611	4.830	S	M
Blackbird	3.257	4.670	S	R
Song thrush	1.701	1.700	S	R
Mistle thrush	1.795	1.330	S	R
Grasshopper warbler	1.198	1.000	S	M
Sedge warbler	3.185	1.900	S	M
Whitethroat	2.273	4.420	S	M
Willow warbler	1.111	1.250	S	M
Chiffchaff	1.000	1.000	S	M
Goldcrest	1.000	1.000	S	R
Spotted flycatcher	1.230	1.000	S	M
Great tit	6.061	2.500	S	R
Blue tit	3.175	1.500	S	R
Yellowhammer	2.000	2.500	S	R
Reed bunting	5.076	5.630	S	R
Chaffinch	1.934	2.370	S	R
Goldfinch	1.493	1.500	S	R
Redpoll	1.000	1.000	S	R
Linnet	5.102	6.500	S	R
House sparrow	3.003	4.500	S	R
Tree sparrow	1.898	2.170	S	R
Starling	41.667	11.620	S	R
Pied flycatcher	1.000	1.000	S	M
Siskin	1.000	1.000	S	R

Table 1: Bird extinction data