

MTHS4005 Coursework description

Student Number 20595239

An ornithologist asks for your assistance in analysing data they have collected on kittiwakes, a type of gull. There are two sub-species of kittiwake, black-legged and red-legged.

The ornithologist provides you with four data sets (.csv files). The data sets are as follows:

- **Observation data** The number of sightings of kittiwakes at an observation point over a 4 week (28 day) period. Each day observations are taken at dawn, noon, mid-afternoon and dusk.
- **Historical data** The number of breeding pairs at 5 sites in 6 different years.
- **Measurement data** The weight (g), wing span (cm) and culmen (beak length) (mm) is collected for 17 black-legged and 12 red-legged kittiwakes.
- **Location data** The number of breeding pairs in 26 colonies is recorded along with potentially important covariate information on mean summer temperature, cliff height (logarithm), sandeel concentration and coastal direction.

The ornithologist has asked for your insights on the following questions:

1. Provide an exploratory analysis of the **Observation data**.
Construct a 99% confidence interval for the mean number of kittiwakes observed at dusk.
20 Marks
2. Does the **Historical data** support the ornithologist's hypothesis that the decline in kittiwake numbers, over time, is independent of site?
The ornithologist would like an estimate for the number of breeding pairs at site Site D in 2014.
16 Marks
3. For the **Measurement data** the ornithologist asks for:
 - a) A visual summary of the data.
 - b) For each sub-species, is wing span and culmen length independent?
 - c) Is there evidence that the weights of birds of the two sub-species are different?
 - d) From the data provided is there evidence that there is a difference between the two sub-species?**24 Marks**

4. For the **Location data** the ornithologist asks you to:

- a) Fit a linear model to predict the number of breeding pairs.
- b) Fit a linear model to the logarithm of the number of breeding pair.
- c) Choose the most appropriate linear model for the data.
- d) Comment on the model fit and affect of the selected covariates on the number of breeding pairs.
- e) Choosing an appropriate model, provide a 98% confidence interval for the number of breeding pairs at a site with coastal direction = East, sandeel concentration = 1.84, mean summer temperature = 17.8 and cliff height (log) = 3.45.

20 Marks

In analysing the data sets clearly state any modelling assumptions you make and any hypotheses.

The remaining **20 Marks** are awarded for presentation. Presentation marks are awarded for:

- Structure of the report, including introduction, conclusion and suitable subsections.
- Grammar and presentation, including use of paragraphs, equations and labels for figures.