# Description and use of the oak leaf data set

Richard Stillman

05/12/2021

The purpose of this document is to describe the oak leaf dataset, and to overview setting research questions and answering these using the data.

## Overview of the dataset

Data were collected along the south-facing edge of an oak woodland. 100 leaves were collected randomly from trees on the edge of this woodland. 50 of these leaves were collected from the southern side of trees, that would be exposed to the sun (called sun leaves), and 50 from the northern side, that would be in shade (called shade leaves).Therefore, the dataset comprises 50 sun leaves and 50 shade leaves. Each row in the dataset represents a different leaf. Each column in the dataset represents a variable to describe the leaves.



Figure 1: Example of a leaf of an oak leaf collected on Bournemouth University campus

## Variables measured directly for the leaves

| Name of variable | Description of variable | Units |
|---|---|---|
| LeafType | Sun" if a sun leaf, "Shade"if a shade leaf | None (factor with two levels) |
| LeafArea_mm2 | Area | $mm^2$ |
| LeafLength_mm | Length | $mm$ |
| LeafMaxWidth | Maximum width | $mm$ |
| LeafThickness_mm | Thickness of leaf | $mm$ |

| Name of variable | Description of variable | Units |
|---|---|---|
| LeafEdgeLength_mm | Length of edge of leaf | $mm$ |
| LeafDryMass_mg | Dry mass of leaf | $mg$ |

## Other variables measured for the leaves

Two additional variables were calculated from the directly measured variables.

Leaf density calculated as:

$Leaf density = \frac{LeafDryMass}{LeafArea}$

Leaf indentation index calculated as

$LeafIndentationIndex = 1 - \frac{4\pi LeafArea}{LeafEdgeLength^2}$

This index has a value of zero if a leaf is circular and increases towards 1 as a leaf becomes increasingly indented.

| Name of variable | Description of variable | Units |
|---|---|---|
| Leaf density | $\frac{LeafDryMass}{LeafArea}$ | $mg\ mm^{-1}$ |
| Leaf indentation index | $1 - \frac{4\pi LeafArea}{LeafEdgeLength^2}$ | Dimensionless (between 0 and 1) |

## Steps to phrasing a research question

A good research question should have a non-obvious answer and may require some background knowledge to set. A question with an obvious answer could be: "Do leaves witha larger area have a longer edge length?". Some background details to help could include: sun leaves will be exposed to more sunlight;being exposed to more sun will have some benefits (e.g., in terms of the potential for photosynthesis) but may have some costs (e.g., the potential for overheating); heat loss can be increased by having a longer edge to area ratio.

## Alternative ways of answering research questions

There are many ways to do this, but two broad categories are to make comparisons between group / class variables or to make comparisons between continuous variables. An example of a group or class variable is $LeafType$. An example of a continuous variable is $LeafArea_m m2$. Analysis of variance can be used for comparisons between group / class variables, and regression for comparisons between continuous variable