

Suggestions for writing up a small observational study

Duncan Golicher

03/12/2021

Introduction

Leaves of trees differ with respect to their overall area and allometric relationships (Viscosi and Cardini 2011). Tree species may show phenotypic plasticity in response to shading that may be related to heat dissipation. (Vogel 1968).

Note that the introduction should focus on setting the scene for the work. I have included just two key phrases here with references that provide a start. Introductions should be concise and relevant

Aims

The study aimed to investigate the effect of shading on English oak (*Quercus robur* L.) leaf size and allometric relationships. It addressed the following questions.

1. How large is the difference in oak leaf area which may be related to a phenotypic response to shading, as estimated from a sample of leaves taken from trees on Bournemouth University campus?
2. Does this sample of leaves provide any evidence that the allometric relationship between oak leaf width and oak leaf length is affected by shading?
3. Add another question
4. Add another question

Note that while the overall framing of the study should be of general interest, the specific questions that the study can address should be more modest and strictly context specific for an observational study

Methods

Concisely explain how you obtained the data. If someone else collected the data then provide a place and date and refer to their technical documentation of the method used. You would ideally also show where the data were obtained from as a map, best placed in the supplementary material (Figure 3) so that the study could be fully reproduced by others. See the description of the data set provided by Richard in the supplementary materials.

Results

Differences in leaf area

Figure 1 shows the pattern of variability in leaf area with respect to leaf exposure to sun or shade. A one way anova confirmed the statistical significance of the difference in means $F(1,98) = 6.4, p=0.013$. The estimated

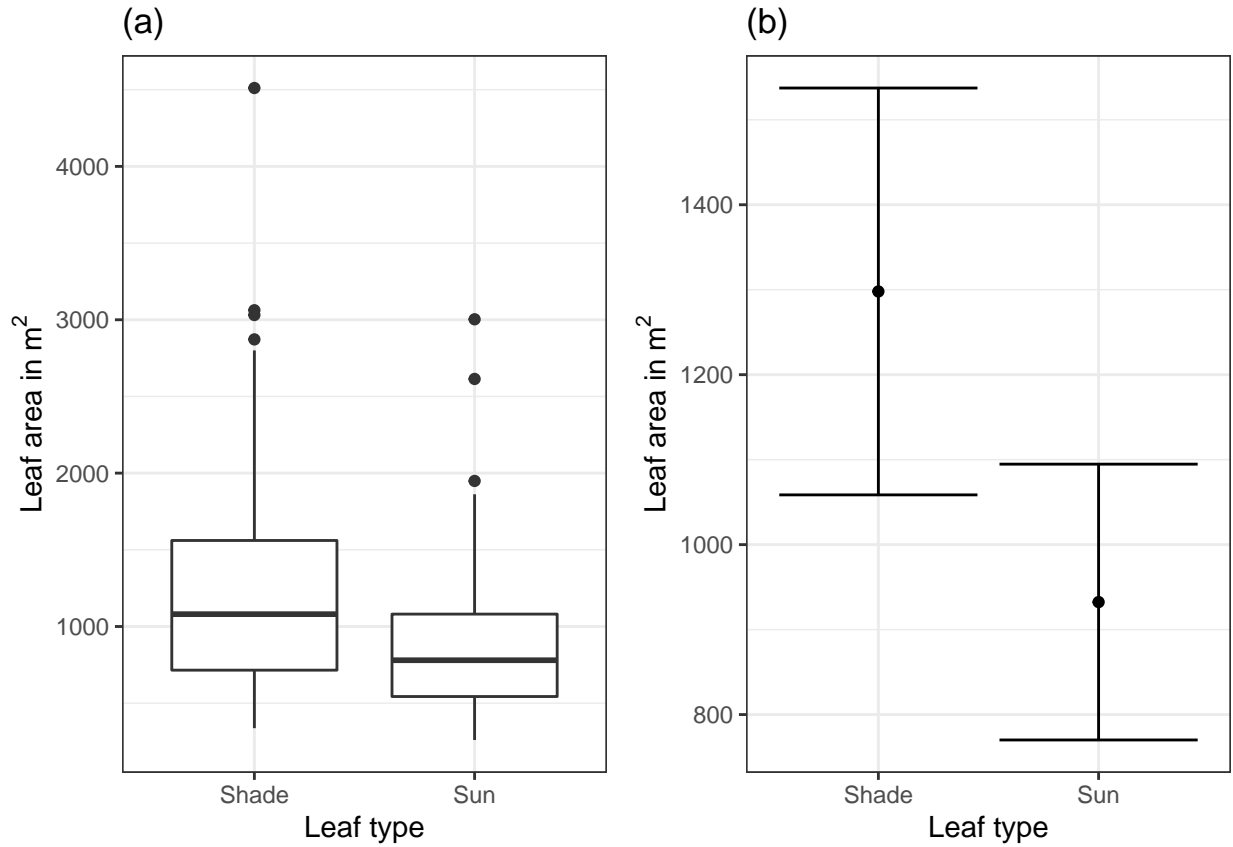


Figure 1: Area of oak leaves measured from sun exposed side of an oak tree ($n=50$) and shaded side ($n=50$) growing on Bournemouth University campus: (a) Leaf areas are approximately normally distributed. Six potentially influential outliers were identifiable, but not removed from the analysis. (b) An unpaired Welch corrected t-test found a statistical significant difference $T = 2.5$ $df = 86$ $p\text{-value} = 0.013$ with a 95% confidence interval of between 79 and 651 mm^2 in mean leaf area

increase in the mean area of shaded leaves was between 6% and 50% of the mean area estimated for the sun exposed leaves with a 95% confidence interval, adjusted for heterogeneity of variance.

Notice that you could use either anova or a t-test here. I've included one way anova as this is more general. If the p-value lies between 0.01 and 0.05, or just below, then include the actual value. If it is clearly less than 0.01 then just give it as $p < 0.01$ or $p < 0.001$. Don't go any deeper with the level of precision. If you had run multiple comparisons, because there were more than two levels, then choose the meaningful results from the HSD table. These will usually be the statistically significant differences, but not always. By running the anova you will have satisfied those who insist on seeing some p-values and you can now stress the confidence intervals)

Relationships between leaf length and leaf width

Describe the key elements of the results shown in figure 2 through interpreting the analysis of covariance in the supplementary material. Link the description to the question you posed in the aims.

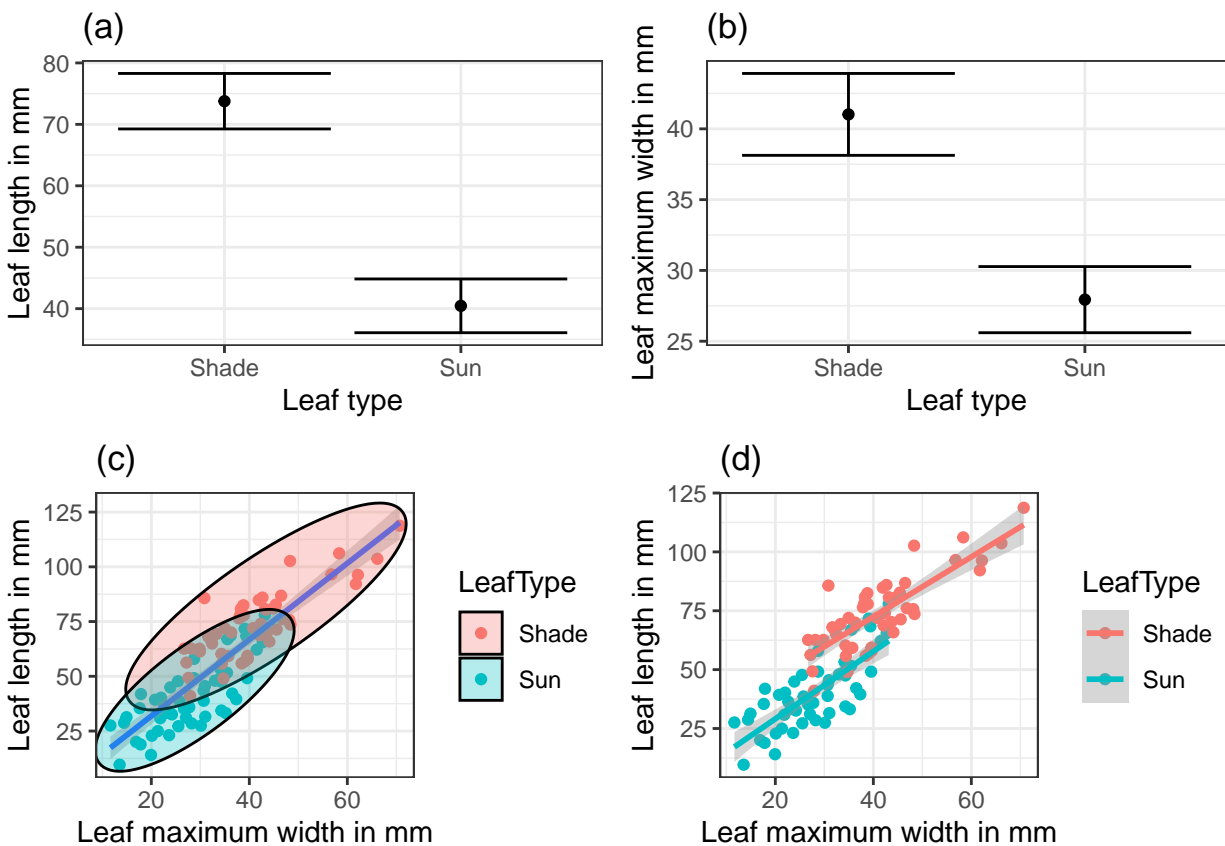


Figure 2: Relationship between maximum width and length of oak leaves measured from sun exposed side of an oak tree ($n=50$) and shaded side ($n=50$) growing on Bournemouth University campus: (a) Difference between mean leaf lengths was estimated to lie between 27 and 40 mm (95% confidence interval) (b) Difference between maximum leaf widths was estimated to lie between 9 and 17 mm (95% confidence interval) (c) Leaves were notably larger overall on the shaded side of the tree (also see figure 1). (d) No statistically significant difference was detected between the slopes of fitted linear regression lines conditioned on leaf type ($p > 0.1$)

Discussion

What does this imply?

What were the strengths and weaknesses of this particular study?

How do these results compare with other studies?

Notes

Many of the key results are best placed in figure captions. You can add the figure captions “by hand” in word, or if you want to automate the process in R add a line **carefully** into the top of the figure chunk. For example to produce figure 2 place this at the top of the chunk.

```
{r,echo=FALSE, fig.cap = "Relationship between maximum width and length of oak leaves measured from sun exposed side of an oak tree (n=50) and shaded side (n=50) growing on Bournemouth University campus: (a) Difference between mean leaf lengths was estimated to lie between 27 and 40 mm (95% confidence interval) (b) Difference between maximum leaf widths was estimated to lie between 9 and 17 mm (95% confidence interval) (c) Leaves were notably larger overall on the shaded side of the tree (also see figure 1). (d) No statistically significant difference was detected between the slopes of fitted linear regression lines conditioned on leaf type (p>0.1)"} }
```

The multiplot function is included in the aqm package. It can be used to group the plots if you assign each plot to an object within the code chunk. This can be a little fiddly to use in practice and it is not required for the assignment. However once mastered it does produce neater results. Aim to make the figure captions as informative as possible (within reason) so that anyone looking at the figure can spot the main points without recourse to the text. Keep the text concise within the results section. Add derived quantitative results that address the question when these are not immediately obvious from the figures. Refer to the figures to support these statements. Respect the intelligence of the reader. A modest data set such as this will still have a few interesting elements to it, but there is no point in padding out the results section with unnecessary trivial details. Keep it as short and to the point as possible. Use p-values with discretion. They can convey information when doubts arise, but often do not. Effect sizes (such as difference in leaf area) often are best expressed as percentages. These can be calculated by R code, but unless the analysis is going to be run frequently on new data it is often quicker just to use a calculator (or the R console as a calculator) rather than trying to code up the calculations in the markdown.

```
## R code to produce figure 1
library(tidyverse)
library(ggforce) # This is used for the ellipses shown in figure 2.
library(aqm) # Used for the data and multiplot function

data("oak_leaf_data")
theme_set(theme_bw())
g0 <-ggplot(d,aes(x=LeafType,y=LeafArea_mm2))
g1<-g0 + geom_boxplot() +labs(title="(a)",y = bquote("Leaf area in"~m^2), x="Leaf type")
g2 <-ci(g0) +labs(title="(b)",y = bquote("Leaf area in"~m^2), x="Leaf type")
multiplot(g1,g2, cols=2)
```

```
## R code to produce figure 2
# To put this all together you would run a code chunk for each figure
# whilst investigating the data, then collect them later.
# Make sure to use a different name for each plot.

g0<-ggplot(d,aes(x=LeafType, y=LeafLength_mm))
g1<-ci(g0)
g1<-g1 +labs(title="(a)",x = "Leaf type", y="Leaf length in mm")

g0<-ggplot(d,aes(x=LeafType, y=LeafMaxWidth_mm))
```

```

g2<-ci(g0)
g2<-g2 +labs(title="(b)",x = "Leaf type", y="Leaf maximum width in mm")

g0<-ggplot(d,aes(x=LeafMaxWidth_mm, y=LeafLength_mm))
g3<-g0 + geom_point(aes(col=LeafType)) + geom_smooth(method="lm")
g3 <- g3+ geom_mark_ellipse(expand=0.02,aes(fill=LeafType))
g3<-g3+labs(title="(c)", y = "Leaf length in mm", x="Leaf maximum width in mm")

g0<-ggplot(d,aes(x=LeafMaxWidth_mm, y=LeafLength_mm, colour=LeafType))
g4<-g0 + geom_point() + geom_smooth(method="lm")
g4<-g4+labs(title="(d)", y = "Leaf length in mm", x="Leaf maximum width in mm")

multiplot(g1,g3, g2,g4, cols=2)

```

Supplementary material

Welch corrected t-test for the difference between mean area of shaded and unshaded leaves. (R Core Team 2020)

```

t.test(d$LeafArea_mm2~d$LeafType)

##
## Welch Two Sample t-test
##
## data: d$LeafArea_mm2 by d$LeafType
## t = 2.5397, df = 86.187, p-value = 0.01289
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 79.43395 651.68605
## sample estimates:
## mean in group Shade mean in group Sun
## 1297.96 932.40

```

An anova can also be used in much the same way. Note that this would not correct for any heterogeneity in variance automatically, but this is rarely of importance

```

mod<-lm(d$LeafArea_mm2~d$LeafType)
anova(mod)

## Analysis of Variance Table
##
## Response: d$LeafArea_mm2
## Df Sum Sq Mean Sq F value Pr(>F)
## d$LeafType 1 3340853 3340853 6.4503 0.01267 *
## Residuals 98 50757916 517938
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

White's correction can be run (Zeileis 2006) , but is ignorable as the degrees of freedom remain unchanged. The p values rounded to two significant digits are the same in all cases (0.013).

```
library(sandwich)
library(car)
Anova(mod,white.adjust='hc3')
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: d$LeafArea_mm2
##           Df      F Pr(>F)
## d$LeafType  1  6.3213 0.01356 *
## Residuals  98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
mod<-lm(data=d,LeafLength_mm~LeafType)
confint(mod)
```

```
##           2.5 %    97.5 %
## (Intercept) 69.37984 78.15776
## LeafTypeSun -39.51594 -27.10210
```

```
mod<-lm(data=d,LeafMaxWidth_mm~LeafType)
confint(mod)
```

```
##           2.5 %    97.5 %
## (Intercept) 38.42304 43.614482
## LeafTypeSun -16.75261 -9.410795
```

Analysis of covariance is a handy way of looking at whether slopes differ.

```
mod<-lm(data=d,LeafLength_mm~LeafMaxWidth_mm*LeafType)
anova(mod)
```

```
## Analysis of Variance Table
##
## Response: LeafLength_mm
##           Df Sum Sq Mean Sq F value    Pr(>F)
## LeafMaxWidth_mm      1  38486   38486 410.3547 < 2.2e-16 ***
## LeafType              1   4164    4164  44.4035  1.68e-09 ***
## LeafMaxWidth_mm:LeafType  1     51     51   0.5486   0.4607
## Residuals            96   9004     94
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
confint(mod)
```

```
##           2.5 %    97.5 %
## (Intercept) 10.2572803 33.0546852
## LeafMaxWidth_mm      1.0005922  1.5403338
## LeafTypeSun         -36.1491004 -6.1846741
## LeafMaxWidth_mm:LeafTypeSun -0.2692417  0.5897968
```



Figure 3: Oak tree sampled for the study on the Bournemouth University Campus, October 12 2020

Description of the oak leaf dataset, and an overview of setting research questions and answering these using the data.

Richard Stillman

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.

Variables measured directly for the leaves

Variables measured directly on 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:

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

Leaf indentation index calculated as

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

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	$\frac{\text{LeafDryMass}}{\text{LeafArea}}$	mg mm ⁻¹
Leaf indentation index	$1 - \frac{4\pi \text{LeafArea}}{\text{LeafEdgeLength}^2}$	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 with a 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_m2*. Analysis of variance can be used for comparisons between group / class variables, and regression for comparisons between continuous variable

References

- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Viscosi, Vincenzo, and Andrea Cardini. 2011. “Leaf Morphology, Taxonomy and Geometric Morphometrics: A Simplified Protocol for Beginners.” Edited by Carles Lalueza-Fox. *PLoS ONE* 6 (10): e25630. <https://doi.org/10.1371/journal.pone.0025630>.
- Vogel, Steven. 1968. “Sun Leaves and Shade Leaves: Differences in Convective Heat Dissipation.” *Ecology* 49 (6): 1203–4. <https://doi.org/10.2307/1934517>.
- Zeileis, Achim. 2006. “Object-Oriented Computation of Sandwich Estimators.” *Journal of Statistical Software* 16 (9). <https://doi.org/10.18637/jss.v016.i09>.



Figure 4: Example of a leaf of an oak leaf collected on Bournemouth University campus