

# Data Analysis Report - your title

Your Names

2024-04-05

## Abstract

Write a summary of your project (around 300 words) here. This section should provide a concise overview of your research question, data, methods, and key findings.

## Introduction

In this section, introduce the problem and question(s) of interest that you aim to address in your analysis. Provide relevant background information and explain the motivation and significance of your research question. Cite manuscripts or websites you use. Here an example of how to cite a paper, Wickham et al. (2019). See “Quarto - Citations & Footnotes” for how to manage citations in Quarto.

## Data and Methods

Describe the data you used in your analysis, including the source, characteristics, and any pre-processing or cleaning steps you performed.

This is a good place where to include results of your exploratory data analysis. See Figure 1 as an example. Check the .qmd file for an example of how to include plots, control their size and include references in the text.

You should also explain the methods you employed. I expect you to define the statistical model you chose for your data in mathematical form, as we did throughout the course.

Crab example: we define as outcome the presence/absence of a satellite for each crab and denote the observed values as  $y_i$  with  $i = 1, \dots, n$  with  $n = 173$ . Let  $y_i = 1$  indicate presence of a satellite and 0 otherwise.

We assume that each  $y_i$  is a realization of a random variables  $Y_i \sim \text{Bernoulli}(p_i)$  independently, and model the probability of having a satellite for each crab using the following logistic regression:

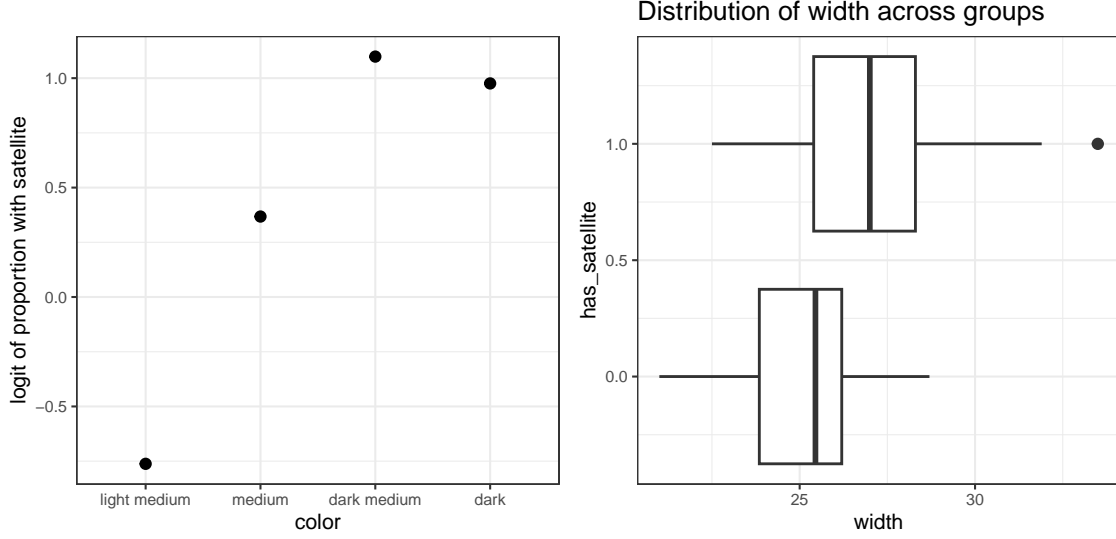


Figure 1: Example figure showing two plots side by side; here in the caption you should write a brief description of what the two plots represents: “Left side panel shows the logit of the proportion of crabs with satellites for each group of crabs with different color. The right side panel shows the distribution of the width among crabs with and without satellites.’’

$$\begin{aligned}
 \text{logit}(p_i) = & \beta_0 + \\
 & \beta_1 \times \text{I}(\text{color} == \text{medium})_i \\
 & \beta_2 \times \text{I}(\text{color} == \text{dark medium})_i \\
 & \beta_3 \times \text{I}(\text{color} == \text{dark})_i \\
 & \beta_4 \times \text{I}(\text{width})_i
 \end{aligned} \tag{1}$$

where  $\text{I}()$  in Equation 1 indicates a dummy variable, taking value 1 when the condition is true and zero otherwise.

## Results

Present your results in this section, including figures, tables, and any other relevant outputs from your analysis.

To display the model summary in a nice format you can use the `summ()` function from the `jtools` package.

Other libraries:

- `xtable`: good if you are familiar with latex (see [xtable documentation](#) for examples)
- `gttable`, implements a “grammar of table” idea (see <https://gt.rstudio.com> for documentation)

Table 1: Logistic regression summary

|                  | Est.   | S.E. | z val. | p    |
|------------------|--------|------|--------|------|
| (Intercept)      | -12.72 | 2.76 | -4.60  | 0.00 |
| colormedium      | 1.11   | 0.59 | 1.87   | 0.06 |
| colordark medium | 1.33   | 0.85 | 1.56   | 0.12 |
| colordark        | 1.40   | 0.55 | 2.56   | 0.01 |
| width            | 0.47   | 0.11 | 4.43   | 0.00 |

Standard errors: MLE

In this section, discuss and interpret your results. Explain the implications and significance of your findings, and relate them to your research question and existing literature or theories.

## Conclusion

Summarize your main findings and conclusions in this section. Highlight the key takeaways and contributions of your analysis, and provide recommendations or suggestions for future research or practical applications based on your results.

## References

“Quarto - Citations & Footnotes.” <https://quarto.org/docs/authoring/footnotes-and-citations.html#bibliography-generation>.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Gromlund, et al. 2019. “Welcome to the Tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.

## Appendix (Optional)

If you want to include your R code as an appendix, you can create a new code chunk and set `#| echo: true` to show the code and `#| eval: false` to avoid that to be run. You can also provide scripts separately as supplementary material.

```
# Your code here
```