Data Analysis for Food Science

2024-05-09

Table of contents

Preface

During the past decades the production of data in relation to research, production, consumer behavior, social network etc. has increased dramatically. Today we are faced with data structures which were unimaginable just 50 years ago. Traditionally, a system under investigation were characterized by a few samples associated with say one to five descriptors and, carefully selected, responses. Today all aspects of the classical system interrogation has blown up, such that we have many more samples (e.g. production monitoring every minute), more descriptors (e.g. consumer characteristics), and by far more response variables (e.g. high throughput omics technologies). Tools developed for handling traditional scenarios still pertain the corner of how to approach today's data analytical challenges, however, by the development of computers, it is possible to carry out challenging mathematical procedures in no time and further produce visual graphics as resources for **translating information into knowledge**. Due to this fact, the traditional tools has gotten a makeover and new tools has been developed.

Food is, as such, an extremely inherent part of the human life, although one could argue that so is e.g. cardiovascular biology and governmental policy making, these subjects either work autonomously or does not demand everyday mental capacity. Everyday all humans need to eat- and drink in some social context, pay attention to the perception of the meal, and further deal with the possible health- and emotional implications of this process. When studying food science all these aspects are relevant.

Food science constitute a broad range of disciplines spanning controlled artificial model systems, over functional modification of real food matrices, production technology, to the relation between food- and meal composition, taste, perception and health. **All by means of data.**

These notes are thought to cover data analysis within food science. That is to; provide a general understanding of the purpose of data analysis, found a theoretical- and practical basis for understanding various numerical and graphical tools and couple generic tools to concrete issues within related disciplines. To this end by theory, examples and exercises.

The book material used in these notes are mostly from the notes for the course; Introduction to Statistics at DTU by P.B. Brockhoff and co workers. Additionally there are relevant chapters from other sources. All exercises are custom made and deal with real problems within food science. Welcome to the course in Fødevaredataanalyse for second year bachelor students in Food Science and Technology - Hope you will enjoy learning about how to use data for getting insight on food systems.

Morten Arendt Vils Rasmussen
August 2024

1 Introduction

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

A histogram can be useful for investigating empirical distributions Figure 1.1.

```
set.seed(4321)

x <- rnorm(100)

ggplot() +
  geom_histogram(aes(x),
                 colour = "white",
                 fill = "steelblue") +
  theme_bw()
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

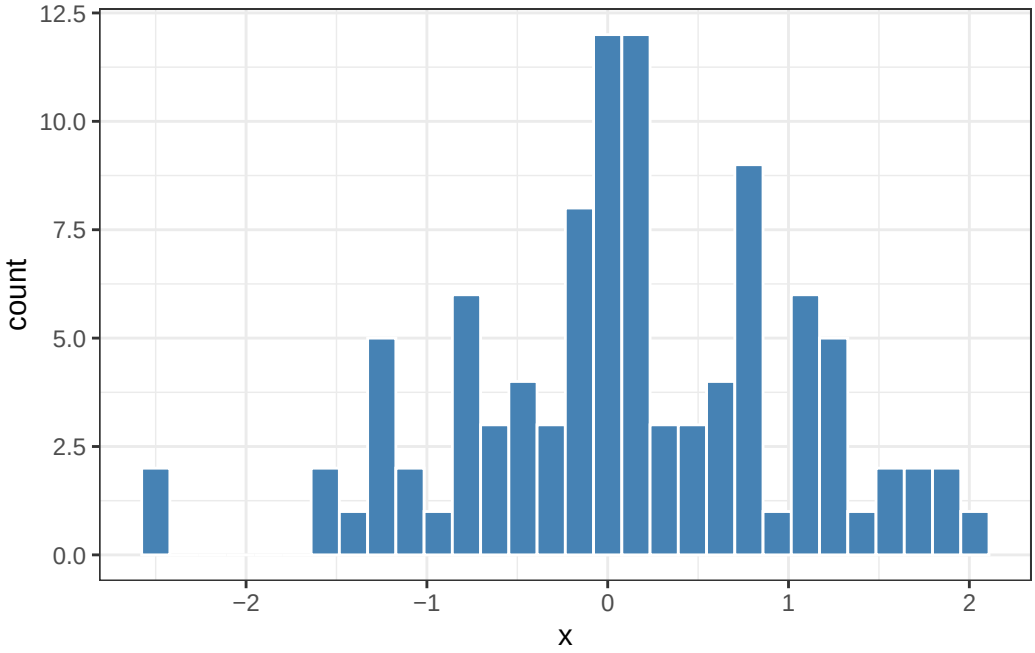


Figure 1.1: This is a histogram

An empirical cumulative distribution is also an effective tool Figure 1.2.

```
plot(ecdf(x))
```

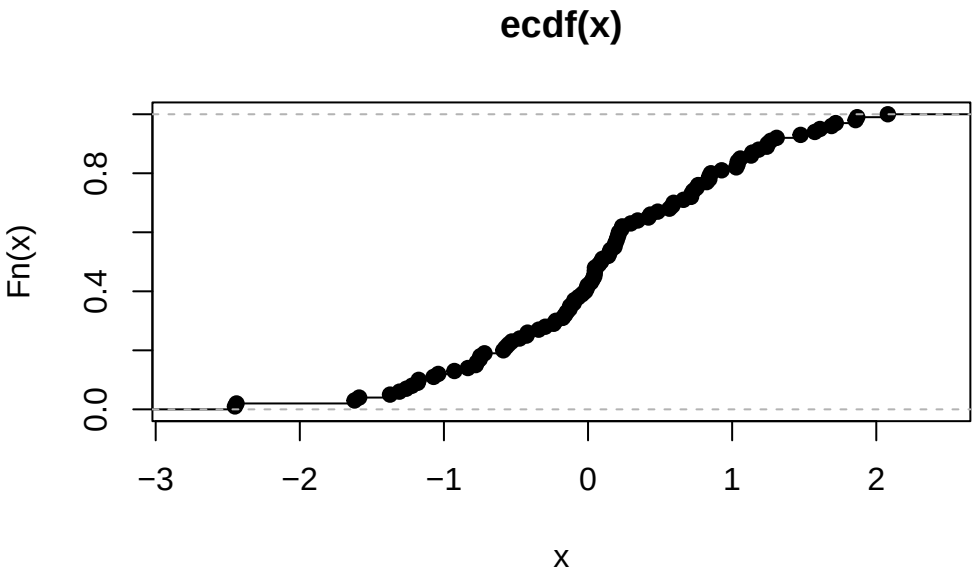


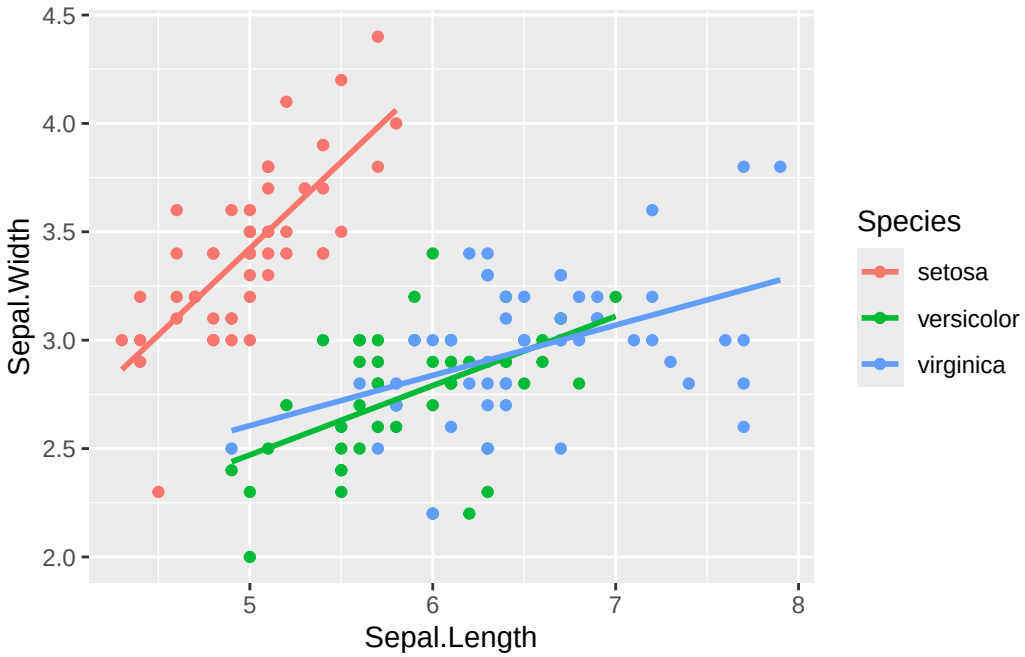
Figure 1.2: This is a ecdf.

2 Summary

In summary, this book has no content whatsoever.

```
ggplot(data = iris,
       aes(x = Sepal.Length,
           y = Sepal.Width,
           color = Species)) +
  geom_point() +
  stat_smooth(method = lm,
              se = F)
```

```
`geom_smooth()` using formula = 'y ~ x'
```



References

Knuth, Donald E. 1984. "Literate Programming." *Comput. J.* 27 (2): 97–111. https://doi.org/10.1093/comjnl/27.2.97.