

R programming

[Modern Statistics for Modern Biology](#)

Sections

[Home](#)

[Note to readers](#)

[Code and data](#)

[Physical Copy](#)

Modern Statistics for Modern Biology

AUTHORS

Susan Holmes

Wolfgang Huber

Home

Please navigate between and within the chapters of this book via the dropdown menu ("Chapters") at the top.

How to cite this book: Modern Statistics for Modern Biology, by Susan Holmes and Wolfgang Huber, Cambridge University Press (2019). ISBN: 9781108705295.

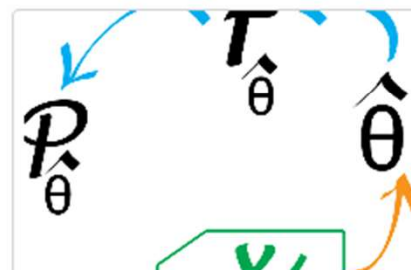
License: [CC BY-NC-SA](#)



Introduction



Generative Models for Discrete Data



Statistical Modeling

Sections

[Home](#)

[Note to readers](#)

[Code and data](#)

[Physical Copy](#)

If you spot anything, please alert us via email to [wolfgang.huber \[at\] embl.org](mailto:wolfgang.huber[at]embl.org).

Cat

All (

Code and data

Package Installation

You can install all the packages required to run every code example in the book via the following command:

```
source("https://www.huber.embl.de/msmb/install_packages.R")
```

Data

- [Data files \(zipped folder\)](#)

last modified: 2023-05-23

md5 hash: `c088c01844487695da215ee9bafdc360`

Code

- [R code \(zipped folder\)](#)

Download an electronic copy of the whole book

- [HTML tree for the book \(zipped\)](#)

```
#source("https://www.huber.embl.de/msmb/install_packages.R")
#절대로 하지 않기
wget https://www.huber.embl.de/msmb/data.tar.gz
```

```
#python
import tarfile
src = "data.tar.gz"
dst = "clean_data.tar.gz"
```

```
with tarfile.open(src, "r:gz") as tin:
    with tarfile.open(dst, "w:gz") as tout:
        for m in tin.getmembers():
            m.name = m.name.lstrip("./")
            f = tin.extractfile(m) if m.isfile() else None
            tout.addfile(m, f)
```

```
tar -xvzf clean_data.tar.gz
```

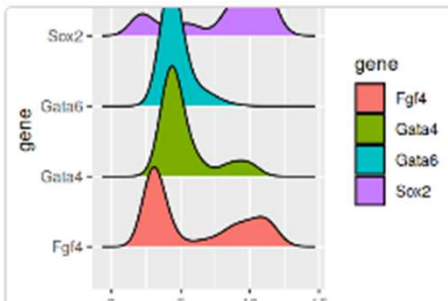
Visualization

```
install.packages("ggplot2")
install.packages("BiocManager")
BiocManager::install("rrcov")
install.packages("remotes")
remotes::install_version(
    "MASS",
    version = "7.3-60"
)
install.packages("dplyr")
```

```
library(tibble)
```



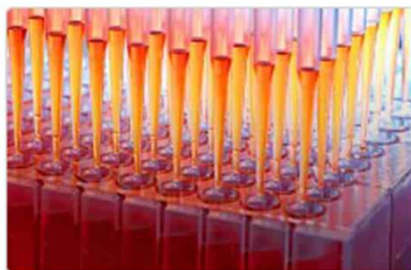
Generative Models for
Discrete Data



Data visualization



Clustering



Testing



Supervised Learning

-1.3. 1.3.1,1.3.2,1.3.3 (poisson, binomial)

-3.2, 3.4 (graph plot, ggplot2)

-5.3.1 (k-mean, hierarchical, dbscan)

-6.2 (t.test, wilcox.test, p.adjust(pvalue, method = "BH"))

-12.2.1 (lda, sva, linear, logistic ...)

sva(data,...)

lm(formula, data,)

```
ctl <- c(4.17,5.58,5.18,6.11,4.50,4.61,5.17,4.53,5.33,5.14)
```

```
trt <- c(4.81,4.17,4.41,3.59,5.87,3.83,6.03,4.89,4.32,4.69)
```

```
group <- gl(2, 10, 20, labels = c("Ctl","Trt"))
```

```
weight <- c(ctl, trt)
```

```
lm.D9 <- lm(weight ~ group)
```

glm(formula, data, family = binomial)