Bio Data Science

The definitive guide

Prof. Dr. Jochen Kruppa-Scheetz

2025-10-26

Table of contents

Welcome

"Life is difficult." — M. Scott Peck's, The Road Less Traveled

This cookbook contains an extensive collection of recipes from the fields of biostatistics, biometrics, statistics, bioinformatics and R programming. These will enable you to gain in-depth knowledge of these subject areas for data analysis, eliminating the need to attend my courses. Feel free to browse through the book to see if anything interests you, and then put together your own menu. I am continuously updating the book and adding to it all the time. In addition to the written content, there are also explanatory YouTube videos available. Either way, I am delighted that you are interested in learning something new, whether through personal interest or exam preparation. Either way, I recommend that you just take a look around. Don't be put off by the sheer volume of material – it just happened that way.

A German version of this OpenBook Bio Data Science is available. The page can be found at the following link: https://jkruppa.github.io/. However, it is much more unstructured. It is more like a cookbook than this book, which has a continuous storyline.

"If there's a book that you want to read, but it hasn't been written yet, then you must write it." — Toni Morrison

> **German Version**
>
> This is the English version of my German Openbook Bio Data Science.

Part I

From science to data by models

Last modified on 26. October 2025 at 20:43:25

> *"What problem have you solved, ever, that was worth solving where you knew all the given information in advance? No problem worth solving is like that. In the real world, you have a surplus of information and you have to filter it, or you don't have sufficient information and you have to go find some."* — *Dan Meyer in Math class needs a makeover*

Here comes the preface text

1 What is science?

Last modified on 26. October 2025 at 20:42:28

> *"A quote." — Dan Meyer*

1.1 Generale background

1.2 Theoretical background

1.3 R packages used

1.4 Data

1.5 Alternatives

Further tutorials and R packages on XXX

1.6 Glossary

term what does it mean.

1.7 The meaning of "Data Science" in this chapter.

- itemize with max. 5-6 words

1.8 Summary

References

2 What is data?

Last modified on 26. October 2025 at 20:42:22

> *"A quote." — Dan Meyer*

2.1 Generale background

2.2 Theoretical background

2.3 R packages used

2.4 Data

2.5 Alternatives

Further tutorials and R packages on XXX

2.6 Glossary

term what does it mean.

2.7 The meaning of "Data Science" in this chapter.

- itemize with max. 5-6 words

2.8 Summary

References

3 What is a model?

Last modified on 26. October 2025 at 20:42:18

> *"A quote." — Dan Meyer*

3.1 Generale background

3.2 Theoretical background

3.3 R packages used

3.4 Data

3.5 Alternatives

Further tutorials and R packages on XXX

3.6 Glossary

term what does it mean.

3.7 The meaning of "Data Science" in this chapter.

- itemize with max. 5-6 words

3.8 Summary

References

Part II

Template Preface

Last modified on 25. October 2025 at 20:17:25

> *"What problem have you solved, ever, that was worth solving where you knew all the given information in advance? No problem worth solving is like that. In the real world, you have a surplus of information and you have to filter it, or you don't have sufficient information and you have to go find some." — Dan Meyer in Math class needs a makeover*

Here comes the preface text

4 Template chapter

Last modified on 26. October 2025 at 13:57:25

> *"A quote." — Dan Meyer*

4.1 Generale background

4.2 Theoretical background

4.3 R packages used

4.4 Data

4.5 Alternatives

Further tutorials and R packages on XXX

4.6 Glossary

term what does it mean.

4.7 The meaning of "Data Science" in this chapter.

- itemize with max. 5-6 words

4.8 Summary

References