Models of Reality

From Data to Science

Prof. Dr. Jochen Kruppa-Scheetz

2025-11-18

Table of contents

Preface **3**
References . . . 3

I From science to data by models **4**

1 What is science? **6**
1.1 General background . . . 7
1.2 Theoretical background . . . 8
1.3 R packages used . . . 8
1.4 Data . . . 8
1.5 Alternatives . . . 8
1.6 Glossary . . . 8
1.7 The meaning of “Models of Reality” in this chapter. . . . 8
1.8 Summary . . . 8
References . . . 8

2 What is data? **9**
2.1 General background . . . 9
2.2 Theoretical background . . . 9
2.3 R packages used . . . 9
2.4 Data . . . 9
2.5 Alternatives . . . 10
2.6 Glossary . . . 10
2.7 The meaning of “Models of Reality” in this chapter. . . . 10
2.8 Summary . . . 10
References . . . 10

3 What is a model? **12**
3.1 General background . . . 12
3.2 Theoretical background . . . 13
3.3 R packages used . . . 13
3.4 Data . . . 13
3.5 Alternatives . . . 13
3.6 Glossary . . . 13

3.7 The meaning of "Models of Reality" in this chapter. . . . 13
3.8 Summary . . . 14
References . . . 14

II Tales of data **15**

4 Factorial data **17**
4.1 General background . . . 17
4.2 Theoretical background . . . 17
4.3 R packages used . . . 17
4.4 Data . . . 17
4.5 Alternatives . . . 18
4.6 Glossary . . . 18
4.7 The meaning of "Models of Reality" in this chapter. . . . 18
4.8 Summary . . . 18
References . . . 18

5 Covariate data **19**
5.1 General background . . . 19
5.2 Theoretical background . . . 19
5.3 R packages used . . . 19
5.4 Data . . . 19
5.5 Alternatives . . . 19
5.6 Glossary . . . 19
5.7 The meaning of "Models of Reality" in this chapter. . . . 19
5.8 Summary . . . 19
References . . . 19

6 Combinded data **20**
6.1 General background . . . 20
6.2 Theoretical background . . . 20
6.3 R packages used . . . 20
6.4 Data . . . 20
6.5 Alternatives . . . 20
6.6 Glossary . . . 20
6.7 The meaning of "Models of Reality" in this chapter. . . . 20
6.8 Summary . . . 20
References . . . 20

III Template Preface **21**

7 Template chapter **23**
7.1 General background 23
7.2 Theoretical background 23
7.3 R packages used 23
7.4 Data 23
7.5 Alternatives 23
7.6 Glossary 23
7.7 The meaning of "Models of Reality" in this chapter. 23
7.8 Summary 24
References 24

Preface

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

It's too early to write a preface. I started writing this book a few weeks ago. From my perspective, it will be a good one, but it will take some time. In the meantime, you are welcome to see what's happening here and what I'm writing. Sometimes things will be in the wrong place. This is a normal part of the writing process. Just so you know, I started writing on 15 November 2025.

Text Smolin (2019)

> *"The first principle is that you must not fool yourself — and you are the easiest person to fool. So you have to be very careful about that."* — Richard Phillips Feynman and Leighton (1992)

Richard P. Feynman (1998)

This book has a German predecessor. You might not recognise this book. That's fine. 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.

References

Part I

From science to data by models

Last modified on 27. October 2025 at 13:16:42

"I'd like to solve the puzzle." — Wheel of Fortune

Here comes the preface text

1 What is science?

Last modified on 17. November 2025 at 20:18:03

> *"Reality is negotiable." — Tim Ferriss*

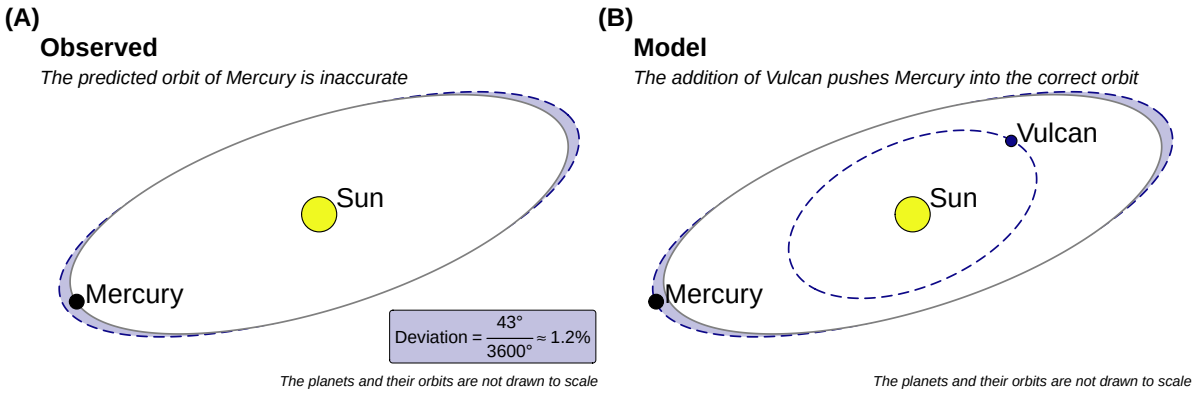

Figure 1.1: foo **(A)** foo **(B)** foo

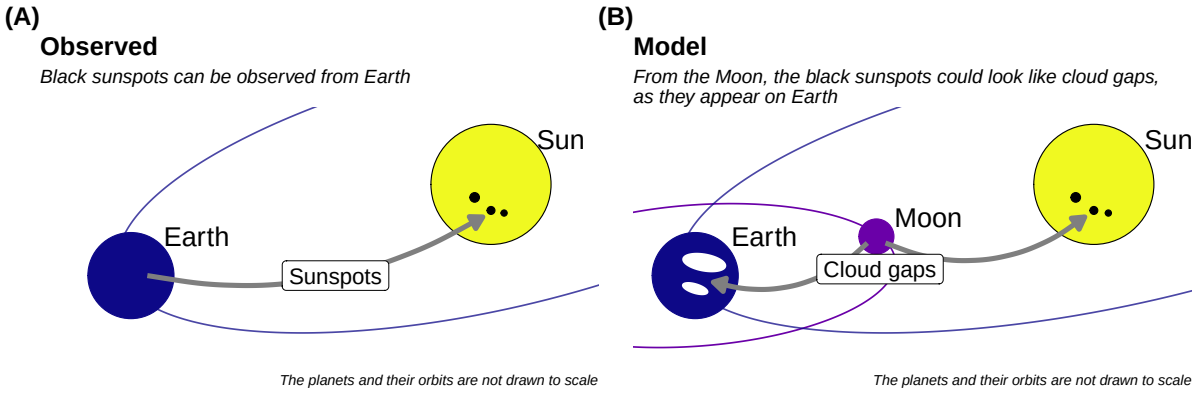

Figure 1.2: foo **(A)** foo **(B)** foo

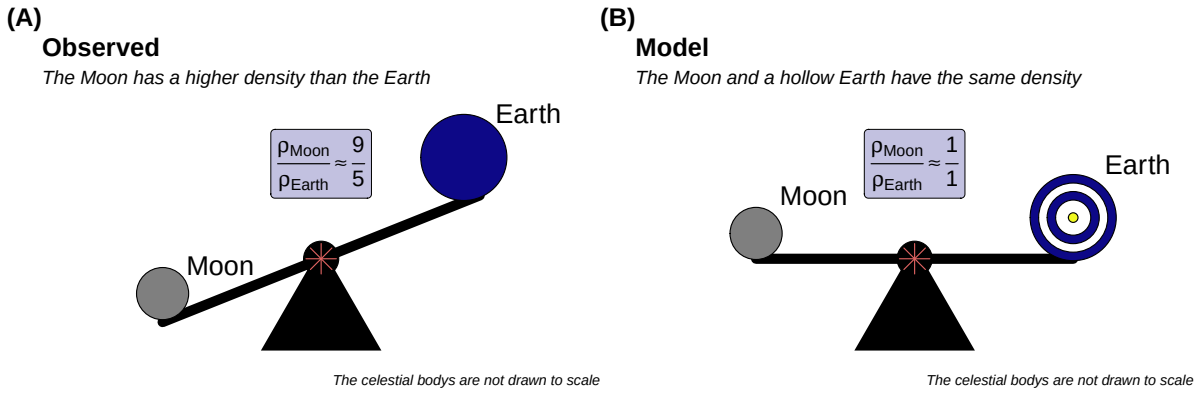

Figure 1.3: foo **(A)** foo **(B)** foo

1.1 General background

> *"Any sufficiently advanced technology is indistinguishable from magic." — Clarke (1968), Clarke's third law*

Full quote

> *"Reality is negotiable. Outside of science and law, all rules can be bent or broken, and it doesn't require being unethical." — Tim Ferriss*

Idea of hypotheses

Science is guessing and falsification

Chalmers (2013) What is this thing called Science?

R. Feynman (1966) What is science

Campbell (1952) What is science?

Clark and Berry (2025) Models Demystified: A Practical Guide from Linear Regression to Deep Learning

Cox and Efron (2017) Statistical Thinking for the 21st Century

Deutsch (2011) The beginning of infinity: Explanations that transform the world

David Deutsch > Quotes

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 "Models of Reality" in this chapter.

- itemize with max. 5-6 words

1.8 Summary

References

2 What is data?

Last modified on 18. November 2025 at 07:21:12

> *"The limits of my language mean the limits of my world." — Ludwig Wittgenstein*

2.1 General background

2.2 Theoretical background

2.3 R packages used

2.4 Data

```
jump_weight_tbl <- tibble(x = c(0.6, 1, 2.3, 3.5, 5.2, 7.1, 8.4, 9.2, 10),
                          y = 0.15*x^3 - 2.2*x^2 + 8.8*x + 3.2 + rnorm(9, 0, 0.5)) |>
  mutate_all(round, 1) |>
  rename(weight_mg = x, jump_length_cm = y)
```

Table 2.1: foo.

weight_mg	jump_length_cm
0.6	7.3
1.0	9.7
2.3	14.0
3.5	13.7
5.2	11.1
7.1	8.3
8.4	10.5
9.2	14.3
10.0	20.8

Equation 2.1

$$y = 0.15 \cdot x^3 - 2.2 \cdot x^2 + 8.8 \cdot x + 3.2 \tag{2.1}$$

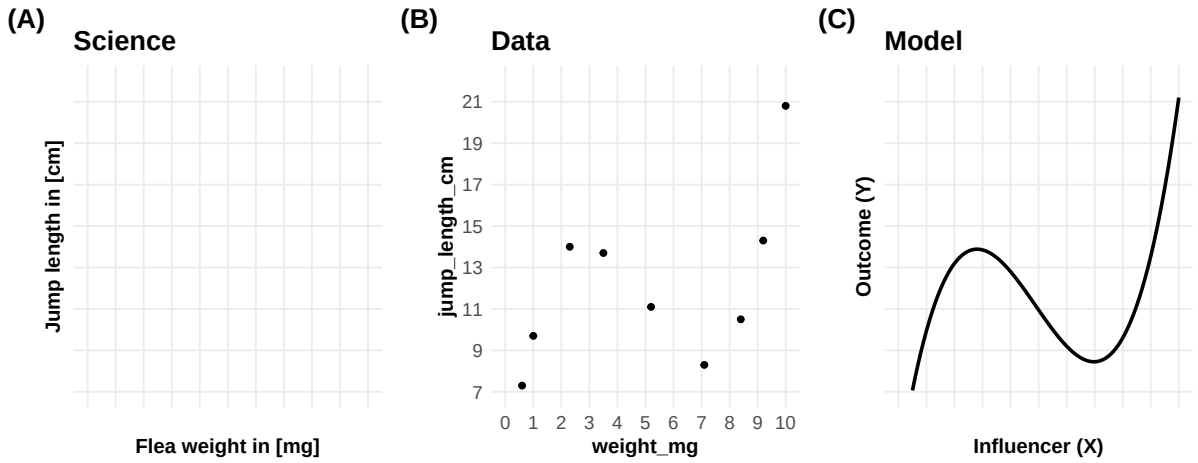

Figure 2.1: foo.

2.5 Alternatives

Further tutorials and R packages on XXX

2.6 Glossary

term what does it mean.

2.7 The meaning of "Models of Reality" in this chapter.

- itemize with max. 5-6 words

2.8 Summary

References

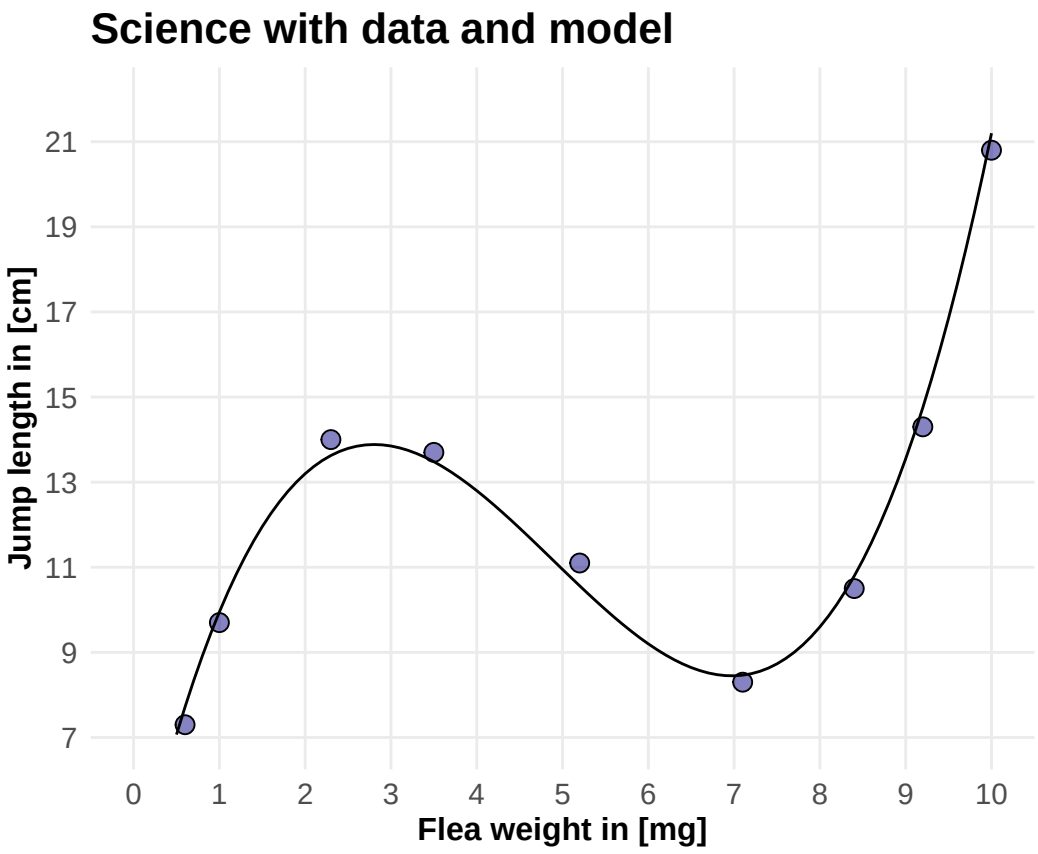

Figure 2.2: foo.

3 What is a model?

Last modified on 17. November 2025 at 20:02:42

"A quote." — Dan Meyer

3.1 General background

Table 3.1: Table of terms used in the statistical modelling. The terms in bold are used here. Depending on the scientific background, the usage of these terms can vary widely.

Application	Symbol	Name	Description
	y	**outcome**, response, endpoint, dependent variable	*The right-hand side (abbr. RHS) of the model. Describing the values measured in an experiment or study.*
	x	**influencer**, influential variable, risk factor, fixed effect, independent variable	*The left-hand side (abbr. LHS) of the model. Describing the influential variables in an experiment or study.*
	z	**random effect**	*A factor that provides a description of an grouping variable, which is not part of the controlled experimental setting.*
Explanation	x	**explanator**, explanatory variable	*The influencer is used to describe or explain the outcome.*
Prediction	x	**predictor**, predictive variable	*The influencer is used to predict the outcome.*
Main effect	x	**focal explanator**, **focal predictor**, focal variable	*In a model with multiple influencers, the focal variable is the variable of primary interest.*
Continuous x	c	**covariate**, covariable	*The influencer is a numeric variable with continuous values.*
Categorical x	f_A	**factor A**, factorial variable, categorical variable	*The influencer is discrete, functioning as a grouping variable, such as an experimental group or a treatment.*

Application	Sym-bol	Name	Description
Factor f_A	$A.1$ to $A.j$	**levels**, groups, treatment groups	*The discrete groups included in one factor A.*

A sentence why we use y and not x for mean and other stuff.

McCullagh (2002) What is a statistical model?

Appleton (1995) What do we mean by a statistical model?

Hand (2019) What Is the Purpose of Statistical Modeling?

Spanos (2006) Where do statistical models come from? Revisiting the problem of specification

Gilchrist (1984) Statistical modelling

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 "Models of Reality" in this chapter.

- itemize with max. 5-6 words

3.8 Summary

References

Part II

Tales of data

Last modified on 18. November 2025 at 19:44:46

> *"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 Factorial data

Last modified on 18. November 2025 at 19:55:01

> *"A quote." — Dan Meyer*

4.1 General background

4.2 Theoretical background

4.3 R packages used

4.4 Data

Based on standard methods in flea research and experimental entomology, as well as a similar published experiment, the following three types of food for adult cat fleas are possible, representing different nutritional conditions:

Blood: This is the natural and optimal food source for adult fleas.Often, defibrinated or anticoagulated animal blood (e.g. bovine blood, rabbit blood) is used for this purpose, which is offered in special in vitro feeding systems (e.g. through a membrane).Expectation: Fleas that are optimally nourished should show the greatest jumping distance.

Sugar water: Often serves as a 'control feed' or as a feed that provides energy (sugar) but lacks essential nutrients (such as proteins from blood).Expectation: Jumping distance could be reduced due to the lack of blood and thus the proteins important for reproduction, which could impair physiological fitness.

Ketchup - a nutrient-poor or unsuitable food: This option is used to simulate poor, incomplete or stressful nutritional conditions.In a similar documented experiment, ketchup (a combination of sugar, vinegar and minimal other substances, but no blood) was used as the third feed.Expectation: It can be assumed that fleas will show the shortest jumping distance under these conditions, as they lack both essential nutrients and the necessary energy.

4.5 Alternatives

Further tutorials and R packages on XXX

4.6 Glossary

term what does it mean.

4.7 The meaning of "Models of Reality" in this chapter.

- itemize with max. 5-6 words

4.8 Summary

References

5 Covariate data

Last modified on 18. November 2025 at 19:54:48

> *"A quote." — Dan Meyer*

5.1 General background

5.2 Theoretical background

5.3 R packages used

5.4 Data

5.5 Alternatives

Further tutorials and R packages on XXX

5.6 Glossary

term what does it mean.

5.7 The meaning of "Models of Reality" in this chapter.

- itemize with max. 5-6 words

5.8 Summary

References

6 Combinded data

Last modified on 18. November 2025 at 19:54:35

> *"A quote." — Dan Meyer*

6.1 General background

6.2 Theoretical background

6.3 R packages used

6.4 Data

6.5 Alternatives

Further tutorials and R packages on XXX

6.6 Glossary

term what does it mean.

6.7 The meaning of "Models of Reality" in this chapter.

- itemize with max. 5-6 words

6.8 Summary

References

Part III

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

7 Template chapter

Last modified on 17. November 2025 at 20:02:35

> *"A quote." — Dan Meyer*

7.1 General background

7.2 Theoretical background

7.3 R packages used

7.4 Data

7.5 Alternatives

Further tutorials and R packages on XXX

7.6 Glossary

term what does it mean.

7.7 The meaning of "Models of Reality" in this chapter.

- itemize with max. 5-6 words

7.8 Summary

References

Appleton, David R. 1995. "What Do We Mean by a Statistical Model?" *Statistics in Medicine* 14 (2): 185–97.

Campbell, Norman Robert. 1952. "What Is Science?"

Chalmers, Alan F. 2013. *What Is This Thing Called Science?* Hackett Publishing.

Clark, Michael, and Seth Berry. 2025. *Models Demystified: A Practical Guide from Linear Regression to Deep Learning*. CRC Press.

Clarke, Arthur C. 1968. "Clarke's Third Law on UFO's." *Science* 159 (3812): 255–55.

Cox, DR, and Bradley Efron. 2017. "Statistical Thinking for 21st Century Scientists." *Science Advances* 3 (6): e1700768.

Deutsch, David. 2011. *The Beginning of Infinity: Explanations That Transform the World*. penguin uK.

Feynman, Richard. 1966. "What Is Science."

Feynman, Richard P. 1998. "Cargo Cult Science." In *The Art and Science of Analog Circuit Design*, 55–61. Elsevier.

Feynman, Richard Phillips, and Ralph Leighton. 1992. *" Surely You're Joking, Mr. Feynman!": Adventures of a Curious Character*. Random House.

Gilchrist, Warren. 1984. *Statistical Modelling*. Wiley Chichester.

Hand, David. 2019. "What Is the Purpose of Statistical Modeling?" *Harvard Data Science Review* 1 (1).

McCullagh, Peter. 2002. "What Is a Statistical Model?" *The Annals of Statistics* 30 (5): 1225–1310.

Smolin, Lee. 2019. *Einstein's Unfinished Revolution: The Search for What Lies Beyond the Quantum*. Penguin.

Spanos, Aris. 2006. "Where Do Statistical Models Come from? Revisiting the Problem of Specification." *Lecture Notes-Monograph Series*, 98–119.