Applied Error Modeling Using R

Stefanie Muff and Lukas F. Keller

2018-12-20

Contents

Preface

This is a *first draft* of a book that deals with effects and cures of measurement error in variables of regression models.

The presence and effects of measurement error and misclassification in covariates and the response of regression models have been recognized already more than a century ago. Thanks to huge efforts of plenty of researchers, in many settings, the consequences of ignoring measurement error or misclassification are known, at least in theory.

When we say "error", we do not only mean actual mistakes in the data that are used to fit regression models.

kind of uncertainty, noise or that are present in the data that we use to fit our models.

- A ton of methods exist
- They go largely unused
- Possible reasons:
  - Problems and methods unknown

  - Often assumed that effect estimates are conservative
  - Solutions often complicated
  - Methods not accessible

N. Breslow, *Lessons in Biostatistics* (2014) (Breslow, 2014) wrote

> Obviously, [. . .] the *best* method of dealing with measurement error was to avoid it.

I say:

> The *second best* method of dealing with measurement error is to properly account for it.

We might develop a package. In this case, the **package-to-be-developed** package can be installed from CRAN or Github:

```
install.packages("package-to-be-developed")
# or the development version
# devtools::install_github("stefaniemuff/package-to-be-developed")
```

Follow me on Twitter! @StefanieMuff

Chapter 1

Introduction

1.1 What is Error?

Explain that measurement error often comes in the shape of uncertainty, which is present in almost all data.

1.2 Why and When do I Have to Worry?

- Triple whammy of ME
- When is error a problem?
- Bias versus variance
- Is it sometimes better not to model the error?

It is surprising how many phenomena in statistics and its applications can be viewed through the measurement error lens. A prominant example is the concept of heritability in genetics and evolutionary biology, as we will explain in Section 3.2.

1.3 Take-Home Messages of This Book

1.4 Outlook

What we are going to do.

Chapter 2

Types of Errors

2.1 Continuous Variables

Two fundamentally different error types

2.1.1 Classical Measurement Error

2.1.2 Berkson Measurement Error

2.2 Categorical and Count Variables

Chapter 3

The Effects of Measurement Error

We will look into effects of ME in the linear regression case.

3.1 Classical Measurement Error

3.2 The Concept of Heritability, Regression to the Mean and Measurement Error

Geneticists, evolutionary biologists and animal breeders will be familiar with the concept of *heritability* (3.1).

$$h^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_E^2} \tag{3.1}$$

- Will use data in Figures 3.1 and 3.2 to explain regression to the mean

(Fuller, 1987; Galton, 1886)

3.3 Berkson Measurement Error

3.4 Error in Categorical and Count Variables

3.5 Error in the response

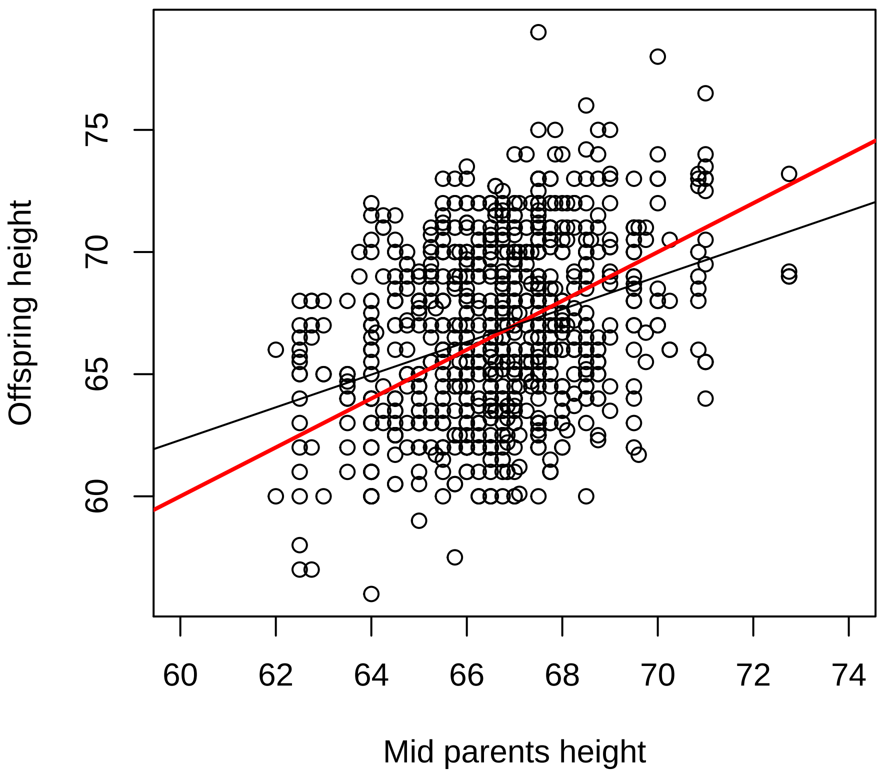

Figure 3.1: Data drawn from ‘http://www.math.uah.edu/stat/data/Galton.txt‘

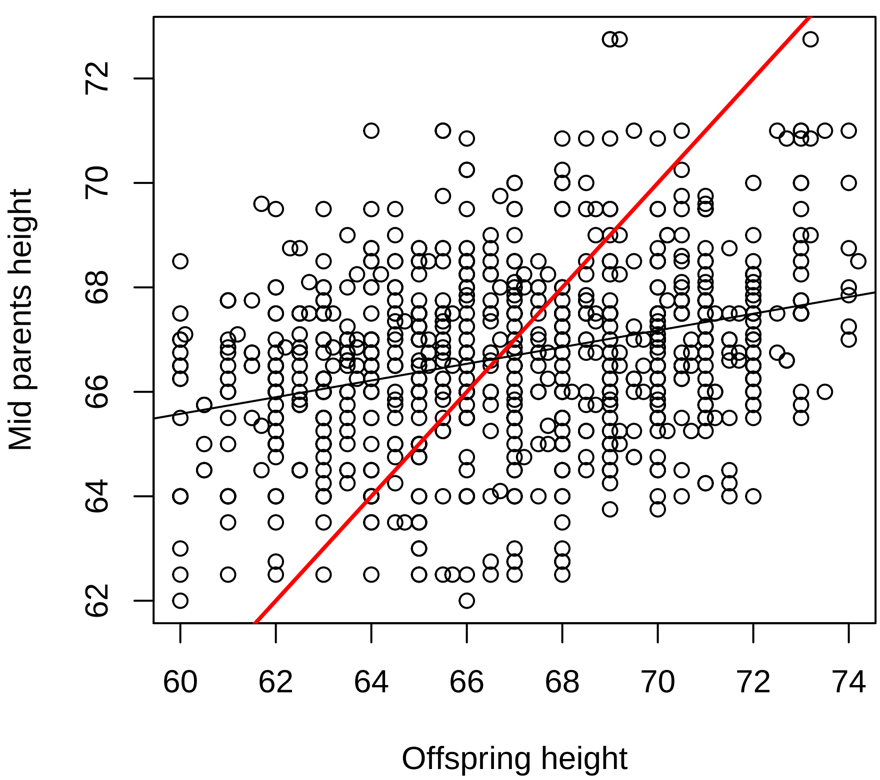

Figure 3.2: Data drawn from ‘http://www.math.uah.edu/stat/data/Galton.txt‘

Chapter 4

Approaches to Account for Measurement Error

4.1 Bayesian Methods

4.2 Simulation Extrapolation (SIMEX)

Chapter 5

Linear Regression Models

Chapter 6

Generalized Linear (Mixed) Models

6.1 Classical error

6.1.1 Error in a covariate

- Correlated covariates

6.1.2 Error in the response

6.2 Berkson error

6.2.1 Error in a covariate

6.2.2 Error in the response

Bibliography

Breslow, N. E. (2014). Lessons in biostatistics. In Lin, X., Genest, C., Banks, D. L., Molenberghs, G., Scott, D. W., and Wang, J., editors, *Past, Present and Future of Statistical Science*, pages 335–347. CRC Press.

Fuller, W. A. (1987). *Measurement Error Models.* John Wiley & Sons, New York.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland*, 15:246–263.