Kurs Bio144:
Datenanalyse in der Biologie

Stefanie Muff & Owen L. Petchey

Lecture 11: Measurement error in regression models

18./19. May 2017

Overview (todo: check)

- ME in the response (y) and in covariates (x) of regression models.
- Effects of ME on regression parameters.
- When do I have to start to worry?
- Simple methods to correct for ME.

Course material covered today

The lecture material of today is partially based on the following literature:

- Chapter 6.1 in "Lineare regression"

Sources of measurement uncertainty / measurement error (ME)

- **Measurement imprecision** in the field or in the lab (length, weight, blood pressure, etc.).
- Errors due to **incomplete** or **biased observations** (e.g., self-reported dietary aspects, health history).
- Biased observations due to **preferential sampling or repeated observations**.
- Rounding error, digit preference.
- **Misclassification error** (e.g., exposure or disease classification).
- ...

"Error" or "uncertainty"?

Why should ME not be ignored?

- It is a fundamental assumption that explanatory variables are measured or estimated without error, for instance for
  - the calculation of correlations.
  - linear regression and ANOVA.
  - Generalized linear and non-linear regressions (e.g. logistic an Poisson).
- Most other modelling assumptions are routinely checked!
- Violation of this assumption may lead to biased parameter estimates, altered standard errors and p-values, incorrect covariate importances, and to misleading conclusions.
- Even standard statistics textbooks do often not mention these problems.

The effects of measurement error (ME)

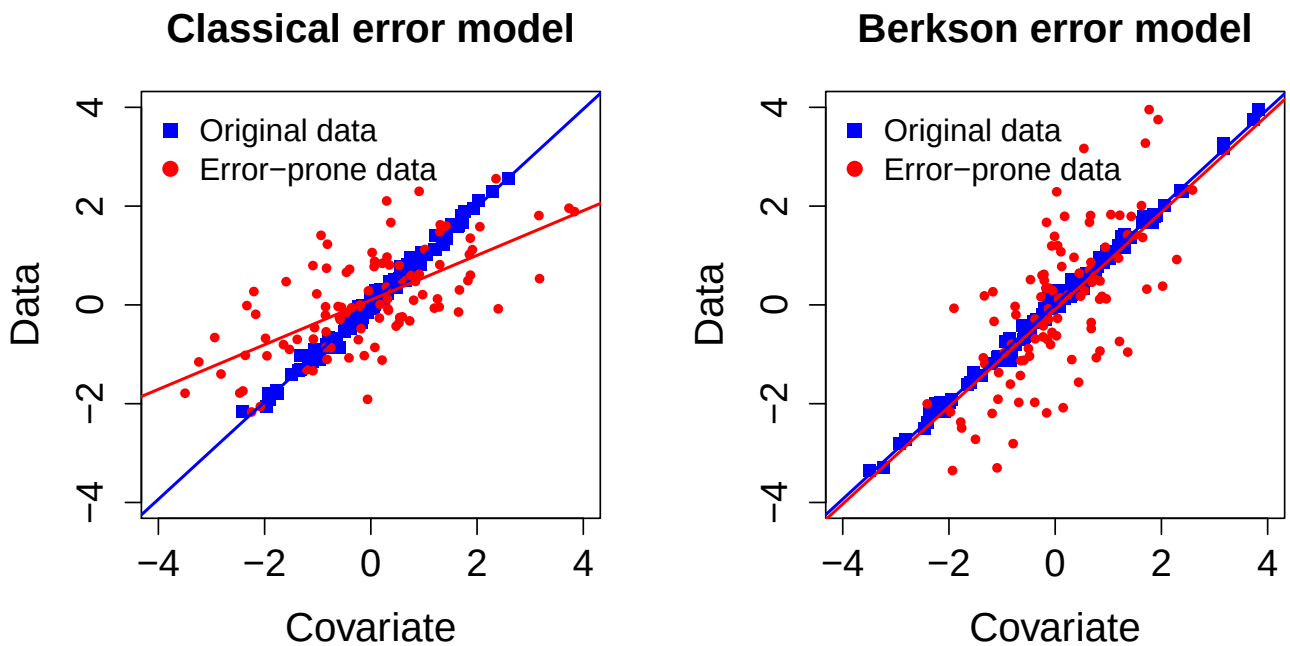
Bias the regression parameters, mainly attenuation (underestimation) of the true effect.

Effect of ME in linear regression

Find regression parameters β_0 and β_x for unobserved **x**:

$$y_i = 1 \cdot x_i + \epsilon_i, \quad \epsilon_i \sim \mathrm{N}(0, \sigma_\epsilon^2) \ .$$

Simulation: $n = 100$, $\sigma_\epsilon^2 = 1/100$, $\sigma_x^2 = \sigma_u^2 = 1$.



Simulations or apps

Shiny apps for classical error in linear, logistic and Poisson regression:

▸ Classical error

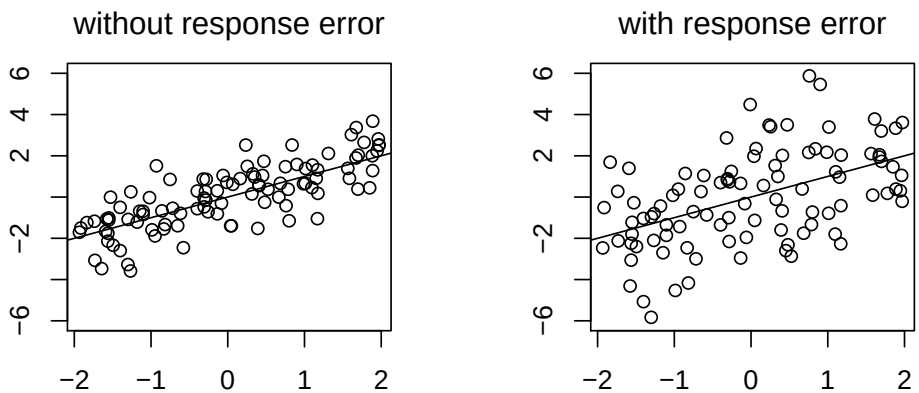
Error in the outcome of regression models

Example: Continuous error in a linear regression outcome.

Note: In the case when the observed response

$$s_i = y_i + v_i \quad v_i \sim \mathrm{N}(0, \sigma_v^2) \ ,$$

the error variance is simply absorbed in the residual variance σ_ϵ^2.



How to correct for error?

(attenuation factor in lin. Reg, SIMEX in some more general cases, Bayesian approach (only mention it))

Summary