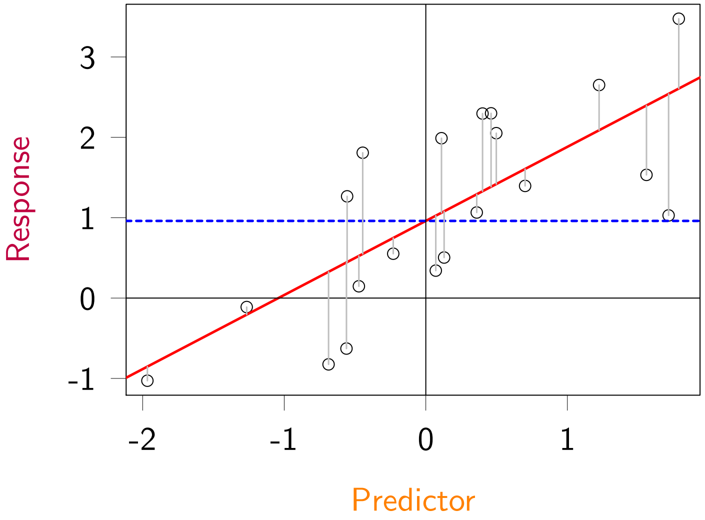
Generalized Linear Models (GLMs)

May 17, 2018

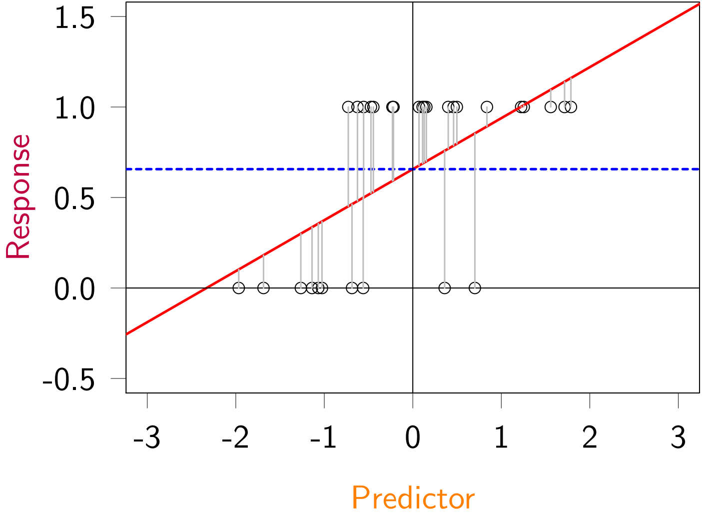
1. Linear model, reminder

A simple linear model

Response = **Intercept** + **Slope** × **Predictor** + **Error**



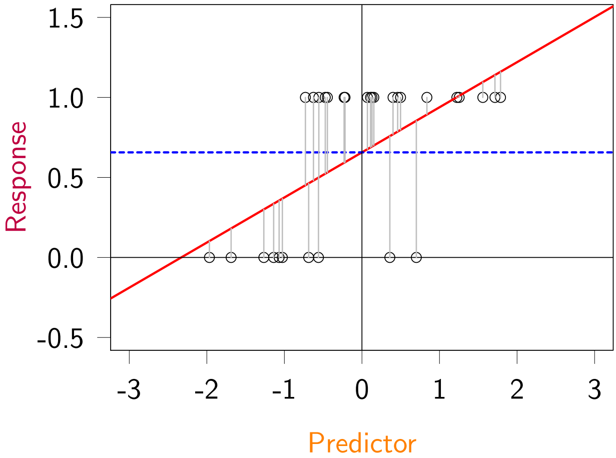
A simple linear model failure: binary data



Linear model basic assumptions

- Linear combination of parameters (including transformation, polynoms, interactions. . .)
  Risk: biologically meaningless
- Predictor not perfectly correlated
  Risk: Model won't run, unstable convergence, or huge SE
- Little error in predictors
  Risk: bias estimates (underestimate with Gaussian error)
- Gaussian error distribution
  Risk: Poor predictions
- Homoscedasticity (constant error variance)
  Risk: Over-optimistic uncertainty, unreliable predictions
- Independence of error
  Risk: Bias and over-optimistic uncertainty

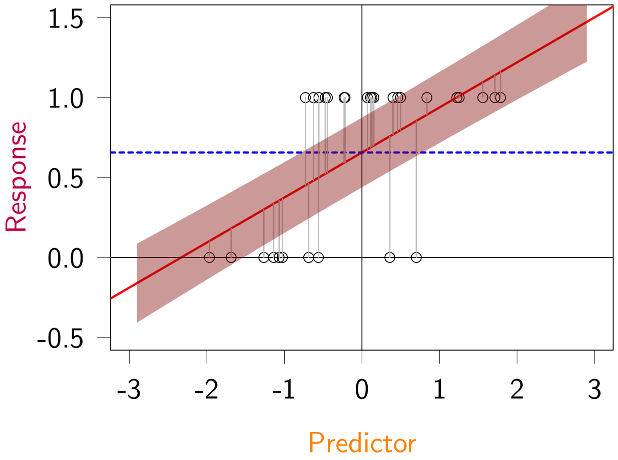
A simple linear model failure: binary data



Assumptions violated:

Non-Gaussian errors, non-constant error variance, correlated errors

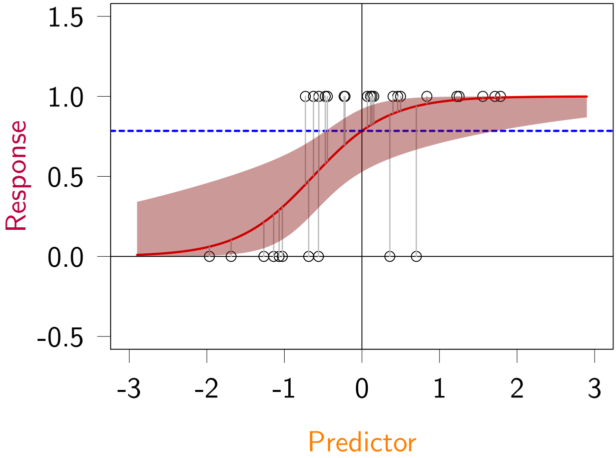
A simple linear model failure: binary data



Practical consequences:

Non-sensical predictions, wrong confidence-interval and p-value, extrapolation ALWAYS fails

What we want our model to do



Good features:

Never out of [0,1], variable uncertainty, non-linear trend, close fit

That is what a Generalized Linear Model does

Vocabulary warning

- General Linear Model (=linear model with several responses, multivariate)
- **Generalized Linear Model (=non-normal errors, and uncertainty dependent on the mean)**

That is what a Generalized Linear Model does

Vocabulary warning

- General Linear Model (=linear model with several responses, multivariate)
- **Generalized Linear Model (=non-normal errors, and uncertainty dependent on the mean)**

What a GLM is:

1. A linear function ($y = \mu + \beta x \ldots$)
2. A probability distribution (Bernouilli, Binomial, Poisson...)
3. A "link function" to convert between the scale of the linear function ($-\infty$ to $+\infty$) and the scale of the data and the probability distribution (often positive integer: 0, 1, 2, 3...)

Logistic regression

- Binary or proportion data

Binomial (and Bernouilli distribution in R):

```
bernouilli_random_sample <- rbinom(n = 10000, size = 1, prob = 0.3)
hist(bernouilli_random_sample)
mean(bernouilli_random_sample); 0.3
var(bernouilli_random_sample); 0.3*(1-0.3)
```

Logistic regression in R:

```
glm(formula = obs ~ 1 + x, family = "binomial", data=data)
```

Logistic regression

- Binary or proportion data
- Binomial probability distribution (= Bernouilly if binary data)

Binomial (and Bernouilli distribution in R):

```
bernouilli_random_sample <- rbinom(n = 10000, size = 1, prob = 0.3)
hist(bernouilli_random_sample)
mean(bernouilli_random_sample); 0.3
var(bernouilli_random_sample); 0.3*(1-0.3)
```

Logistic regression in R:

```
glm(formula = obs ~ 1 + x, family = "binomial", data=data)
```

Logistic regression

- Binary or proportion data
- Binomial probability distribution (= Bernouilly if binary data)
- Link function often logit: $y = \log(\frac{probability}{1-probability})$

Binomial (and Bernouilli distribution in R):

```
bernouilli_random_sample <- rbinom(n = 10000, size = 1, prob = 0.3)
hist(bernouilli_random_sample)
mean(bernouilli_random_sample); 0.3
var(bernouilli_random_sample); 0.3*(1-0.3)
```

Logistic regression in R:

```
glm(formula = obs ~ 1 + x, family = "binomial", data=data)
```

Logistic regression

- Binary or proportion data
- Binomial probability distribution (= Bernouilly if binary data)
- Link function often logit: $y = \log(\frac{probability}{1-probability})$
- Back-transformation inverse-logit: $probability = \frac{1}{1+exp(-y)}$

Binomial (and Bernouilli distribution in R):

```
bernouilli_random_sample <- rbinom(n = 10000, size = 1, prob = 0.3)
hist(bernouilli_random_sample)
mean(bernouilli_random_sample); 0.3
var(bernouilli_random_sample); 0.3*(1-0.3)
```

Logistic regression in R:

```
glm(formula = obs ~ 1 + x, family = "binomial", data=data)
```

Logistic regression

- Binary or proportion data
- Binomial probability distribution (= Bernouilly if binary data)
- Link function often logit: $y = \log(\frac{probability}{1-probability})$
- Back-transformation inverse-logit: $probability = \frac{1}{1+exp(-y)}$
- Linear function $y = intercept + slope_1 predictor_1 + slope_2 predictor_2 + \ldots$

Binomial (and Bernouilli distribution in R):

```
bernouilli_random_sample <- rbinom(n = 10000, size = 1, prob = 0.3)
hist(bernouilli_random_sample)
mean(bernouilli_random_sample); 0.3
var(bernouilli_random_sample); 0.3*(1-0.3)
```

Logistic regression in R:

```
glm(formula = obs ~ 1 + x, family = "binomial", data=data)
```