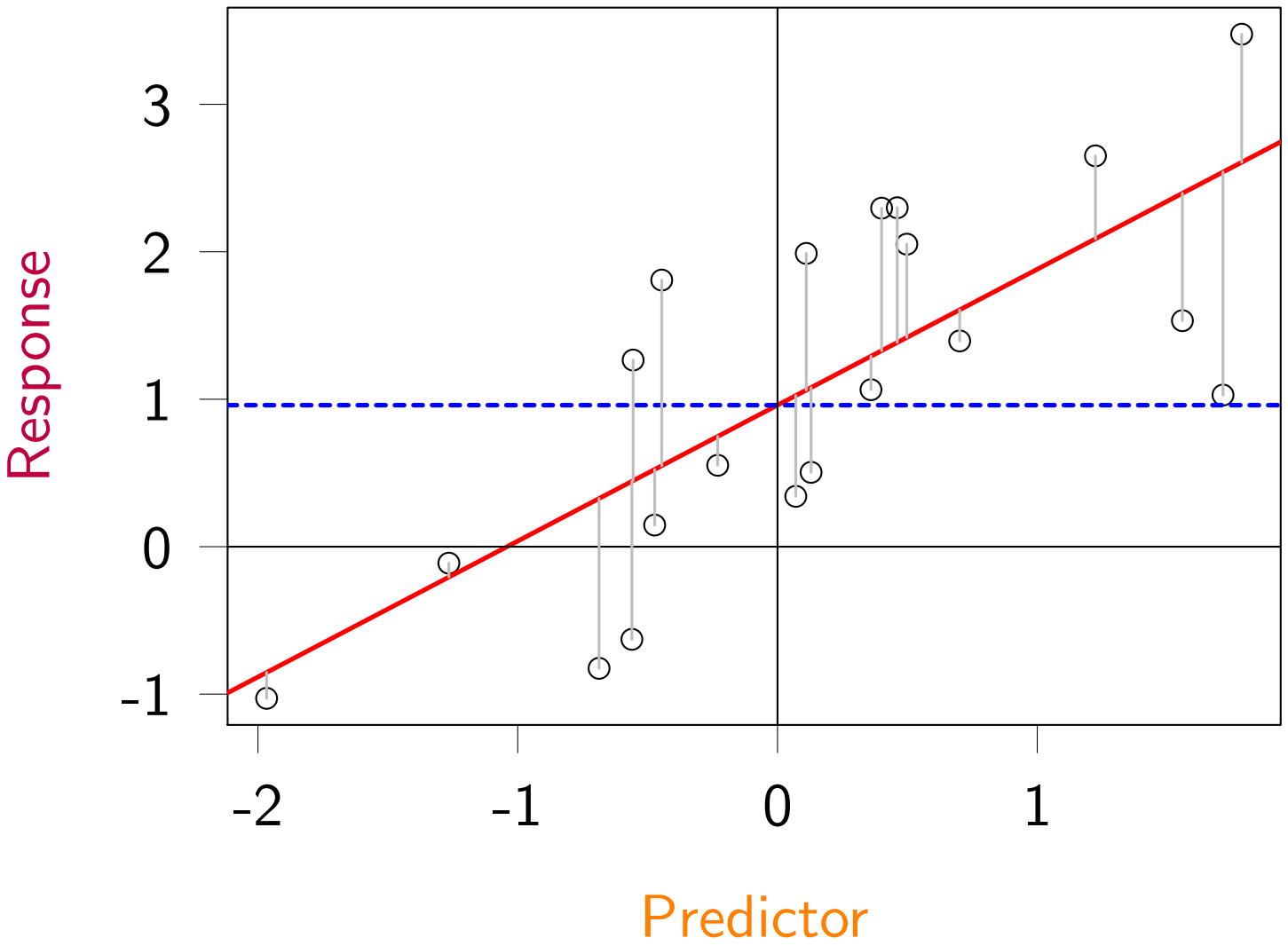
Multiple regressions and interactions

April 4, 2018

A simple linear model

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



A simple linear model

$$\textbf{Response} = \textbf{Intercept} + \textbf{Slope1} \times \textbf{Predictor1} + \textbf{Slope2} \times \textbf{Predictor2} + \textbf{Error}$$

In R:

```
lm(response ~ 1 + predictor1 + predictor2, data=data)
```

Warnings

Vocabulary warning!

- *correlation*: linear association between two variables "how well does x explain y ?"

Warnings

Vocabulary warning!

- *correlation*: linear association between two variables "how well does x explain y ?"
- *interaction*: non-additive effect of two or more variables "does the effect of x_1 on y change as a function of x_2?"

Warnings

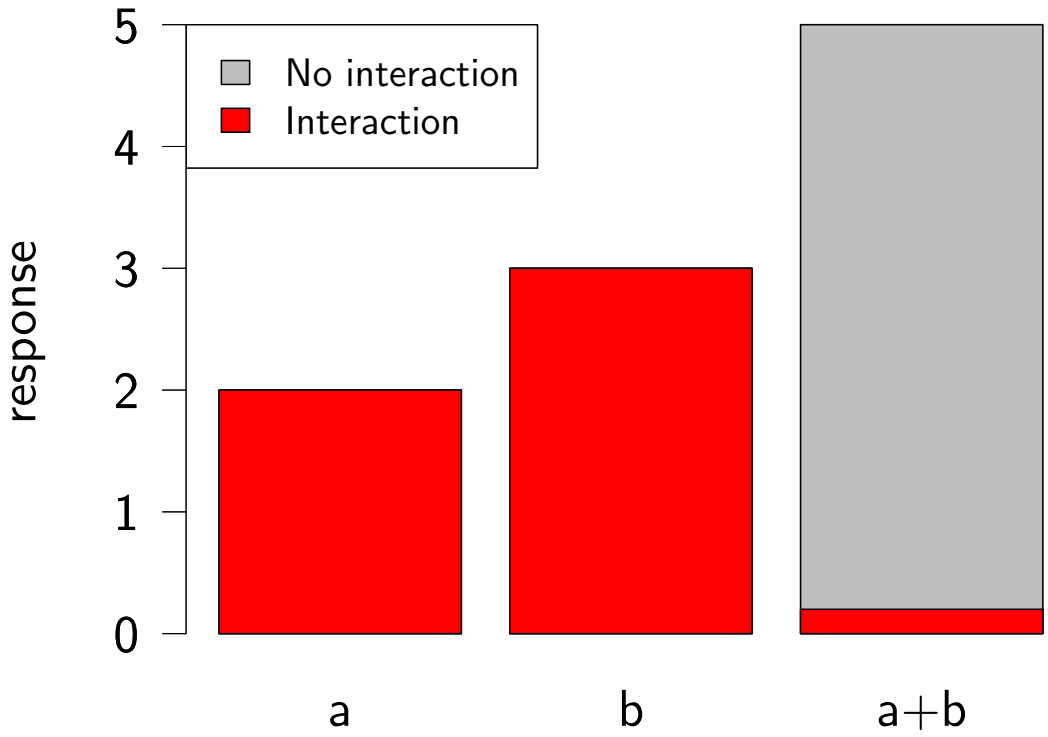
Vocabulary warning!

- *correlation*: linear association between two variables "how well does x explain y ?"
- *interaction*: non-additive effect of two or more variables "does the effect of x_1 on y change as a function of x_2?"

Warnings

Vocabulary warning!

- *correlation*: linear association between two variables "how well does x explain y ?"
- *interaction*: non-additive effect of two or more variables "does the effect of x_1 on y change as a function of x_2?"



Fitting an interaction

```
lm(y ~ 1 + x1 * x2)
lm(y ~ 1 + x1 + x2 + x1:x2)
```

Fitting an interaction

```
lm(y ~ 1 + x1 * x2)
lm(y ~ 1 + x1 + x2 + x1:x2)
```

```
summary(lm(y~ 1 + x1*x2))
```

```
Call:
lm(formula = y ~ 1 + x1 * x2)

Residuals:
    Min      1Q  Median      3Q     Max
-1.8719 -0.6777 -0.1086  0.5897  2.3166

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.14098    0.09578  11.913  < 2e-16 ***
x1          -0.49281    0.10834  -4.549 1.58e-05 ***
x2           0.53434    0.09881   5.408 4.67e-07 ***
x1:x2        0.35911    0.11449   3.137  0.00227 **
---
```

Fitting an interaction

Why the multiplication sign?

Fitting an interaction

Why the multiplication sign?

```
x1Xx2 <- x1*x2
```

Fitting an interaction

Why the multiplication sign?

```
x1Xx2 <- x1*x2
```

```
summary(lm(y  1 + x1 + x2 + x1Xx2))

Call: lm(formula = y  1 + x1 + x2 + x1Xx2)
Residuals: Min 1Q Median 3Q Max -1.8719 -0.6777 -0.1086 0.5897 2.3166
Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 1.14098 0.09
11.913 < 2e-16 *** x1 -0.49281 0.10834 -4.549 1.58e-05 *** x2 0.53434 0.098
5.408 4.67e-07 *** x1Xx2 0.35911 0.11449 3.137 0.00227 ** --- Signif. codes
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.9468 on 96 degrees of freedom Multiple R-squared
0.4252, Adjusted R-squared: 0.4072 F-statistic: 23.67 on 3 and 96 DF, p-val
1.49e-11
```

Fitting an interaction

Why the multiplication sign?
Really a multiplication, but can be complicated for factors

Warnings

Modeling warning!

- ~~DO NOT COMPARE P-VALUES OF TWO MODELS TO TEST FOR AN INTERACTION~~

Exercise

1. Load the data masssex.csv
2. Fit a simple regression explaining movement by mass for each sex separately. Is the relationship different between sexes?
3. Fit the multiple regression explaining movement by mass, sex, and mass:sex, using the full dataset. Is the relationship different between sexes?
4. Try to understand the discreapancy by plotting the data

Warnings

1.

```
masssex <- read.csv(file="masssex.csv")
```

Warnings

1.

```
masssex <- read.csv(file="masssex.csv")
```

2.

```
summary(lm(movement ~ mass, data=masssex[masssex$sex==0,]))
summary(lm(movement ~ mass, data=masssex[masssex$sex==1,]))
```

Warnings

1.

```
masssex <- read.csv(file="masssex.csv")
```

2.

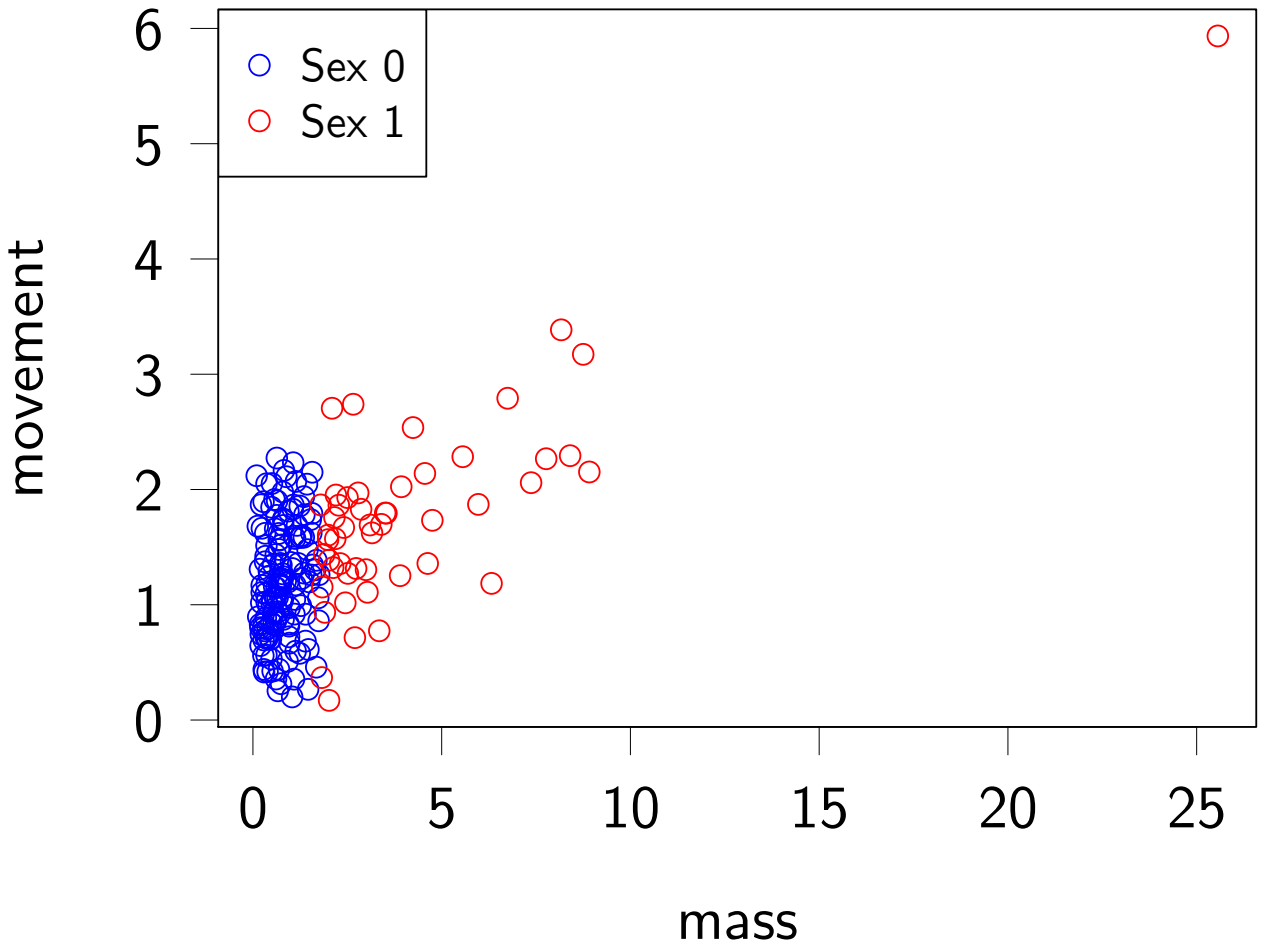
```
summary(lm(movement ~ mass, data=masssex[masssex$sex==0,]))
summary(lm(movement ~ mass, data=masssex[masssex$sex==1,]))
```

3.

```
summary(lm(movement ~ mass*sex, data=masssex))
```

Warnings

4.



Next times

- April 20th Kevin on ggplot
- May 4th Nina on Structural Equation Modeling
- then, mixed models and GLM
- **Other requests?**