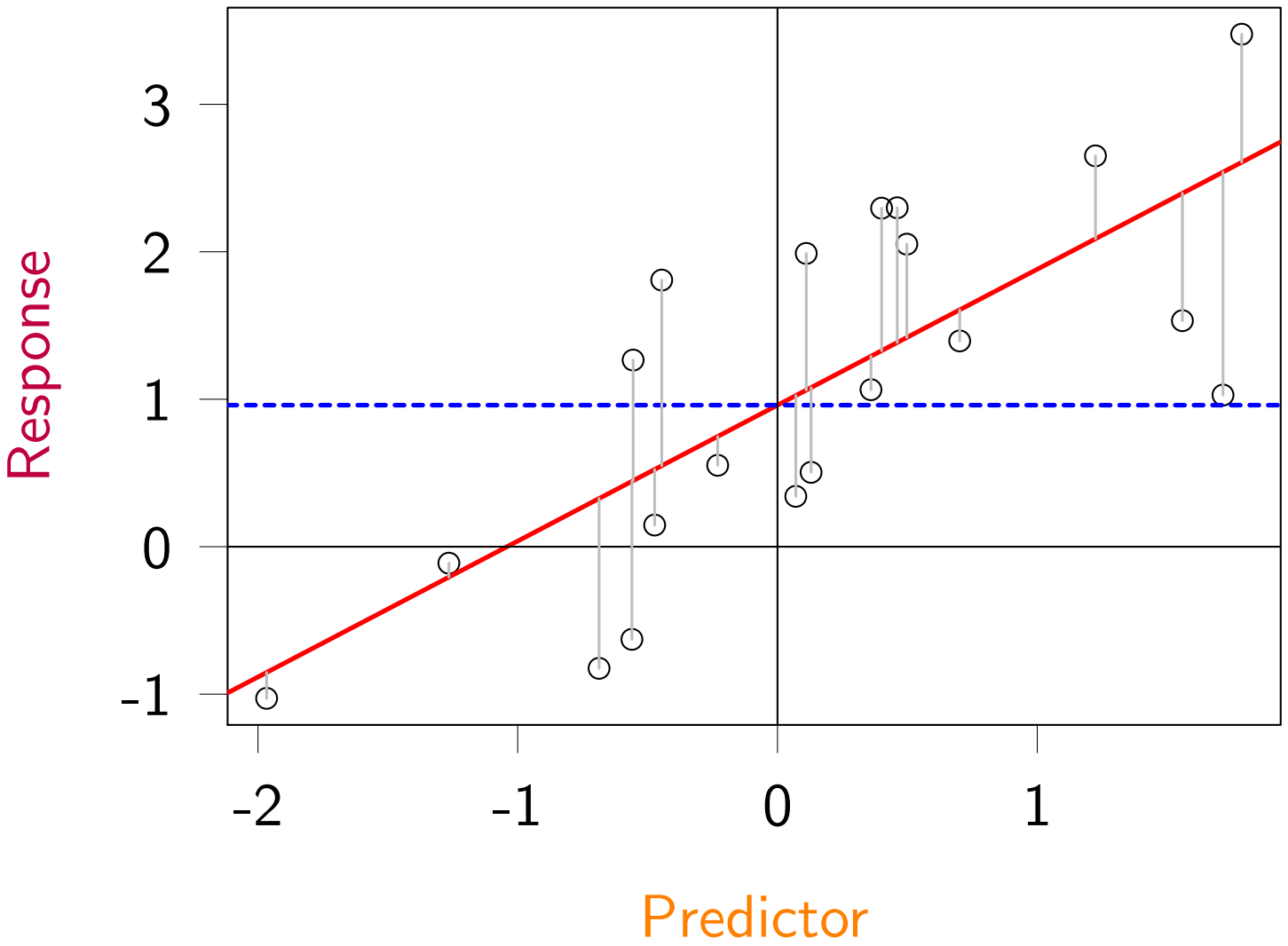
Multiple regressions and interactions

April 5, 2018

A simple linear model

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



A multiple linear model

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

In R:

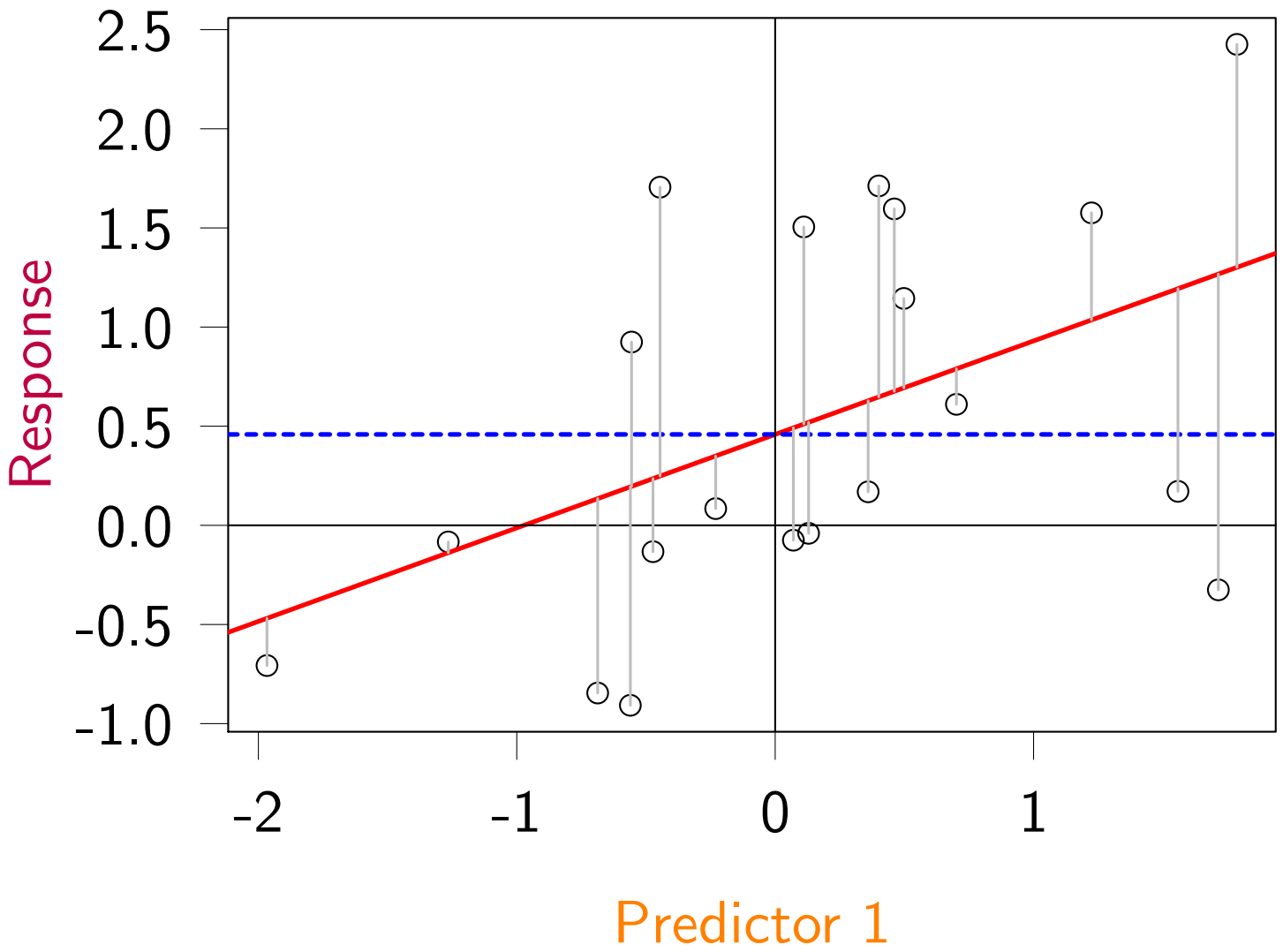
```
lm(response ~ 1 + predictor1 + predictor2, data=data)
```

Sequential regression

We want to explain a response by three predictors

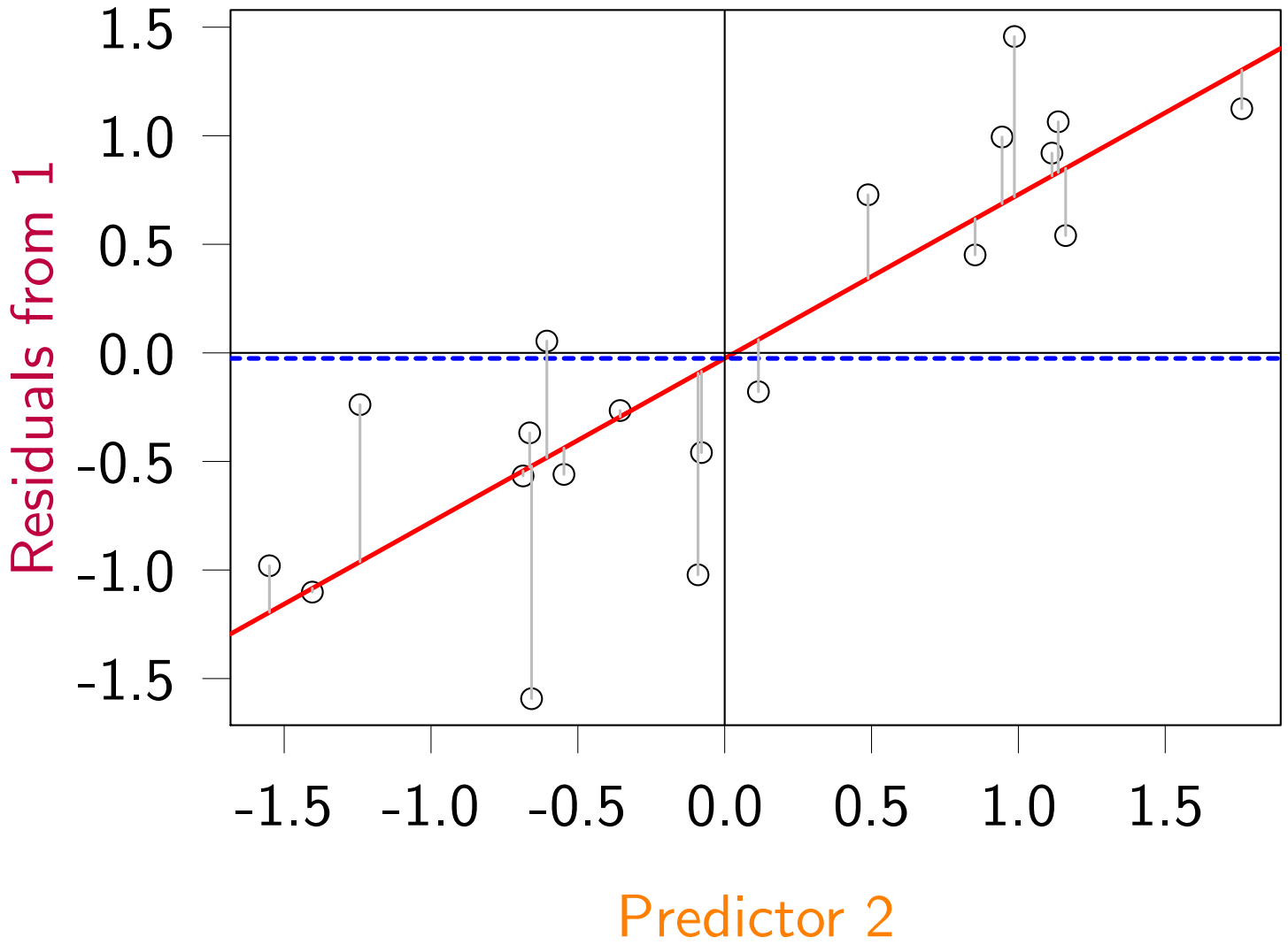
Sequential regression

We want to explain a response by three predictors



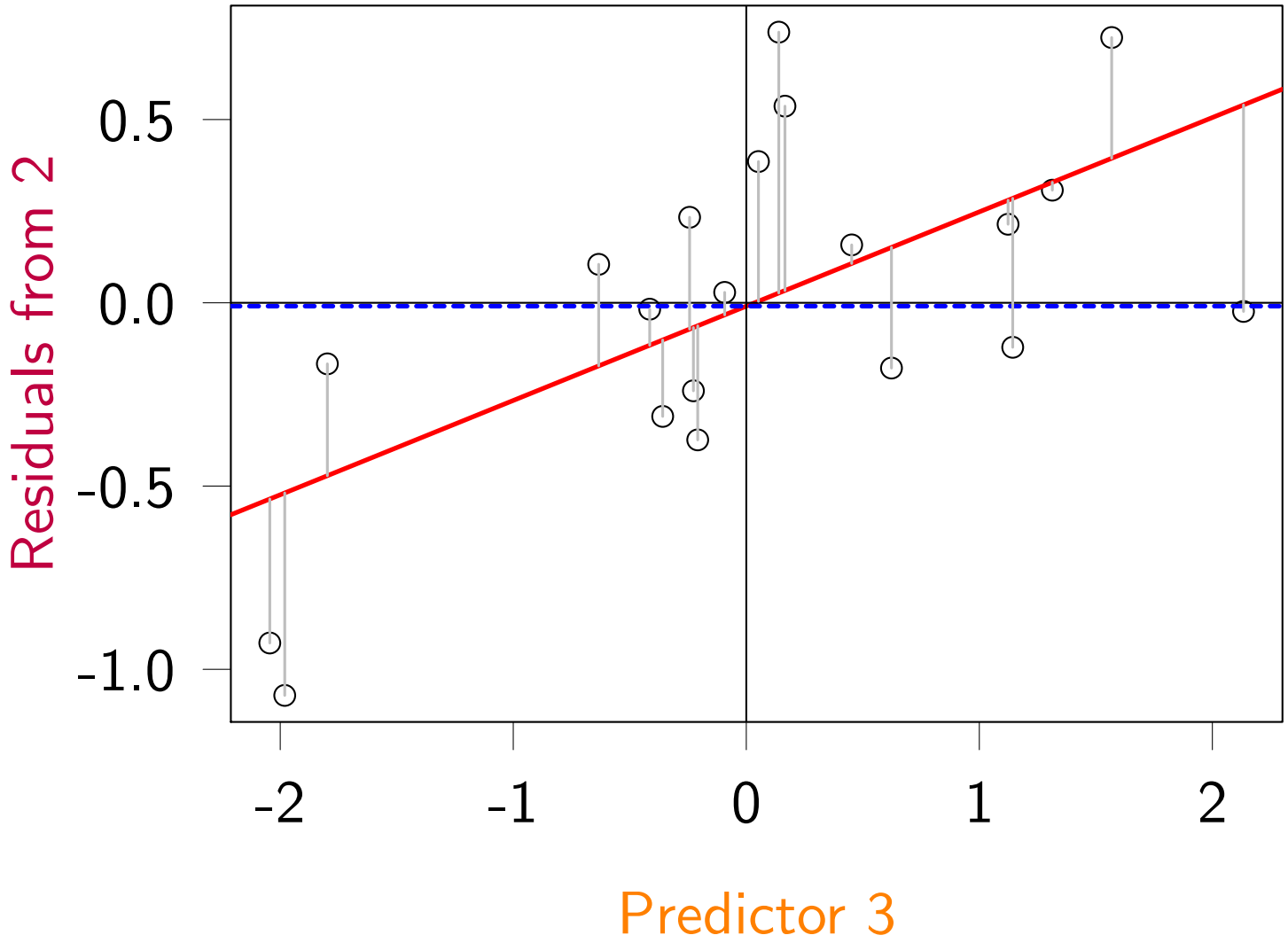
Sequential regression

We want to explain a response by three predictors



Sequential regression

We want to explain a response by three predictors



Sequential regression

```
m1 <- lm(y ~ x1)
m2 <- lm(m1$residuals ~ x2)
m3 <- lm(m2$residuals ~ x3)
```

Sequential regression

But,

```
  m1 <- lm(y ~ x1)
  m2 <- lm(m1$residuals ~ x2)
  m3 <- lm(m2$residuals ~ x3)

  coefficients(m1); coefficients(m2); coefficients(m3)
```

```
(Intercept)          x1
  0.4585067   0.4715738
(Intercept)          x2
-0.02542972  0.75420781
 (Intercept)          x3
-0.009178959  0.257305942
```

is different from

```
    m1 <- lm(y ~ x3)
    m2 <- lm(m1$residuals ~ x2)
    m3 <- lm(m2$residuals ~ x1)

    coefficients(m1); coefficients(m2); coefficients(m3)
```

Sequential regression

Also what happens with classical ANOVA (aov in R)

```
  summary(aov(y ~ x1 + x2 + x3))

            Df Sum Sq Mean Sq F value   Pr(>F)
x1           1  3.997   3.997  394.05 1.07e-12 ***
x2           1 13.998  13.998 1379.87  < 2e-16 ***
x3           1  0.120   0.120   11.82  0.00338 **
Residuals   16  0.162   0.010
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

  summary(aov(y ~ x2 + x3 + x1))

            Df Sum Sq Mean Sq  F value  Pr(>F)
x2           1 17.931  17.931 1767.562 < 2e-16 ***
x3           1  0.183   0.183   18.003 0.00062 ***
x1           1  0.002   0.002    0.176 0.68076
Residuals   16  0.162   0.010
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Simultaneous regression

In contrast lm() optimizes relationships simultaneously
Hence order does not matter:

```
  coefficients(lm(y ~ x1 + x2 + x3))

(Intercept)          x1          x2          x3
 0.48948612 -0.01357404  1.03015700  0.08395938

  coefficients(lm(y ~ x2 + x3 + x1))

(Intercept)          x2          x3          x1
 0.48948612  1.03015700  0.08395938 -0.01357404
```

Simultaneous regression

```
    vcov(lm(y ~ x1 + x2 + x3))

             (Intercept)            x1            x2            x3
(Intercept)  5.295645e-04 -0.0001478660  3.906623e-05 -7.631575e-05
x1          -1.478660e-04  0.0010493201 -4.483007e-04  4.028972e-04
x2           3.906623e-05 -0.0004483007  7.877264e-04 -5.987783e-05
x3          -7.631575e-05  0.0004028972 -5.987783e-05  5.963792e-04
```

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?"*. Adds a predictor (or several) to a model.

Warnings

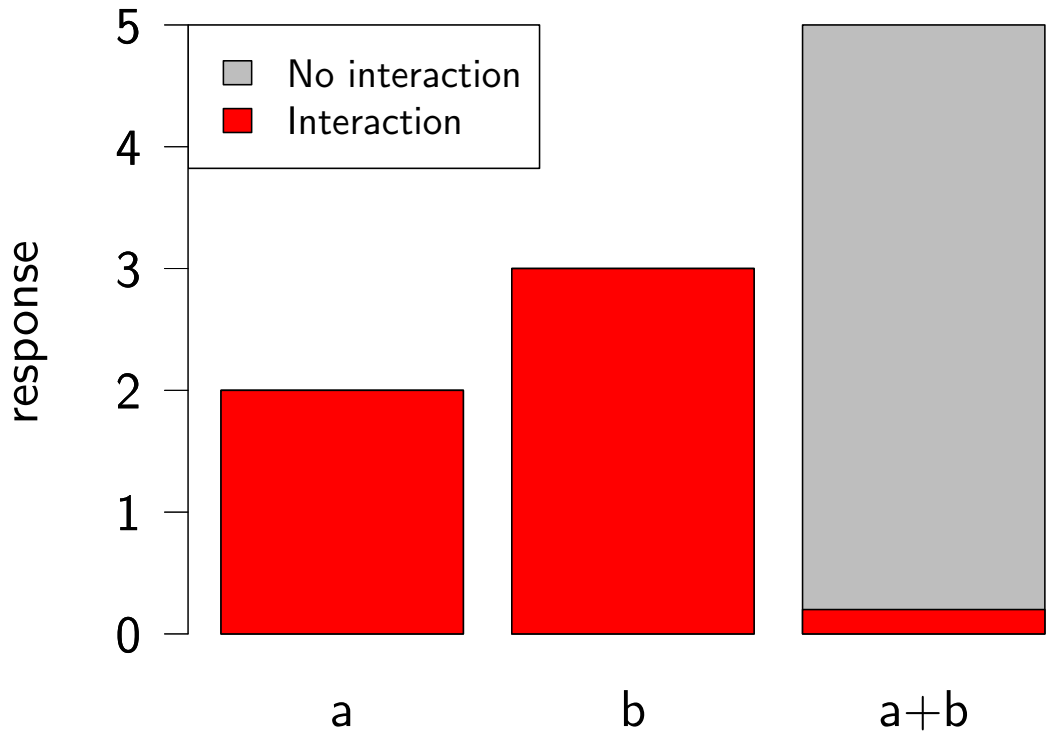
Vocabulary warning!

- **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?"*. Adds a predictor (or several) to a model.

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?"*. Adds a predictor (or several) to a model.



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    0.09578  11.913  < 2e-16 ***
x1          -0.49281    0.10834  -4.549 1.58e-05 ***
x2           0.53434    0.09881   5.408 4.67e-07 ***
x1Xx2        0.35911    0.11449   3.137  0.00227 **
---
```

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.

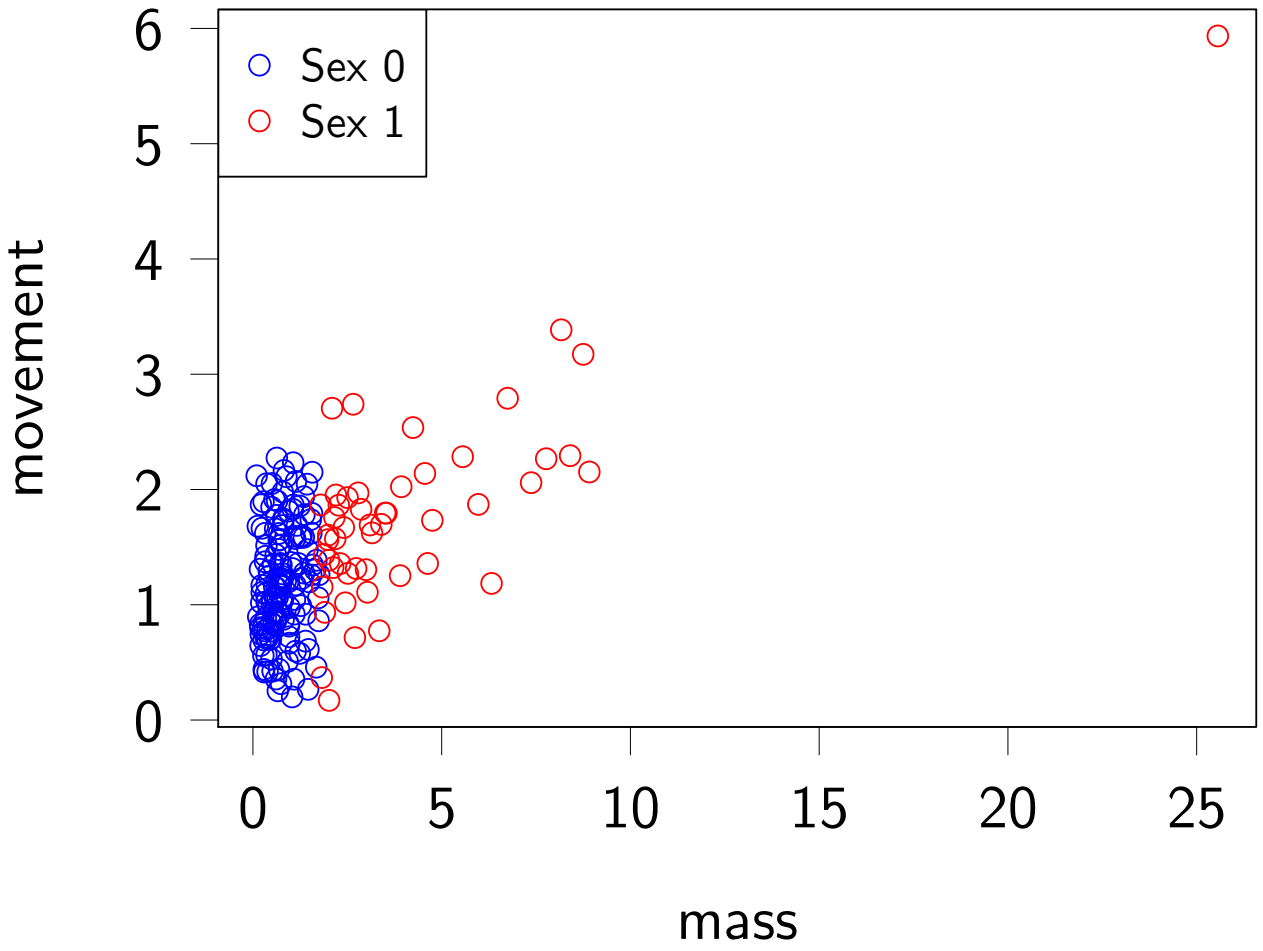
```
summary(lm(movement ~ mass, data=masssex[masssex$sex==0,]))
summary(lm(movement ~ mass, data=masssex[masssex$sex==1,]))
```

3.

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

Warnings

4.



Prediction

Exercise

1. Load plantsize.csv and plot the data
2. Fit an additive model explaining plant size by x and y coordinates

```
plantsize <- read.csv("plantsize.csv")
m0 <- lm(plantsize ~ x_location + y_location, data=plantsize)
```

Prediction

Exercise

1. Load plantsize.csv and plot the data
2. Fit an additive model explaining plant size by x and y coordinates
3. Create a prediction for plant size as a function of x for two values of y

```
plantsize <- read.csv("plantsize.csv")
m0 <- lm(plantsize ~ x_location + y_location, data=plantsize)
```

Prediction

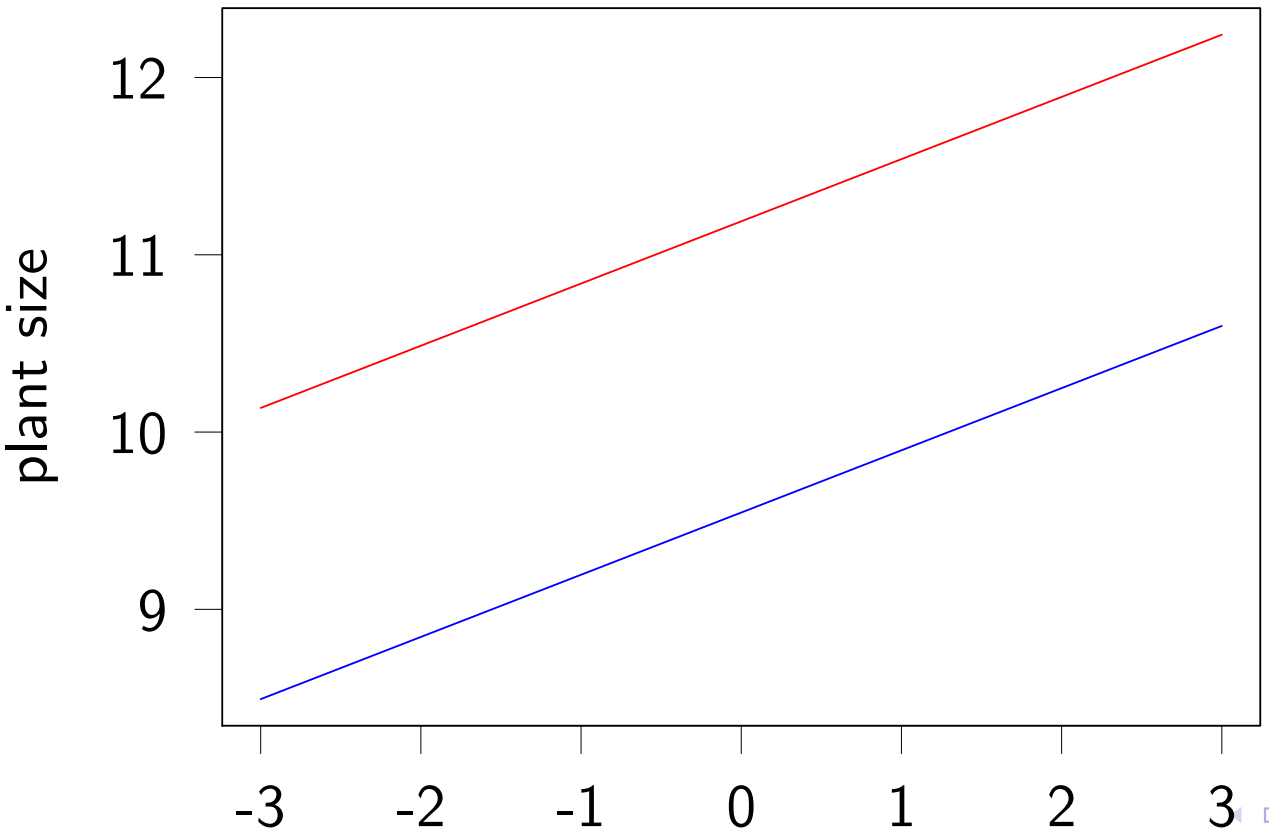
3.1. Predict

```
newdata <- data.frame(x_location = rep(seq(-3,3, length.out = 100),2),
                      y_location = c(rep(-3, 100), rep(4,100)))
newdata$prediction <- predict(m0, newdata = newdata)
```

Prediction

3.2 Visualize

```
setPar()
plot(newdata$x_location[newdata$y_location==-3], newdata$prediction[newda
     xlab="x location", ylab="plant size", type="l", ylim = range(newdata
lines(newdata$x_location[newdata$y_location==4], newdata$prediction[newda
```

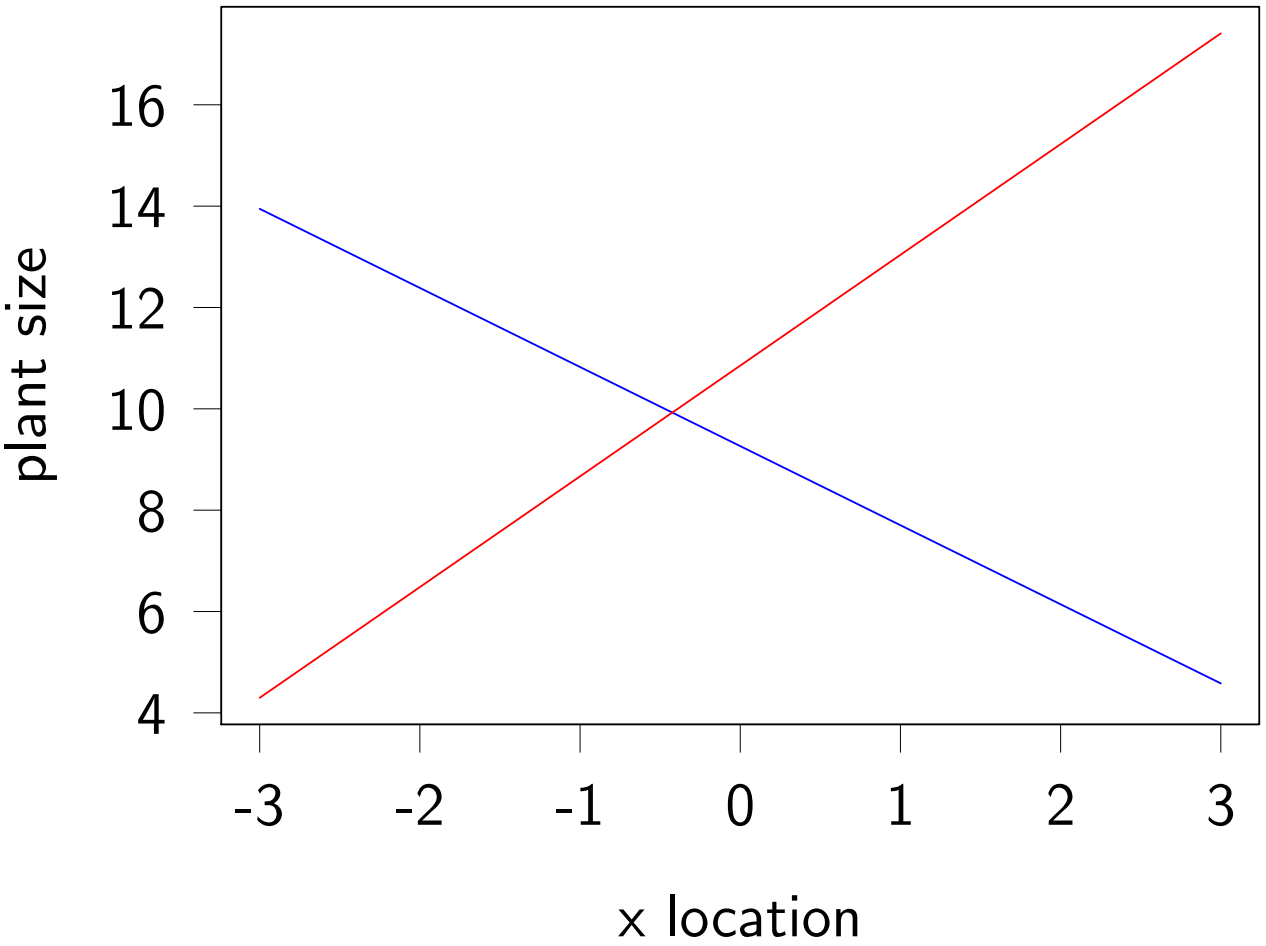


Prediction with interaction

Exercise

1. Load plantsize.csv and plot the data
2. Fit an additive model explaining plant size by x and y coordinates
3. Create a prediction for plant size as a function of x for two values of y and plot it
4. Fit an interaction between x and y coordinates
5. Create a new prediction with interaction, and plot it

Prediction with interaction

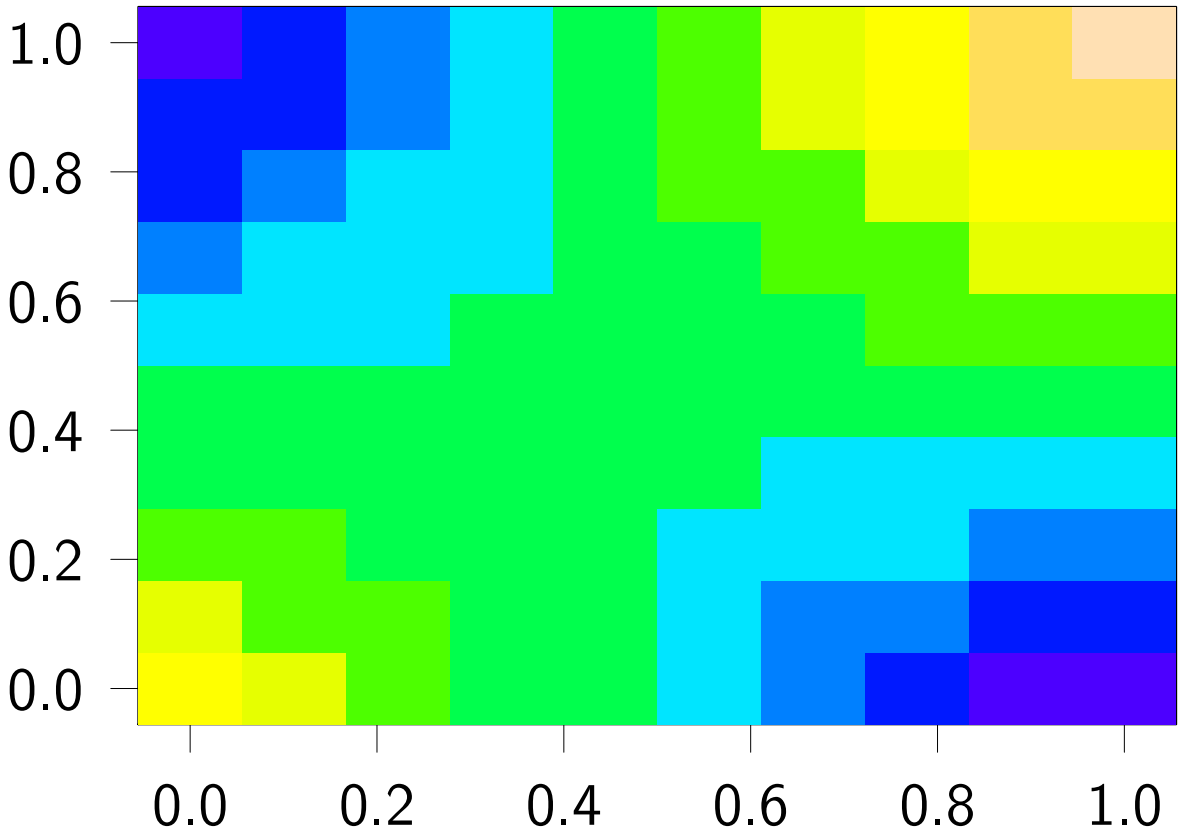


Prediction with interaction

Exercise

1. Load plantsize.csv and plot the data
2. Fit an additive model explaining plant size by x and y coordinates
3. Create a prediction for plant size as a function of x for two values of y and plot it
4. Fit an interaction between x and y coordinates
5. Create a new prediction with interaction, and plot it
6. Compare estimates and p-values across models. Do you think x location has an effect or not?

Prediction with interaction



Next times

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