Biostatistics I: Linear Regression

Interactions

Nicole S. Erler
Department of Biostatistics, Erasmus Medical Center
n.erler@erasmusmc.nl
@N_Erler

The Multiple Linear Regression Model

$$\mathbf{y} = \underbrace{\beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \ldots + \beta_p \mathbf{x}_p}_{\substack{\text{additive linear systematic component} \\ \text{(linear predictor)}}} + \boldsymbol{\varepsilon} \qquad \mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}, \quad \mathrm{var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$$

Requirement:
The model is **linear in the regression coefficients** and the error term.

In general:
As long as we can write the model as $y_i = f(\mathbf{x}_i)^\top \boldsymbol{\beta} + \varepsilon_i$ we have a linear model.

The Multiple Linear Regression Model

$$\mathbf{y} = \underbrace{\beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \ldots + \beta_p \mathbf{x}_p}_{\substack{\text{additive linear systematic component} \\ \text{(linear predictor)}}} + \boldsymbol{\varepsilon} \qquad \mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}, \quad \mathrm{var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$$

Requirement:
The model is **linear in the regression coefficients** and the error term.

In general:
As long as we can write the model as $y_i = f(\mathbf{x}_i)^\top \boldsymbol{\beta} + \varepsilon_i$ we have a linear model.

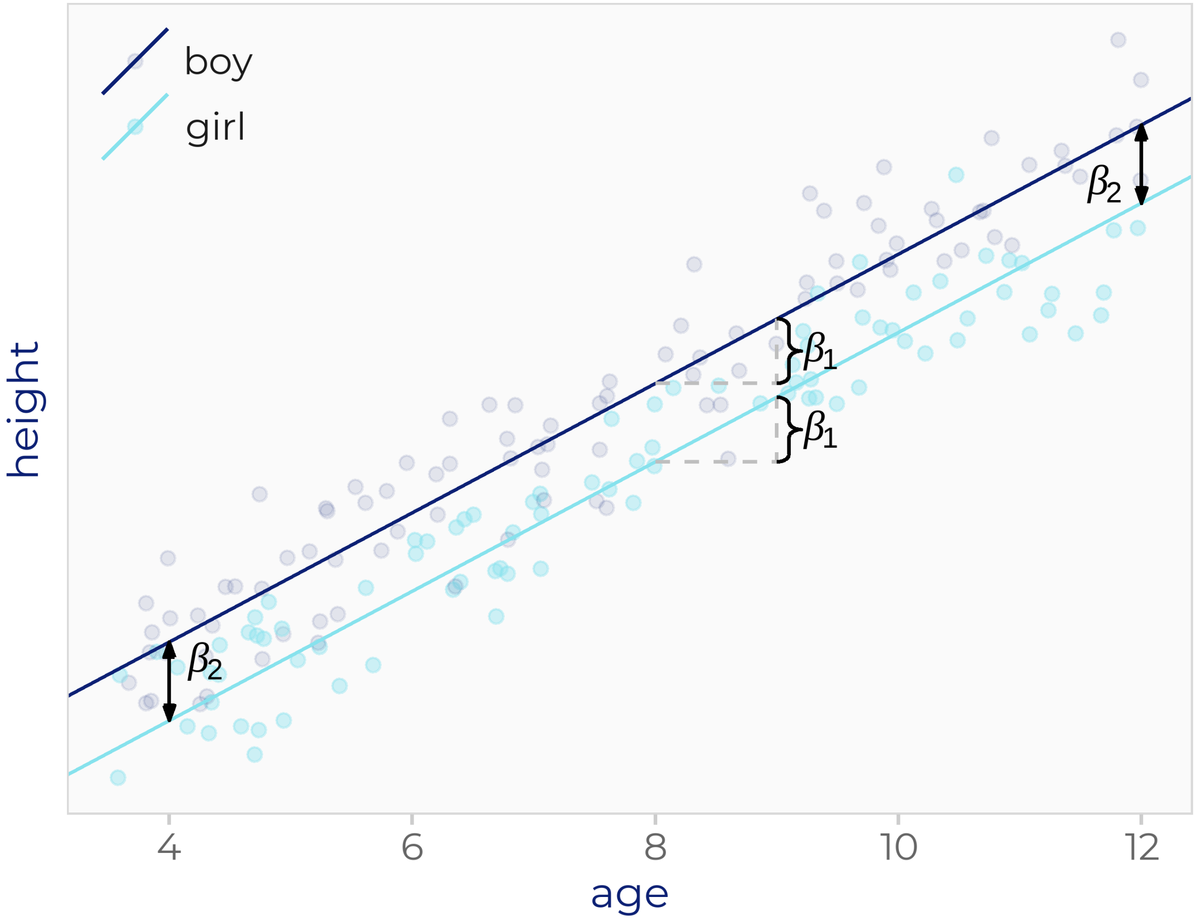
Interpretation:
β_j is the expected change in y that is associated with an increase in x_j (or $f(x_j)$) of 1 unit **while all other covariates are kept constant**.

Example: Child Growth

$$\text{height}_i = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{sex}_i + \varepsilon_i$$

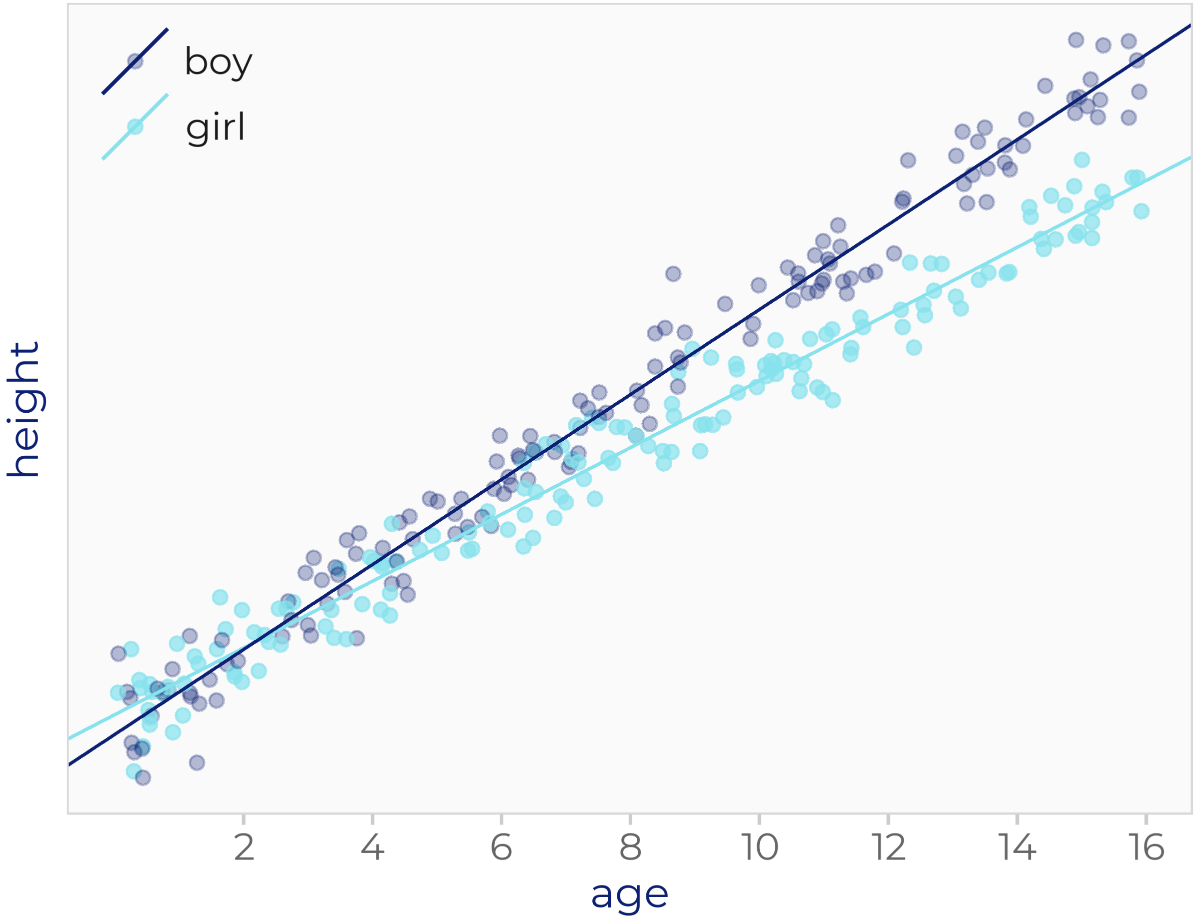
Implied assumption:
Boys and girls grow equally fast.

⇨ The regression lines for boys and girls are **parallel**.



Example: Child Growth

The regression lines for boys and girls are **not parallel**.

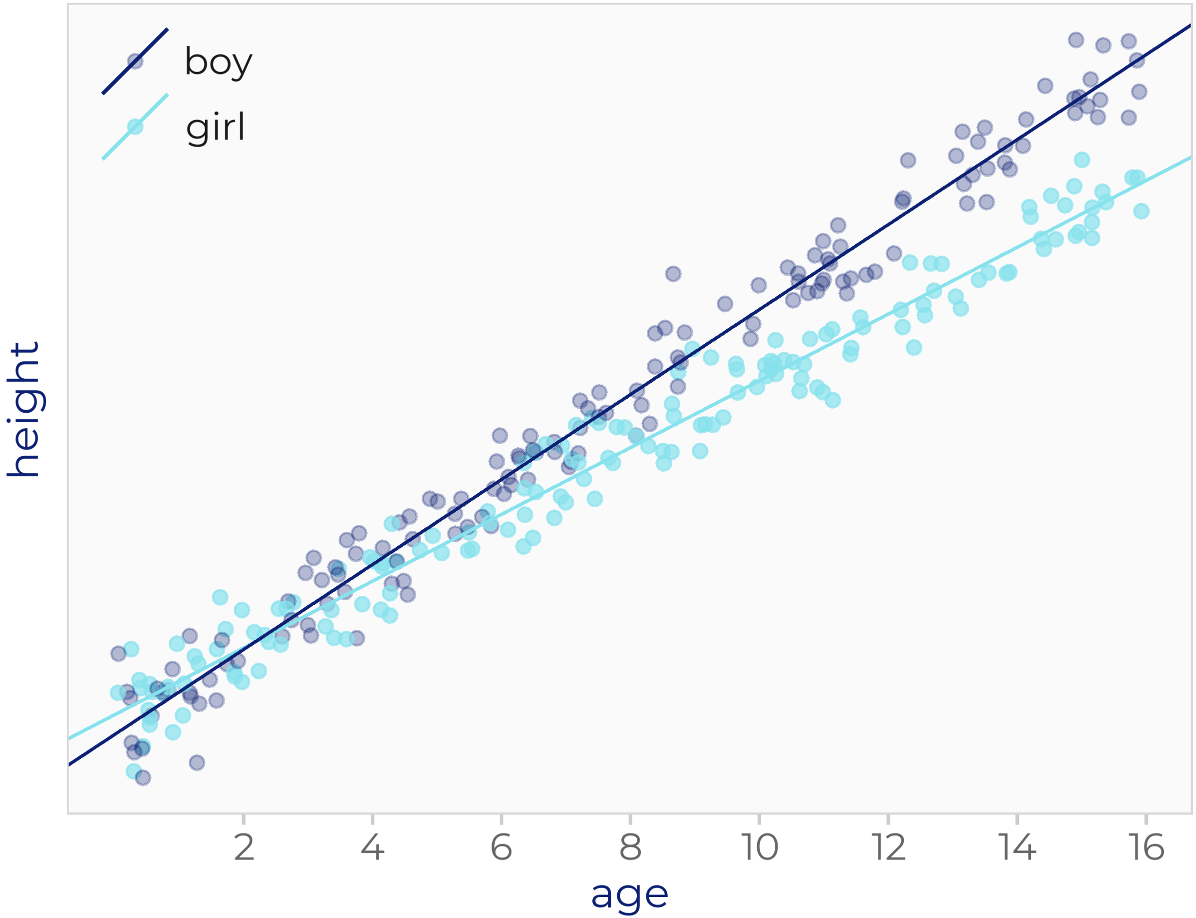


Example: Child Growth

The regression lines for boys and girls are **not parallel**.

⇨Age and sex **interact**.

Age **modifies** the effect of sex.
Sex **modifies** the effect of age.



Interaction Terms

To relax the assumption of independent effects, we can include **interaction terms** by adding new terms that are the **product of two (or more) covariates**:

$$\text{height}_i = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{sex}_i + \beta_3 \text{age}_i \, \text{sex}_i + \varepsilon_i$$

Interaction Terms

To relax the assumption of independent effects, we can include **interaction terms** by adding new terms that are the **product of two (or more) covariates**:

$$\text{height}_i = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{sex}_i + \beta_3 \text{age}_i \, \text{sex}_i + \varepsilon_i$$

When the model includes an interaction term, **the interpretation** of the corresponding effects **changes**:

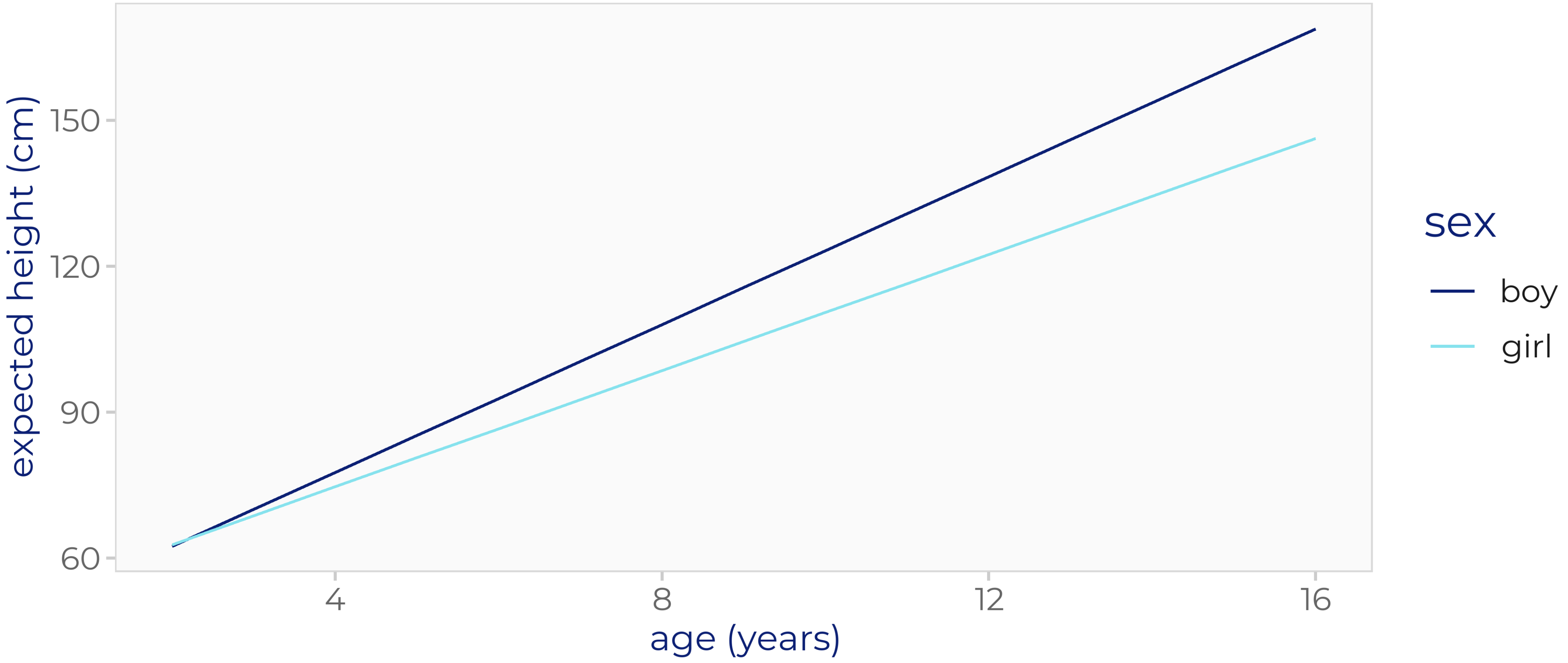
	β
(Intercept)	47.3
age	7.6
sex_{girl}	3.6
$\text{age} \times \text{sex}_{\text{girl}}$	-1.6

- At birth (age = 0) boys are, on average, 47.3 cm tall.
- At birth (age = 0) girls are 3.6 cm taller than boys.
- Boys grow 7.6 cm per year.
- Girls grow 7.6 - 1.6 = 6.0 cm per year.
- Boys grow 1.6 cm per year faster than girls.

⇨ We **cannot interpret** the effects of age and sex **separately** from each other.

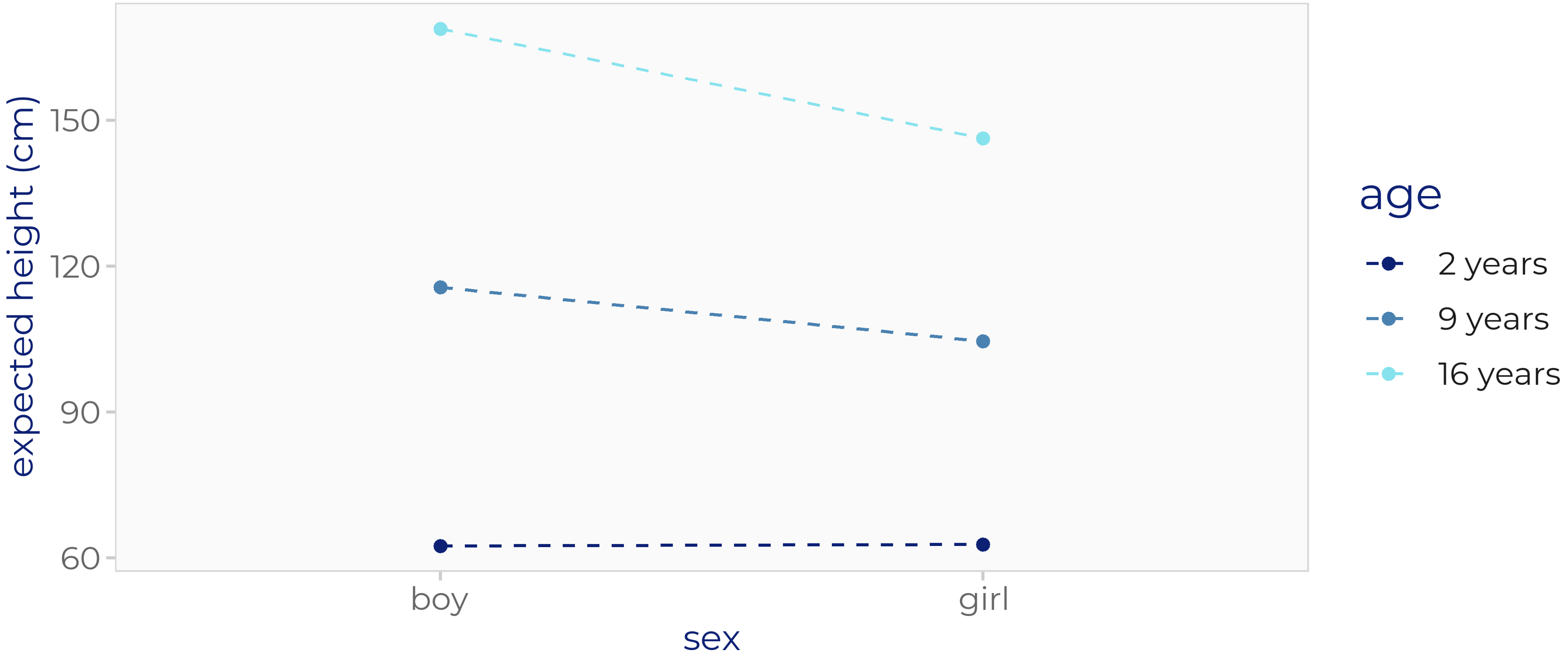
Interaction: Visualization (1)

Sex modifies the effect of age:



Interaction: Visualization (2)

Age modifies the effect of sex:



Interactions between Multiple Covariates

$$\text{height}_i = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{sex}_i + \beta_3 \text{kcal}_i + \beta_4 \text{age}_i\, \text{sex}_i + \beta_5 \text{age}_i\, \text{kcal}_i + \varepsilon_i$$

Interactions between Multiple Covariates

$$\text{height}_i = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{sex}_i + \beta_3 \text{kcal}_i + \beta_4 \text{age}_i \, \text{sex}_i + \beta_5 \text{age}_i \, \text{kcal}_i + \varepsilon_i$$

Interpretation of the coefficients: (assuming dummy coding for sex)

- β_0: expected height of a boy (reference category) with age = 0 and kcal = 0

Interactions between Multiple Covariates

$$\text{height}_i = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{sex}_i + \beta_3 \text{kcal}_i + \beta_4 \text{age}_i \, \text{sex}_i + \beta_5 \text{age}_i \, \text{kcal}_i + \varepsilon_i$$

Interpretation of the coefficients: (assuming dummy coding for sex)

- β_0: expected height of a boy (reference category) with age = 0 and kcal = 0
- β_1: effect of age when kcal = 0 and sex = 0 (reference category)

Interactions between Multiple Covariates

$$\text{height}_i = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{sex}_i + \beta_3 \text{kcal}_i + \beta_4 \text{age}_i\, \text{sex}_i + \beta_5 \text{age}_i\, \text{kcal}_i + \varepsilon_i$$

Interpretation of the coefficients: (assuming dummy coding for sex)

- β_0: expected height of a boy (reference category) with age = 0 and kcal = 0
- β_1: effect of age when kcal = 0 and sex = 0 (reference category)
- β_2: effect of sex when age = 0

Interactions between Multiple Covariates

$$\text{height}_i = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{sex}_i + \beta_3 \text{kcal}_i + \beta_4 \text{age}_i\, \text{sex}_i + \beta_5 \text{age}_i\, \text{kcal}_i + \varepsilon_i$$

Interpretation of the coefficients: (assuming dummy coding for sex)

- β_0: expected height of a boy (reference category) with age = 0 and kcal = 0
- β_1: effect of age when kcal = 0 and sex = 0 (reference category)
- β_2: effect of sex when age = 0
- β_3: effect of kcal when age = 0

Interactions between Multiple Covariates

$$\text{height}_i = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{sex}_i + \beta_3 \text{kcal}_i + \beta_4 \text{age}_i\, \text{sex}_i + \beta_5 \text{age}_i\, \text{kcal}_i + \varepsilon_i$$

Interpretation of the coefficients: (assuming dummy coding for sex)

- β_0: expected height of a boy (reference category) with age = 0 and kcal = 0
- β_1: effect of age when kcal = 0 and sex = 0 (reference category)
- β_2: effect of sex when age = 0
- β_3: effect of kcal when age = 0
- β_4: change in the effect of age when sex = 1 vs when sex = 0, or
  change in the effect of sex when age is increase by 1 unit

Interactions between Multiple Covariates

$$\text{height}_i = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{sex}_i + \beta_3 \text{kcal}_i + \beta_4 \text{age}_i\, \text{sex}_i + \beta_5 \text{age}_i\, \text{kcal}_i + \varepsilon_i$$

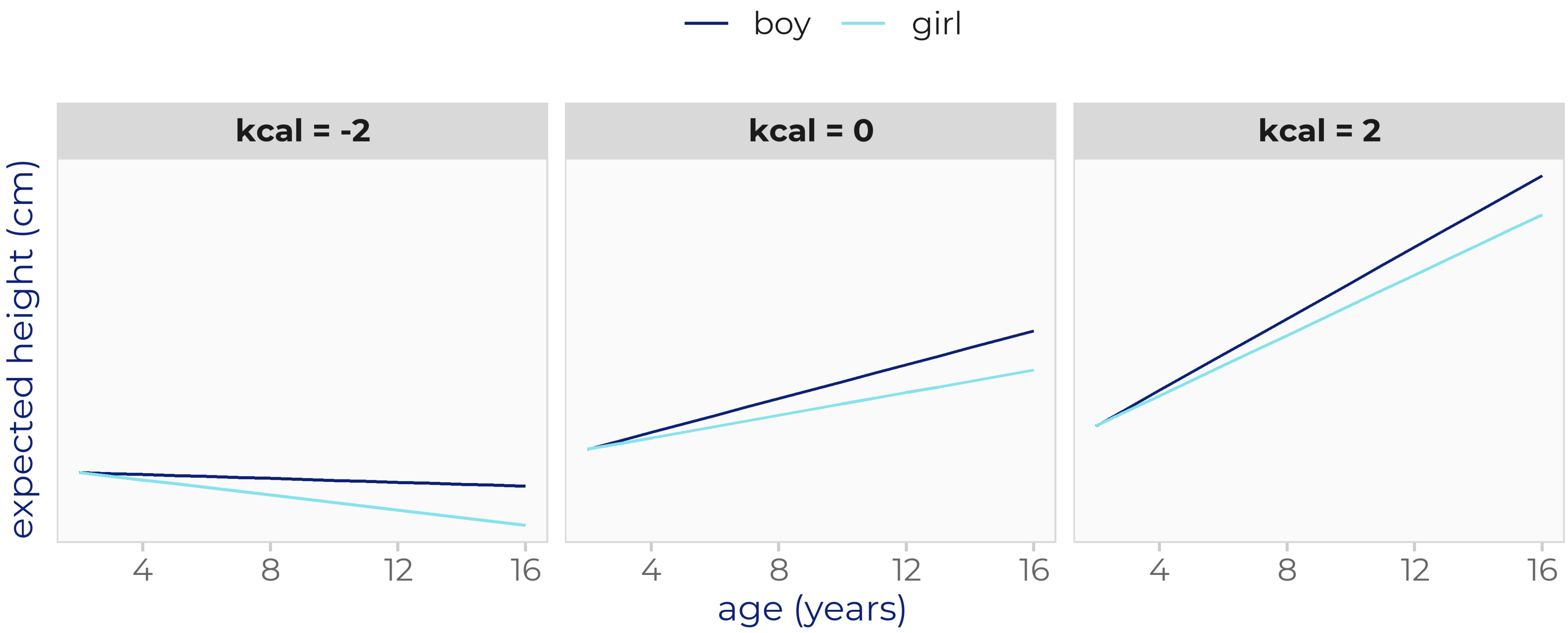
Interpretation of the coefficients: (assuming dummy coding for sex)

- β_0: expected height of a boy (reference category) with age = 0 and kcal = 0
- β_1: effect of age when kcal = 0 and sex = 0 (reference category)
- β_2: effect of sex when age = 0
- β_3: effect of kcal when age = 0
- β_4: change in the effect of age when sex = 1 vs when sex = 0, or
  change in the effect of sex when age is increase by 1 unit
- β_5: change in the effect of age when kcal is increased by 1 unit, or
  change in the effect of kcal when age is increase by 1 unit

... and everything else remains constant.

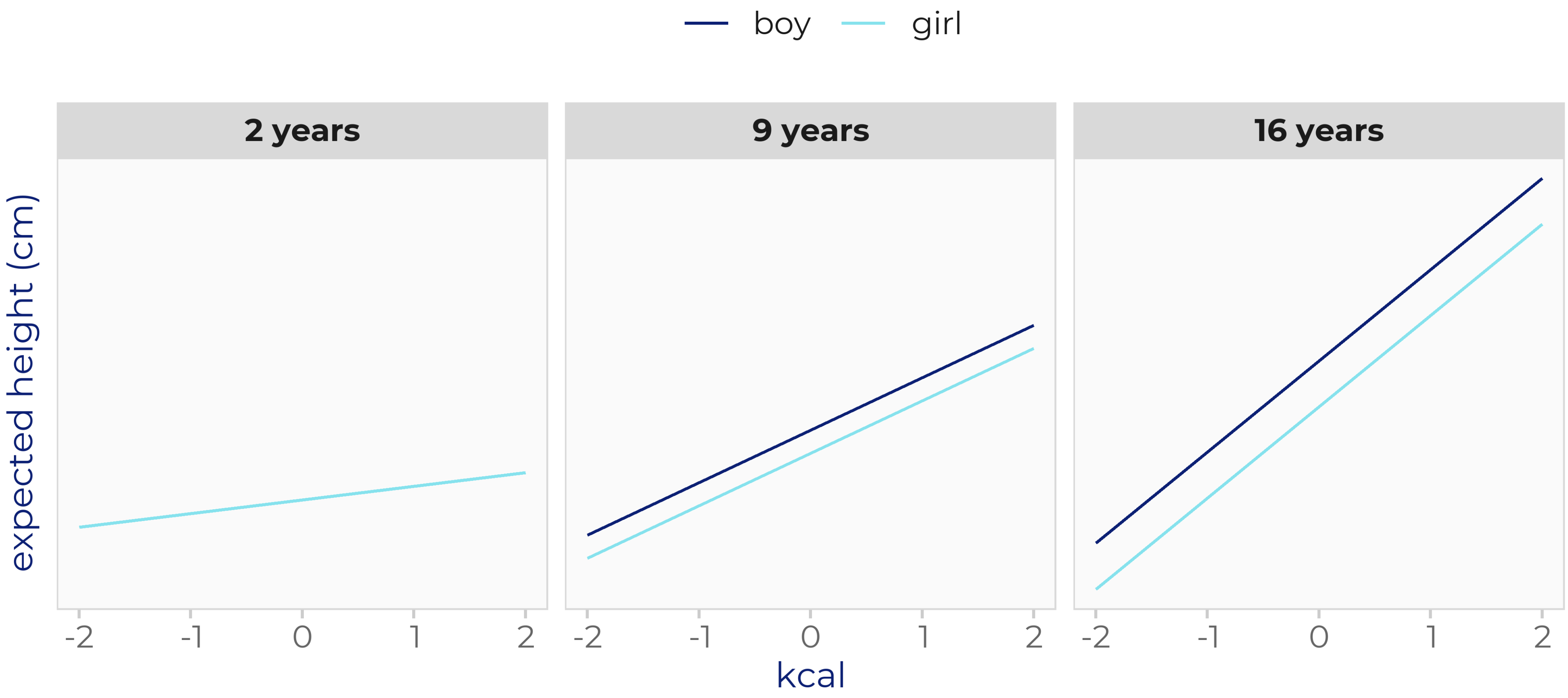
Interactions between Multiple Covariates

Sex and kcal modify the effect of age:



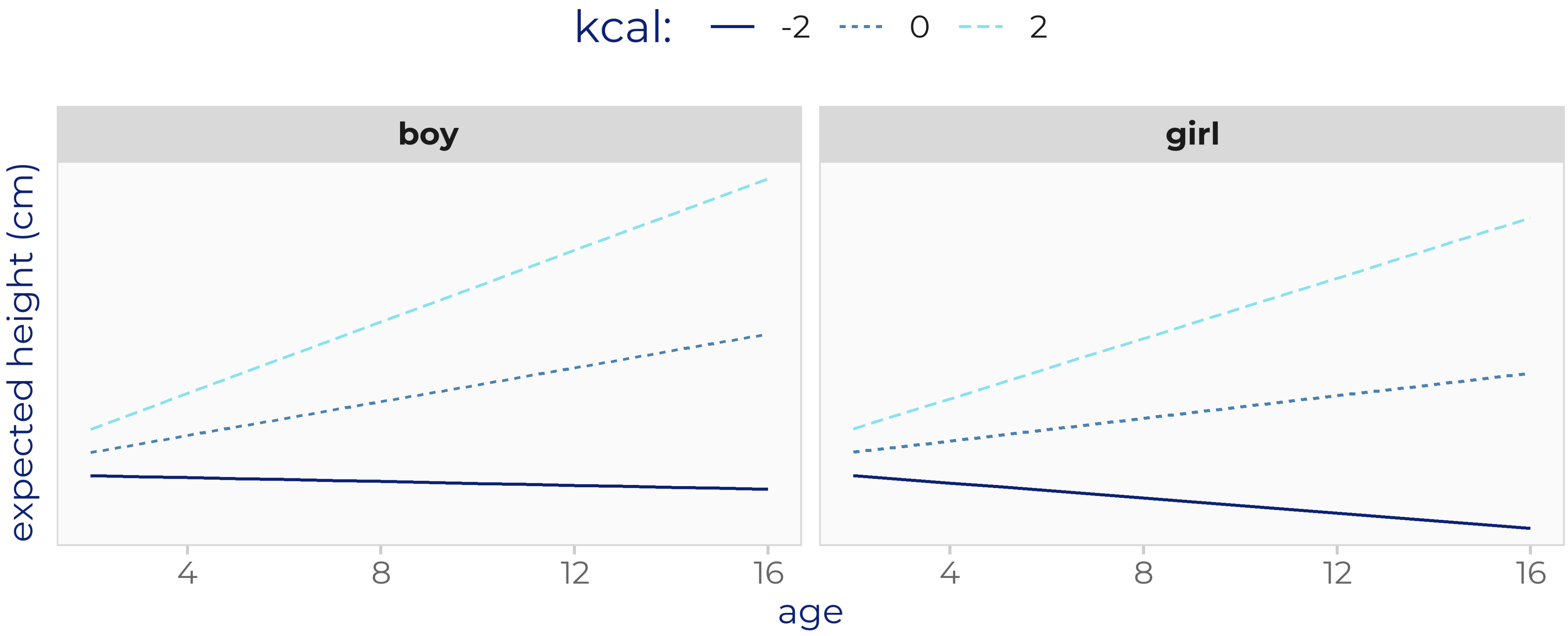
Interactions between Multiple Covariates

Age modifies the effect of kcal:



Interactions between Multiple Covariates

Kcal modifies the effect of age:



Higher Order Interactions

For example, a 3-way interaction between age, sex and kcal (and interactions between age and sex and age and kcal):

$$\beta_0 + \beta_1 \mathbf{age}_i + \beta_2 \mathbf{sex}_i + \beta_3 \mathbf{kcal}_i + \beta_4 \mathbf{age}_i \, \mathbf{sex}_i + \beta_5 \mathbf{age}_i \, \mathbf{kcal}_i + \beta_6 \mathbf{age}_i \, \mathbf{sex}_i \, \mathbf{kcal}_i$$

Interpretation of the coefficients:

- Does the 3-way interaction influence the interpretation of β_4 and β_5?
- What is the interpretation of β_6?

Higher Order Interactions: Interpretation

	age	**sex**	**kcal**	β_0	$\beta_1$age	$\beta_2$sex	$\beta_3$kcal	$\beta_4$age sex	$\beta_5$age kcal	$\beta_6$age sex kcal
scenario 1	0	0	0	β_0						
scenario 2	1	0	0	β_0	$+ \beta_1$					
scenario 3	0	1	0	β_0		$+ \beta_2$				
scenario 4	0	0	1	β_0			$+ \beta_3$			
scenario 5	1	1	0	β_0	$+ \beta_1$	$+ \beta_2$		$+ \beta_4$		
scenario 6	1	0	1	β_0	$+ \beta_1$		$+ \beta_3$		$+ \beta_5$	
scenario 7	0	1	1	β_0		$+ \beta_2$	$+ \beta_3$			
scenario 8	1	1	1	β_0	$+ \beta_1$	$+ \beta_2$	$+ \beta_3$	$+ \beta_4$	$+ \beta_5$	$+ \beta_6$

Higher Order Interactions: Interpretation

	age	sex	kcal	β_0		$\beta_1$age		$\beta_2$sex		$\beta_3$kcal		$\beta_4$age sex		$\beta_5$age kcal		$\beta_6$age sex kcal
scenario 1	0	0	0	β_0												
scenario 2	1	0	0	β_0	$+$	β_1										
scenario 3	0	1	0	β_0			$+$	β_2								
scenario 4	0	0	1	β_0					$+$	β_3						
scenario 5	1	1	0	β_0	$+$	β_1	$+$	β_2			$+$	β_4				
scenario 6	1	0	1	β_0	$+$	β_1			$+$	β_3			$+$	β_5		
scenario 7	0	1	1	β_0			$+$	β_2	$+$	β_3						
scenario 8	1	1	1	β_0	$+$	β_1	$+$	β_2	$+$	β_3	$+$	β_4	$+$	β_5	$+$	β_6

Effect of age when

sex = 0 and kcal = 0 (scenario 2 vs 1): β_1

sex = 1 and kcal = 0 (scenario 5 vs 3): $\beta_1 + \beta_4$

sex = 0 and kcal = 1 (scenario 6 vs 4): $\beta_1 + \beta_5$

sex = 1 and kcal = 1 (scenario 8 vs 7): $\beta_1 + \beta_4 + \beta_5 + \beta_6$

Higher Order Interactions: Interpretation

Effect of age when

sex = 0 and kcal = 0	(scenario 2 vs 1):	β_1
sex = 1 and kcal = 0	(scenario 5 vs 3):	$\beta_1 + \beta_4$
sex = 0 and kcal = 1	(scenario 6 vs 4):	$\beta_1 + \beta_5$
sex = 1 and kcal = 1	(scenario 8 vs 7):	$\beta_1 + \beta_4 + \beta_5 + \beta_6$

⇨ β_4 : change in the effect of age when sex = 1 vs sex = 0 (and kcal = 0)
(age-sex interaction when kcal = 0)

⇨ $\beta_4 + \beta_6$: change in the effect of age when sex = 1 vs sex = 0 (and kcal = 1)
(age-sex interaction when kcal = 1)

Higher Order Interactions: Interpretation

Effect of age when

sex = 0 and kcal = 0	(scenario 2 vs 1):	β_1
sex = 1 and kcal = 0	(scenario 5 vs 3):	$\beta_1 + \beta_4$
sex = 0 and kcal = 1	(scenario 6 vs 4):	$\beta_1 + \beta_5$
sex = 1 and kcal = 1	(scenario 8 vs 7):	$\beta_1 + \beta_4 + \beta_5 + \beta_6$

⇨ β_4 : change in the effect of age when sex = 1 vs sex = 0 (and kcal = 0)
(age-sex interaction when kcal = 0)

⇨ $\beta_4 + \beta_6$: change in the effect of age when sex = 1 vs sex = 0 (and kcal = 1)
(age-sex interaction when kcal = 1)

⇨ β_6 : **change in age-sex interaction** when kcal is increased by one unit

Higher Order Interactions: Interpretation

Effect of age when

sex = 0 and kcal = 0	(scenario 2 vs 1):	β_1
sex = 1 and kcal = 0	(scenario 5 vs 3):	$\beta_1 + \beta_4$
sex = 0 and kcal = 1	(scenario 6 vs 4):	$\beta_1 + \beta_5$
sex = 1 and kcal = 1	(scenario 8 vs 7):	$\beta_1 + \beta_4 + \beta_5 + \beta_6$

⇨ β_5 : change in effect of age when kcal = 1 vs kcal = 0 (and sex = 0)
(age-kcal interaction when sex = 0)

⇨ $\beta_5 + \beta_6$: change in effect of age when kcal = 1 vs kcal = 0 (and sex = 1)
(age-kcal interaction when sex = 1)

Higher Order Interactions: Interpretation

Effect of age when

sex = 0 and kcal = 0	(scenario 2 vs 1):	β_1
sex = 1 and kcal = 0	(scenario 5 vs 3):	$\beta_1 + \beta_4$
sex = 0 and kcal = 1	(scenario 6 vs 4):	$\beta_1 + \beta_5$
sex = 1 and kcal = 1	(scenario 8 vs 7):	$\beta_1 + \beta_4 + \beta_5 + \beta_6$

⇨ β_5 : change in effect of age when kcal = 1 vs kcal = 0 (and sex = 0)
(age-kcal interaction when sex = 0)

⇨ $\beta_5 + \beta_6$: change in effect of age when kcal = 1 vs kcal = 0 (and sex = 1)
(age-kcal interaction when sex = 1)

⇨ β_6 : **change in age-kcal interaction** when sex = 1 vs sex = 0

Higher Order Interactions: Interpretation

	age	sex	kcal	β_0		$\beta_1$age		$\beta_2$sex		$\beta_3$kcal		$\beta_4$age sex		$\beta_5$age kcal		$\beta_6$age sex kcal
scenario 1	0	0	0	β_0												
scenario 2	1	0	0	β_0	$+$	β_1										
scenario 3	0	1	0	β_0			$+$	β_2								
scenario 4	0	0	1	β_0					$+$	β_3						
scenario 5	1	1	0	β_0	$+$	β_1	$+$	β_2			$+$	β_4				
scenario 6	1	0	1	β_0	$+$	β_1			$+$	β_3			$+$	β_5		
scenario 7	0	1	1	β_0			$+$	β_2	$+$	β_3						
scenario 8	1	1	1	β_0	$+$	β_1	$+$	β_2	$+$	β_3	$+$	β_4	$+$	β_5	$+$	β_6

Effect of sex when

age = 0 and kcal = 0 (scenario 3 vs 1): β_2

age = 1 and kcal = 0 (scenario 5 vs 2): $\beta_2 + \beta_4$

age = 0 and kcal = 1 (scenario 7 vs 4): β_2

age = 1 and kcal = 1 (scenario 8 vs 6): $\beta_2 + \beta_4 + \beta_6$

Higher Order Interactions: Interpretation

Effect of sex when

age = 0 and kcal = 0	(scenario 3 vs 1):	β_2
age = 1 and kcal = 0	(scenario 5 vs 2):	$\beta_2 + \beta_4$
age = 0 and kcal = 1	(scenario 7 vs 4):	β_2
age = 1 and kcal = 1	(scenario 8 vs 6):	$\beta_2 + \beta_4 + \beta_6$

⇨ β_4 : change in effect of sex when age = 1 vs age = 0 (and kcal = 0)
(sex-age interaction when kcal = 0)

⇨ $\beta_4 + \beta_6$: change in effect of sex when age = 1 vs age = 0 (and kcal = 1)
(sex-age interaction when kcal = 1)

Higher Order Interactions: Interpretation

Effect of sex when

age = 0 and kcal = 0	(scenario 3 vs 1):	β_2
age = 1 and kcal = 0	(scenario 5 vs 2):	$\beta_2 + \beta_4$
age = 0 and kcal = 1	(scenario 7 vs 4):	β_2
age = 1 and kcal = 1	(scenario 8 vs 6):	$\beta_2 + \beta_4 + \beta_6$

⇨ β_4 : change in effect of sex when age = 1 vs age = 0 (and kcal = 0)

(sex-age interaction when kcal = 0)

⇨ $\beta_4 + \beta_6$: change in effect of sex when age = 1 vs age = 0 (and kcal = 1)

(sex-age interaction when kcal = 1)

⇨ β_6 : **change in sex-age interaction** when kcal is increased by one unit

Higher Order Interactions: Interpretation

Effect of sex when

age = 0 and kcal = 0	(scenario 3 vs 1):	β_2
age = 1 and kcal = 0	(scenario 5 vs 2):	$\beta_2 + \beta_4$
age = 0 and kcal = 1	(scenario 7 vs 4):	β_2
age = 1 and kcal = 1	(scenario 8 vs 6):	$\beta_2 + \beta_4 + \beta_6$

⇨ β_6 : change in effect of sex when kcal = 1 vs kcal = 0 (and age = 1)
(sex-kcal interaction when age = 1)

Since there is no sex-kcal interaction when age = 0:

⇨ β_6 : **change in sex-kcal interaction** when age is increased by one unit

Higher Order Interactions: Interpretation

The interpretation with respect to kcal is analogue to the interpretation with respect to sex.

Higher Order Interactions: Interpretation

The interpretation with respect to kcal is analogue to the interpretation with respect to sex.

Interpretation of the 3-way interaction β_6:

- Change in the age-sex interaction when kcal is increased by 1 unit, or
- change in the age-kcal interaction when sex is 1 vs 0, or
- change in the effect of sex when kcal is increased by 1 unit and age $\neq$ 0, or
- change in the sex-kcal interaction when age is increased by 1 unit.

Higher Order Interactions: Interpretation

The interpretation with respect to kcal is analogue to the interpretation with respect to sex.

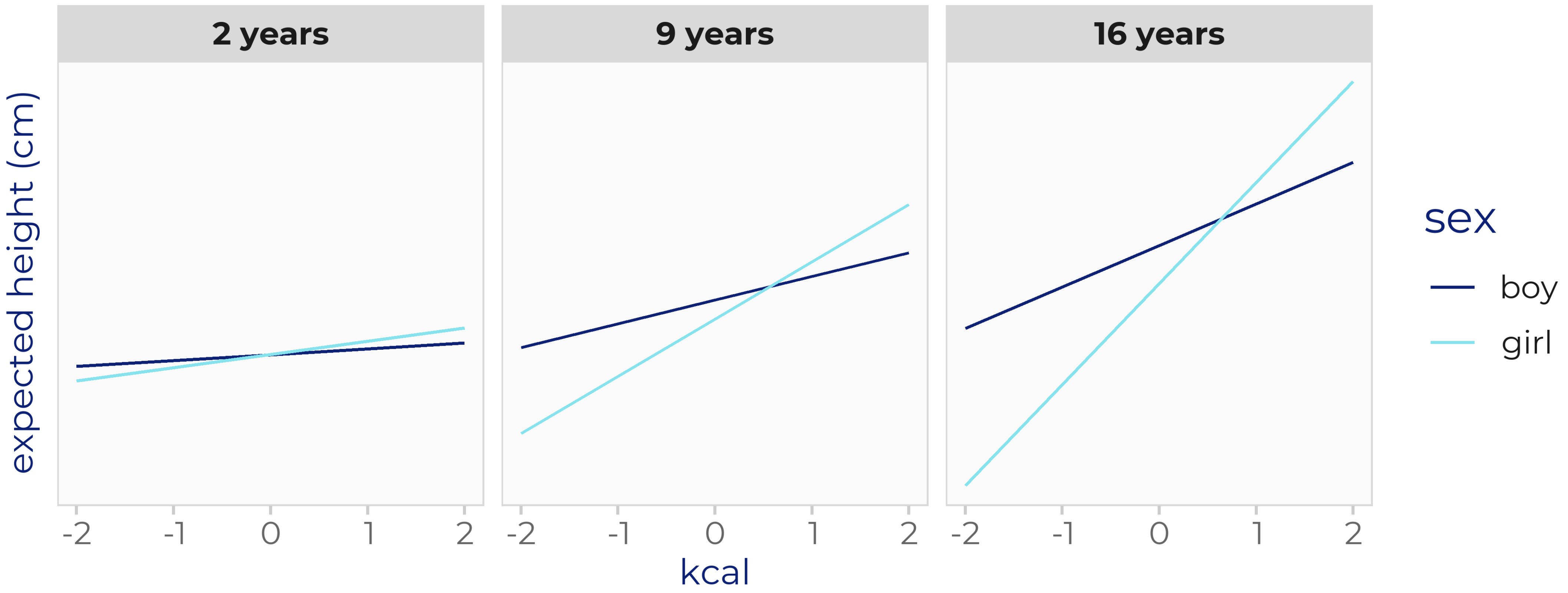
Interpretation of the 3-way interaction β_6:

- Change in the age-sex interaction when kcal is increased by 1 unit, or
- change in the age-kcal interaction when sex is 1 vs 0, or
- change in the effect of sex when kcal is increased by 1 unit and age $\neq$ 0, or
- change in the sex-kcal interaction when age is increased by 1 unit.

The 2-way interactions now have to be interpreted with respect to the value of the third variable.

Higher Order Interactions: Visualization

- Sex and age modify the effect of kcal.
- The interaction between sex and kcal is modified by age.



Interactions

Including interactions **increases the flexibility** of the model.

But:

- The **interpretation** of the parameters can get **very complicated** when
  - variables are involved in **multiple interaction terms**
  - and for **higher-level interactions**.
- We need **sufficient sample size** to get reasonably precise parameter estimates.
  - Interactions increase the number of parameters.
  - Interaction effects are usually smaller than the main effects.

Visualization via effect plots can facilitate the interpretation.

Remember:
The effects of variables involved in interaction terms cannot be interpreted independent of each other!

Interactions in R

Specification directly via the model `formula`:

- interaction: **`a:b`**
- interaction and main effects: **`a*b`** (same as **`a + b + a:b`**)

For example

```
fit1 <- lm(height ~ age * sex + kcal, data = child)
```

```
coef(fit1)
```

```
## (Intercept)         age     sexgirl        kcal age:sexgirl
##     47.3326      7.5928      3.5590     -0.3737     -1.6284
```

Interactions in R

Interaction with multiple variables:

```
fit2a <- lm(height ~ age * (sex + kcal), data = child)
```

is the same as

```
fit2b <- lm(height ~ age * sex + age * kcal, data = child)
```

```
coef(fit2a)
```

```
## (Intercept)         age     sexgirl        kcal age:sexgirl    age:kcal
##    47.29520     7.59695     3.56224    -0.08005    -1.62874    -0.03703
```

```
coef(fit2b)
```

```
## (Intercept)         age     sexgirl        kcal age:sexgirl    age:kcal
##    47.29520     7.59695     3.56224    -0.08005    -1.62874    -0.03703
```

Interactions in R

Higher order interactions can be specified using the **"^"** symbol:

```
fit3a <- lm(height ~ (age + sex + kcal)^3, data = child)
```

which is equivalent to

```
fit3b <- lm(height ~ age * sex * kcal, data = child)
```

```
coef(fit3a)
```

```
##       (Intercept)               age           sexgirl              kcal 
##          47.39104           7.58540           3.39474          -0.76393 
##       age:sexgirl          age:kcal      sexgirl:kcal age:sexgirl:kcal 
##          -1.60785           0.07809           1.25691          -0.22077
```

```
coef(fit3b)
```

```
##       (Intercept)               age           sexgirl              kcal 
##          47.39104           7.58540           3.39474          -0.76393 
##       age:sexgirl          age:kcal      sexgirl:kcal age:sexgirl:kcal 
##          -1.60785           0.07809           1.25691          -0.22077
```

Interactions in R

Remember: we can also remove terms from the model:

```
fit3c <- lm(height ~ (age + sex + kcal)^3 - sex:kcal, data = child)
coef(fit3c)
```

```
##       (Intercept)               age           sexgirl              kcal
##          47.30875           7.59390           3.53907          -0.06275
##       age:sexgirl          age:kcal  age:sexgirl:kcal
##          -1.62253           0.01287          -0.10274
```

Alternatively:

```
fit3d <- lm(height ~ age*(sex + kcal) + age:sex:kcal, data = child)
coef(fit3d)
```

```
##       (Intercept)               age           sexgirl              kcal
##          47.30875           7.59390           3.53907          -0.06275
##       age:sexgirl          age:kcal  age:sexgirl:kcal
##          -1.62253           0.01287          -0.10274
```