Lecture 5 | 31.03.2026

Linear regression model
with (linear) interactions

Overview: Multiple regression model

- Mathematical relationship between a continuous dependent variable Y and a set of explanatory (independent) variables $\boldsymbol{X} = (X_1, \dots, X_p)^\top$ (may be continuous, binary, or categorical – or an arbitrary combination)
- Typically expressed for some general function $f : \mathbb{R}^p \longrightarrow \mathbb{R}$ but for the linear regression model we use more specific notation of the form

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_{p-1} X_{p-1} + \varepsilon & &= \boldsymbol{X}^\top \boldsymbol{\beta} + \varepsilon \\ E[Y|\boldsymbol{X}] &= \beta_0 + \beta_1 X_1 + \beta_{p-1} X_{p-1} & &= \boldsymbol{X}^\top \boldsymbol{\beta} \end{aligned}$$

  where $f : \mathbb{R}^p \to \mathbb{R}$ is linear in the unknown parameters $\beta_1, \dots, \beta_p \in \mathbb{R}$
- The corresponding model for a random sample $\{(Y_i, \boldsymbol{X}_i);\ i = 1, \dots, n\}$ drawn from some joint distribution function $F_{(Y, \boldsymbol{X})}$ takes the form

$$Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta} + \varepsilon_i, \text{for } i = 1, \dots, n \text{or, alternatively } \boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

  where we assume (by default) the presence of the intercept parameter $\beta_0 \in \mathbb{R}$ in the model (i.e., $X_{i0} = 1$ almost surely)

Quantifying the effect of X on Y

- One of the main goals of the regression model (regression analysis in general) is to quantify the effect of some given explanatory variable on Y
- Formally, the explanatory variables in $\boldsymbol{X} \in \mathbb{R}^p$ typically have an effect on the whole (conditional) distribution of Y... however, for simplicity, we only focus on some distributional characteristics of this distribution...
- A typical characteristic related to the linear regression model is the conditional mean/expectation of Y given $\boldsymbol{X}$. Therefore, the effect of $\boldsymbol{X}$ on Y is also explained/interpreted in terms of the corresponding difference of the conditional expected values for two different values of $\boldsymbol{X}$
- The evaluation of the effect may be quantitative (in terms of the estimation of an unknown parameter) or it can be qualitative (in terms of a formal statement whether the effect is statistically important or not)

Interpretation: Association vs. causality

$\hookrightarrow$ the regression model is typically a model that explains only an association (relationship) between two (or more) subpopulations that differ with respect to the values of the explanatory covariates

Interpretation: Association vs. causality

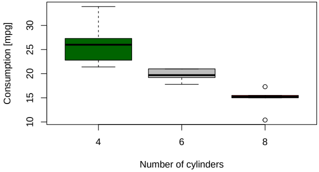
$\hookrightarrow$ the regression model is typically a model that explains only an association (relationship) between two (or more) subpopulations that differ with respect to the values of the explanatory covariates

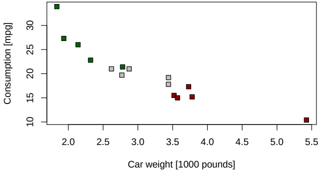
- **Associative interpretation**
  - Comparing two sub-populations that differ wrt to different values of $\boldsymbol{X}$
  - Interpreting the effect of $\boldsymbol{X}$ in terms of the comparison of two subjects
- **Causal interpretation**
  - Comparing the same sub-population before and after the change in $\boldsymbol{X}$
  - Interpreting the effect of $\boldsymbol{X}$ in terms of a change within the subject

$\hookrightarrow$ it is a common mistake that the associative regression model is (unintentionally) interpreted as a causal model... however, for a causal interpretation we usually need much stricter assumptions (a randomized trial for instance)

Example: Simple regression model variants

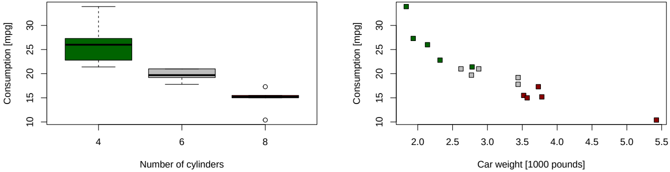
- Consumption of 15 US cars (given the number of cylinders and the car's weight) (5 cars with 4 cylinders, 5 cars with 6 cylinders and 5 cars with 8 cylinders)





Example: Simple regression model variants

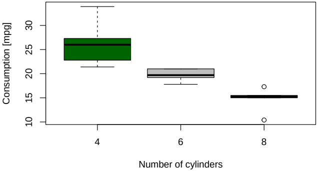
- Consumption of 15 US cars (given the number of cylinders and the car's weight) (5 cars with 4 cylinders, 5 cars with 6 cylinders and 5 cars with 8 cylinders)

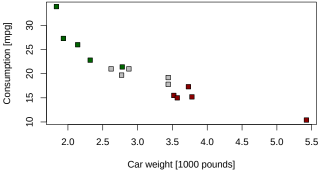



```
lm(formula = mpg ~ cyl + wt)

Estimate Std. Error t value Pr(>|t|)
(Intercept)    34.843      2.677  13.014 5.03e-08 ***
cyl6           -3.324      1.806  -1.841   0.0927 .
cyl8           -4.532      2.781  -1.630   0.1314
wt             -3.889      1.109  -3.508   0.0049 **
```

Example: Simple regression model variants

- Consumption of 15 US cars (given the number of cylinders and the car's weight) (5 cars with 4 cylinders, 5 cars with 6 cylinders and 5 cars with 8 cylinders)

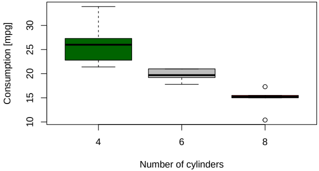


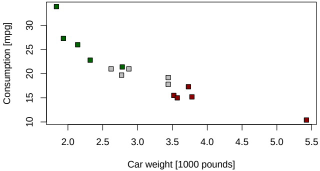



```
lm(formula = mpg ~ cyl + wt)

Estimate Std. Error t value Pr(>|t|)
(Intercept)    34.843       2.677  13.014 5.03e-08 ***
cyl6           -3.324       1.806  -1.841   0.0927 .
cyl8           -4.532       2.781  -1.630   0.1314
wt             -3.889       1.109  -3.508   0.0049 **


lm(formula = mpg ~ cyl + I(wt - 3), data = mtcars2)

Estimate Std. Error t value Pr(>|t|)
(Intercept)    23.177       1.412  16.419 4.39e-09 ***
cyl6           -3.324       1.806  -1.841   0.0927 .
cyl8           -4.532       2.781  -1.630   0.1314
I(wt - 3)      -3.889       1.109  -3.508   0.0049 **
```

Example: Simple regression model variants

- Consumption of 15 US cars (given the number of cylinders and the car's weight) (5 cars with 4 cylinders, 5 cars with 6 cylinders and 5 cars with 8 cylinders)





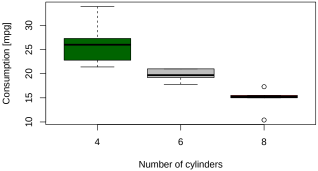

```
lm(formula = mpg ~ cyl + wt)

Estimate Std. Error t value Pr(>|t|)
(Intercept)   34.843      2.677  13.014 5.03e-08 ***
cyl6          -3.324      1.806  -1.841   0.0927 .
cyl8          -4.532      2.781  -1.630   0.1314
wt            -3.889      1.109  -3.508   0.0049 **
```

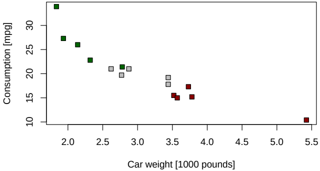
```
lm(formula = mpg ~ cyl + wt,  contrasts = list(cyl = contr.sum))

Estimate Std. Error t value Pr(>|t|)
(Intercept)  32.2238     3.5719   9.021 2.05e-06 ***
cyl1          2.6189     1.4001   1.870   0.0882 .
cyl2         -0.7053     0.9119  -0.773   0.4556
wt           -3.8886     1.1085  -3.508   0.0049 **
```

```
lm(formula = mpg ~ cyl + I(wt - 3), data = mtcars2)

Estimate Std. Error t value Pr(>|t|)
(Intercept)   23.177      1.412  16.419 4.39e-09 ***
cyl6          -3.324      1.806  -1.841   0.0927 .
cyl8          -4.532      2.781  -1.630   0.1314
I(wt - 3)     -3.889      1.109  -3.508   0.0049 **
```

Example: Simple regression model variants

- Consumption of 15 US cars (given the number of cylinders and the car's weight) (5 cars with 4 cylinders, 5 cars with 6 cylinders and 5 cars with 8 cylinders)






```
lm(formula = mpg ~ cyl + wt)

Estimate Std. Error t value Pr(>|t|)
(Intercept)    34.843      2.677  13.014 5.03e-08 ***
cyl6           -3.324      1.806  -1.841   0.0927 .
cyl8           -4.532      2.781  -1.630   0.1314
wt             -3.889      1.109  -3.508   0.0049 **
```

```
lm(formula = mpg ~ cyl + wt,  contrasts = list(cyl = contr.sum))

Estimate Std. Error t value Pr(>|t|)
(Intercept)  32.2238     3.5719   9.021 2.05e-06 ***
cyl1          2.6189     1.4001   1.870   0.0882 .
cyl2         -0.7053     0.9119  -0.773   0.4556
wt           -3.8886     1.1085  -3.508   0.0049 **
```

```
lm(formula = mpg ~ cyl + I(wt - 3), data = mtcars2)

Estimate Std. Error t value Pr(>|t|)
(Intercept)    23.177      1.412  16.419 4.39e-09 ***
cyl6           -3.324      1.806  -1.841   0.0927 .
cyl8           -4.532      2.781  -1.630   0.1314
I(wt - 3)      -3.889      1.109  -3.508   0.0049 **
```

```
lm(formula = mpg ~ cyl + I(wt-3), contrasts = list(cyl = contr.sum)

Estimate Std. Error t value Pr(>|t|)
(Intercept)  20.5580     0.6628  31.018 4.64e-12 ***
cyl1          2.6189     1.4001   1.870   0.0882 .
cyl2         -0.7053     0.9119  -0.773   0.4556
I(wt - 3)    -3.8886     1.1085  -3.508   0.0049 **
```

Correlation among explanatory variables

Ideal scenario

- balanced data
- uncorrelated predictors
- each coeffcient β_j can be estimated separately
- interpretation of the estimated coefficients is relatively stable

Typical real-data situations

- unbalanced data
- correlated predictor variables (multicolinearity)
- variance of the estimated parameters typically increases
- the interpretation of the estimated coefficients becomes rather vague

$\hookrightarrow$ in general, the value of the parameter β_j stands for a difference in the expected (conditional) value of Y which comes with a unit difference within the X_j covariate while all other predictors are kept fixed.
In practice, however, predictor variables typically change simultaneously and also, moreover, they even affect one another...

Example: Body fat vs. weight and height

- **Body fat vs. person's height** (Model 1)

```
lm(formula = fat ~ height, data = Policie)

Coefficients:
             Estimate  Std.Error t value Pr(>|t|)
(Intercept) -47.6791    23.9707  -1.989   0.0524 .
height        0.3405     0.1343   2.535   0.0146 *
```

- **Body fat vs. person's weight** (Model 2)

```
lm(formula = fat ~ weight, data = Policie)

Coefficients:
              Estimate  Std.Error t value Pr(>|t|)
(Intercept) -20.75217    3.42327  -6.062 2.02e-07 ***
weight        0.42674    0.04266  10.003 2.51e-13 ***
```

What about a multiple model?

- **Body fat vs. person's height and weight** (Model 3)

```
lm(formula = fat ~ height + weight, data = Policie)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.55309   15.24621   1.086   0.2831
height      -0.24362    0.09728  -2.504   0.0158 *
weight       0.50418    0.05095   9.896 4.49e-13 ***
```

What about a multiple model?

- **Body fat vs. person's height and weight** (Model 3)

```
lm(formula = fat ~ height + weight, data = Policie)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.55309   15.24621   1.086   0.2831
height      -0.24362    0.09728  -2.504   0.0158 *
weight       0.50418    0.05095   9.896 4.49e-13 ***
```

- **What is the estimated effect of the height on the overall body fat?**
- **What is the estimated effect of the weight on the overall body fat?**
- **How well the conclusions correspond among different models?**

What about a multiple model?

- **Body fat vs. person's height and weight** (Model 3)

```
lm(formula = fat ~ height + weight, data = Policie)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.55309   15.24621   1.086   0.2831
height      -0.24362    0.09728  -2.504   0.0158 *
weight       0.50418    0.05095   9.896 4.49e-13 ***
```

- **What is the estimated effect of the height on the overall body fat?**
- **What is the estimated effect of the weight on the overall body fat?**
- **How well the conclusions correspond among different models?**
- The estimated correlation between the weight and height is 0.6068!

How to overcome the problems?

- **Body fat vs. person's height and weight with interactions** (Model 4)

```
lm(formula = fat ~ height + weight + height:weight)

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   -48.604790  87.698149  -0.554    0.582
height          0.123659   0.496447   0.249    0.804
weight          1.324727   1.088637   1.217    0.230
height:weight  -0.004608   0.006106  -0.755    0.454
```

How to overcome the problems?

- **Body fat vs. person's height and weight with interactions** (Model 4)

```
lm(formula = fat ~ height + weight + height:weight)

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -48.604790  87.698149  -0.554    0.582
height         0.123659   0.496447   0.249    0.804
weight         1.324727   1.088637   1.217    0.230
height:weight -0.004608   0.006106  -0.755    0.454
```

- **What is the interaction term? How to explain it?**
- **What are the main pros/cons of the model with interactions?**

How to overcome the problems?

- **Body fat vs. person's height and weight with interactions** (Model 4)

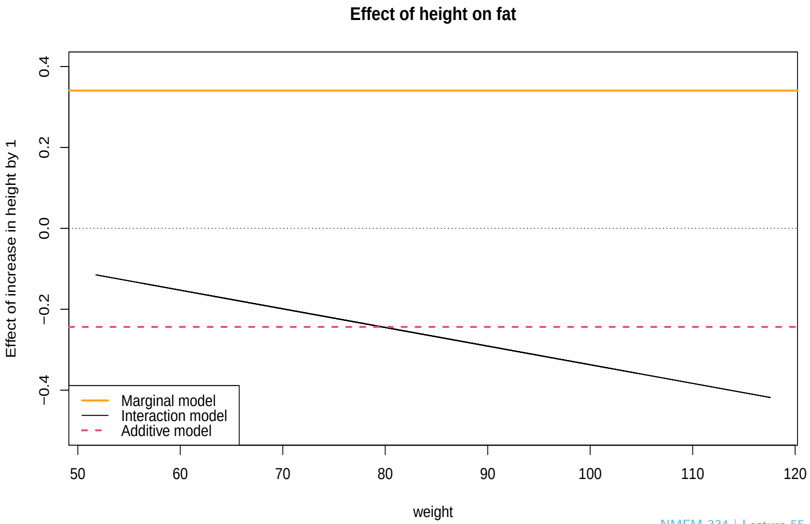
```
lm(formula = fat ~ height + weight + height:weight)

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   -48.604790  87.698149  -0.554    0.582
height          0.123659   0.496447   0.249    0.804
weight          1.324727   1.088637   1.217    0.230
height:weight  -0.004608   0.006106  -0.755    0.454
```

- **What is the interaction term? How to explain it?**
- **What are the main pros/cons of the model with interactions?**

Note, that Model 1 and Model 2 are principially different—they are using different set of covariates to explain the conditional expectation of `fat`. On the other hand, Model 3 and Model 4 are also different but both are using the same information provided by the same set of covariates. Model 3 is actually a **submodel** of Model 4.

Moreover, both Model 1 and Model 2 are also submodels of Model 3 or Model 4.

Illustration of the Models 1, 3, and 4



Regression model with interactions: Formally

- **Implementation in the R software**
  - using the expression `height:weight` (interactions only)
  - using the expression `height * weight` (hierarchically well structured)
  - defining new covariate as a product of `height` and `weight`

- **Formulation within a linear regression model**
  - using a regression model expression: $Y \approx \beta_0 + \beta_1 X_h + \beta_2 X_w + \beta_3 X_h X_w$
  - using a new covariate $Y \approx \beta_0 + \beta_1 X_h + \beta_2 X_w + \beta_3 Z$ where $Z = X_h \times X_w$

- **More general formulations and models**
  - effect of height: $Y \approx \beta_0 + (\beta_1 + \beta_3 X_w) X_h + \beta_2 X_w$
  - effect of weight: $Y \approx \beta_0 + (\beta_2 + \beta_3 X_h) X_w + \beta_3 X_h$

Regression model with interactions: Formally

- **Implementation in the R software**
  - using the expression `height:weight` (interactions only)
  - using the expression `height * weight` (hierarchically well structured)
  - defining new covariate as a product of `height` and `weight`

- **Formulation within a linear regression model**
  - using a regression model expression: $Y \approx \beta_0 + \beta_1 X_h + \beta_2 X_w + \beta_3 X_h X_w$
  - using a new covariate $Y \approx \beta_0 + \beta_1 X_h + \beta_2 X_w + \beta_3 Z$ where $Z = X_h \times X_w$

- **More general formulations and models**
  - effect of height: $Y \approx \beta_0 + (\beta_1 + \beta_3 X_w) X_h + \beta_2 X_w$
  - effect of weight: $Y \approx \beta_0 + (\beta_2 + \beta_3 X_h) X_w + \beta_3 X_h$

- **Thus**
  - parameter β_3 can be seen as a linear function of X_w (or X_h respectively)
  - more generaly, β_3 is a function of X_w (or X_h respectively)
  - it can be written as $\beta_3(X_w)$ (or $\beta_3(X_h)$ respectively), where $\beta_3(x) = cx$

Regression model with interactions: Formally

- **Implementation in the R software**
  - using the expression `height:weight` (interactions only)
  - using the expression `height * weight` (hierarchically well structured)
  - defining new covariate as a product of `height` and `weight`

- **Formulation within a linear regression model**
  - using a regression model expression: $Y \approx \beta_0 + \beta_1 X_h + \beta_2 X_w + \beta_3 X_h X_w$
  - using a new covariate $Y \approx \beta_0 + \beta_1 X_h + \beta_2 X_w + \beta_3 Z$ where $Z = X_h \times X_w$

- **More general formulations and models**
  - effect of height: $Y \approx \beta_0 + (\beta_1 + \beta_3 X_w) X_h + \beta_2 X_w$
  - effect of weight: $Y \approx \beta_0 + (\beta_2 + \beta_3 X_h) X_w + \beta_3 X_h$

- **Thus**
  - parameter β_3 can be seen as a linear function of X_w (or X_h respectively)
  - more generaly, β_3 is a function of X_w (or X_h respectively)
  - it can be written as $\beta_3(X_w)$ (or $\beta_3(X_h)$ respectively), where $\beta_3(x) = cx$
  - Is it necessary to stay within the linearity constraints here?
    What if $\beta_3(x) = g(x)$ for some general (smooth/continuous) function g?

Regression model with interactions: Formally

- **Implementation in the R software**
  - using the expression `height:weight` (interactions only)
  - using the expression `height * weight` (hierarchically well structured)
  - defining new covariate as a product of `height` and `weight`

- **Formulation within a linear regression model**
  - using a regression model expression: $Y \approx \beta_0 + \beta_1 X_h + \beta_2 X_w + \beta_3 X_h X_w$
  - using a new covariate $Y \approx \beta_0 + \beta_1 X_h + \beta_2 X_w + \beta_3 Z$ where $Z = X_h \times X_w$

- **More general formulations and models**
  - effect of height: $Y \approx \beta_0 + (\beta_1 + \beta_3 X_w) X_h + \beta_2 X_w$
  - effect of weight: $Y \approx \beta_0 + (\beta_2 + \beta_3 X_h) X_w + \beta_3 X_h$

- **Thus**
  - parameter β_3 can be seen as a linear function of X_w (or X_h respectively)
  - more generaly, β_3 is a function of X_w (or X_h respectively)
  - it can be written as $\beta_3(X_w)$ (or $\beta_3(X_h)$ respectively), where $\beta_3(x) = cx$
  - Is it necessary to stay within the linearity constraints here?
    What if $\beta_3(x) = g(x)$ for some general (smooth/continuous) function g?

When we investigate the effect of the height on the overall amount of fat the other covariate (weight) acts as an **effect modifier** in the underlying model (and vise versa)

When to use a model with interactions?

- **Effect modifier**
  If there is a belief that the effect of one specific covariate $X \in \mathbb{R}$ will be different in different sub-populations that we want/need to control for in the model by using the remaining covariates

- **Colinearity issues**
  If the model design is not optimal and it is possible that some covariates may be substantially correlated (i.e., linearly dependent – multicolinearity) then the interaction(s) may help to improve the model

- **Model interpretability**
  Interactions can be also used for the purpose of some better model interpretation (despite the fact that typically the interactions make the model interpretability more complex/challenging)
  (typically in situations with purely categorical covariates for instance)

When to use a model with interactions?

- **Effect modifier**
  If there is a belief that the effect of one specific covariate $X \in \mathbb{R}$ will be different in different sub-populations that we want/need to control for in the model by using the remaining covariates

- **Colinearity issues**
  If the model design is not optimal and it is possible that some covariates may be substantially correlated (i.e., linearly dependent – multicolinearity) then the interaction(s) may help to improve the model

- **Model interpretability**
  Interactions can be also used for the purpose of some better model interpretation (despite the fact that typically the interactions make the model interpretability more complex/challenging)
  (typically in situations with purely categorical covariates for instance)

Interactions are not necessarily between just two explanatory covariates (so-called double interactions, or first-order interactions). In practice, higher-order interactions between three and even more covariates can be implemented in the model
(but the model interpretation becomes very challenging, almost impossible...)

Interpretation of the interaction term

- Consider a simple regression model with one interaction

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2) + \varepsilon$$

- We are primariliy interested in the effect of X_1 on $E[Y|X_1, X_2]$ thus, we can rewrite the model in the equivalent form

$$Y = \beta_0 + (\beta_1 + \beta_3 X_2) X_1 + \beta_2 X_2 + \varepsilon$$

- To describe the effect of X_1 on $E[Y|X_1, X_2]$ we need to estimate/predict $(\beta_1 + \beta_3 X_2)$ which, however, depends on a (random) value of X_2 – taking (hypotetically) infinitelly many values   Which ones to use?
- For a given $X_2 = x_2$, the effect of X_1 on $E[Y|X_1, X_2]$ reduces to $\beta_1 + \beta_3 x_2$ which can be easily estimated
- For $X_2 \in \{0, 1\}$ (binary), the effect of X_1 on $E[Y|X_1, X_2]$ reduces to two disctinct effects in two separate grous defined by X_2

Towards model building process...

- **Nonlinear transformations**
  Many different transformation functions $t \in \mathcal{G}$ can be considered within the regression model

$$Y = \beta_0 + \beta_1 t_1(X_1) + \beta_2 t_2(X_2) + \varepsilon$$

  to improve its flexibility but different transformations (different choices of $t_1, t_2 \in \mathcal{G}$) change the overall model (its properties, interpretation, etc.) and the models are not directly comparable among each other

- **Linear transformations**
  There is a specific class of transformations that preserve most of the model qualitites (if the model is well-formulated) and these are linear transformations of the form $t(x) = a + bx$, i.e.,

$$Y = \beta_0 + \beta_1(a_1 + b_1 X_1) + \beta_2(a_2 + b_2 X_2) + \varepsilon$$

  for $a_1, a_2, b_1, b_2 \in \mathbb{R}$ – models under such transformations are equivalent (if $b_1 \neq 0 \neq b_2$) and can be directly compared among each other...

Linear transformations of the covariates

Typically they are used to

- to improve the stability of the estimated parameters
  (e.g., measuring the distance between Prague and Brno in millimeters/kilometers)
- for better (visual) representation of the model outputs
  (mostly using different units, scales, proportions for better visualization)
- to improve the interpretation of the final model
  (typically, we want to have a reasonable interpretation of the intercept and interactions)

Linear transformations of the covariates

Typically they are used to

- to improve the stability of the estimated parameters
  (e.g., measuring the distance between Prague and Brno in millimeters/kilometers)
- for better (visual) representation of the model outputs
  (mostly using different units, scales, proportions for better visualization)
- to improve the interpretation of the final model
  (typically, we want to have a reasonable interpretation of the intercept and interactions)

However, it only works with a **hierarchically well-formulated** model...

- What is a hierarchically well-formualted/structured model?
- What are the consequences of a non-hierarchical model?

Model hierarchy

A **hierarchically well-formulated linear regression model** (often called a hierarchical or marginality-respecting model) follows a simple but important principle:
If a higher-order term is included (e.g., interactions or polynomials), all corresponding lower-order terms must also be included in the model.

- **Advantages**
  - linear transformations of the covariates does not effect the model
  - different models are better compared within their hierarchical structure
  - systematic model building procedures are well defined and work well

- **Disadvantages**
  - some models can not be fitted under the restriction of hierarchy
  - models with various irregularities (discontinuous, non-smooth
  - sometimes it is necessary to use a model without the intercept

Model hierarchy

A **hierarchically well-formulated linear regression model** (often called a hierarchical or marginality-respecting model) follows a simple but important principle:
If a higher-order term is included (e.g., interactions or polynomials), all corresponding lower-order terms must also be included in the model.

- **Advantages**
  - linear transformations of the covariates does not effect the model
  - different models are better compared within their hierarchical structure
  - systematic model building procedures are well defined and work well

- **Disadvantages**
  - some models can not be fitted under the restriction of hierarchy
  - models with various irregularities (discontinuous, non-smooth
  - sometimes it is necessary to use a model without the intercept

↪ when fitting a linear regression model, we always need to be aware of its structure – whether we are building a model that is hierarchically well formulated or not... and depending on the model we have different tools available for the fitting process and the consecutive inference as well

Summary

- **Models with interactions**
  - they may help to overcome some issues with the covariates
  - they may improve the overall flexibility of the model
  - interpretation of the model usually becomes more challenging
- **Linear transformations of the covariates**
  - they help with the model/computational stability
  - when used wisely, they improve the interpretation of the final model
  - they require a hierarchically well formulated model to work properly
- **Hierarchically well-formulated (HWF) models**
  - have some specific advantages and disadvantages
  - inference in a hierarchical model is more straightfoward
  - however, some practical applications may require a non-hierarchical model
- **Key message**
  - some models are in a mutual relationship (i.e., model vs. submodel)
  - a sequence of submodels is the main tool for the model building purposes
  - for a proper comparisons of the submodels, we need HWF models