Lecture 6 | 24.03.2025

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 $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$ (may be continuous, binary, or categorical – or any combination)
- Typically expressed for some general function $f : \mathbb{R}^p \longrightarrow \mathbb{R}$ but for the linear regression model we use a more specific notation of the form

$$Y = \beta_0 + \beta_1 X_1 + \beta_{p-1} X_{p-1} + \varepsilon \qquad = \boldsymbol{X}^\top \boldsymbol{\beta} + \varepsilon$$

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

- The corresponding model for a random sample $\{(Y_i, \boldsymbol{X}_i);\ i = 1, \ldots, n\}$ drawn from some joint distribution function $F_{(Y,\boldsymbol{X})}$ takes the form

$$Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta} + \varepsilon_i$$

for $i = 1, \ldots, n$ 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 variable may have an effect on the whole (conditional) distribution of Y... however, for simplicity, we only focus on some distributional characteristics instead
- Typical characteristic related to the linear regression model is the conditional mean of Y given X. Therefore, the effect of X on Y is also explained/interpreted in terms of the corresponding change of the conditional expected value when the value of X changes
- 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 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 value of the explanatory covariate(s)

Interpretation: Association vs. causality

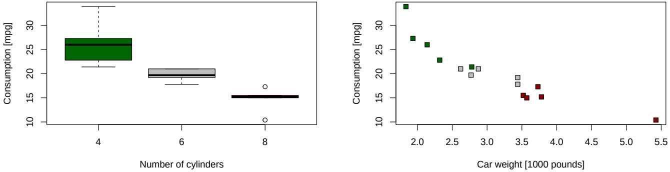
↪ the regression model is typically a model that explains only an association (relationship) between two (or more) subpopulations that differ with respect to the value of the explanatory covariate(s)

- **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

↪ it is a very 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)

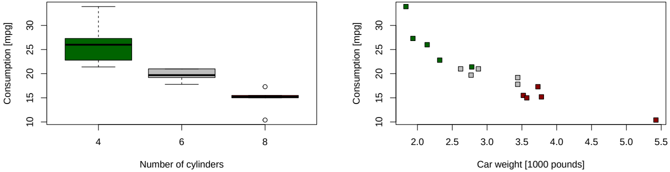
Example: Simple regression model variants

- 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

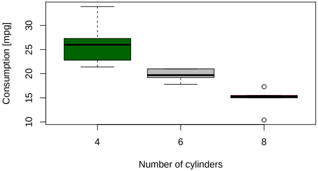
- Consumption of 15 US cars (given the number of cylinders and the car's weight) (5 cars with 4 cylinders, 5 cars with 6 cylinders and 5 cars with 8 cylinders)

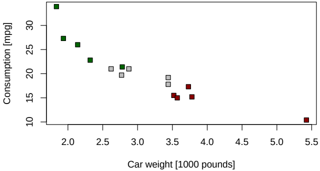



```
lm(formula = mpg ~ cyl + wt)

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)

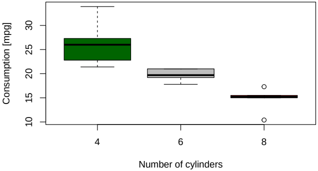


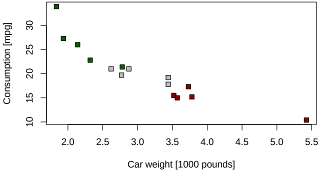



```
lm(formula = mpg ~ cyl + wt)

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)





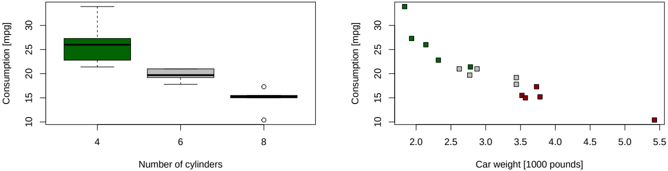

```
lm(formula = mpg ~ cyl + wt)

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 fixed

- **Typical real situations**
  - unbalanced data
  - correlated predictor variables (multicolinearity)
  - variance of the estimated parameters typically increases
  - the interpretation of the estimated coefficients become vague

$\hookrightarrow$ briefly saying, the estimated parameter β_j stands for a change in the expected (conditional) value of Y which comes with a unit change of X_j covariate, however, with all other predictors being fixed. In practice, the predictor variables typically change simultaneously. variables

Example: Body fat vs. weight and height

- **Body fat vs. person's height**

```
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**

```
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**

```
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

```
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**

```
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? Interactions!

- **Body fat vs. person's height and weight with the interaction**

```
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
```

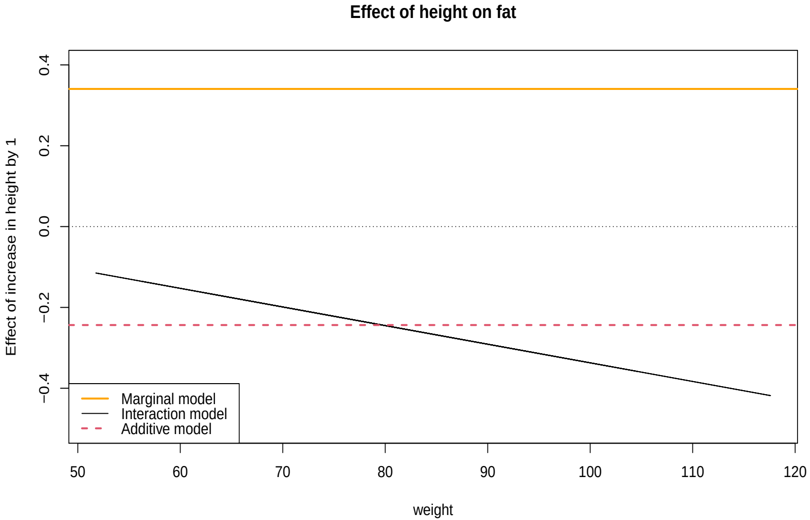
How to overcome the problems? Interactions!

- **Body fat vs. person's height and weight with the interaction**

```
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?**
- **Is the model good one?**
- **What are the main advantages and disadvantages of the model with interactions?**

Illustration of the models



Regression model with interactions: Formally

- **Implementation in the R software**
  - using the expression `height:weight`
  - using the expression `height * weight`
  - 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`
  - using the expression `height * weight`
  - 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: $Y \approx \beta_0 + (\beta_2 + \beta_3 X_h) X_w + \beta_3 X_h$

  - 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)
  - thus, we can write $\beta_3(X_w)$ (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`
  - using the expression `height * weight`
  - 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: $Y \approx \beta_0 + (\beta_2 + \beta_3 X_h) X_w + \beta_3 X_h$

  - 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)
  - thus, we can write $\beta_3(X_w)$ (or $\beta_3(X_h)$ respectively), where $\beta_3 x = cx$
  - so, is it necessary to stay with the linearity restrictions? What if $\beta(x) = g(x)$ for some general function g?

Regression model with interactions: Formally

Implementation in the R software

- using the expression `height:weight`
- using the expression `height * weight`
- 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: $Y \approx \beta_0 + (\beta_2 + \beta_3 X_h) X_w + \beta_3 X_h$

- 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)
- thus, we can write $\beta_3(X_w)$ (or $\beta_3(X_h)$ respectively), where $\beta_3 x = cx$
- so, is it necessary to stay with the linearity restrictions? What if $\beta(x) = g(x)$ for some general function g?

$\hookrightarrow$ Thus, when being interested in the effect of height on the overall amount of fat, the other covariate (weight) acts as a **effect modifier** in the model (and vise versa)

When to use a model with interactions?

- **Effect modifier**
  When there is an expectation that the effect of one specific covariate X_j 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 there is a belief that some covariates may be substantially correlated (linearly dependent – multicolinearity) then the interaction(s) may help to improve the model
- **Model interpretability** Interactions can be also used just for the purpose of some better model interpretability (despite the fact that mostly interactions make the model interpretability more complex/challenging)

When to use a model with interactions?

- **Effect modifier**
  When there is an expectation that the effect of one specific covariate X_j 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 there is a belief that some covariates may be substantially correlated (linearly dependent – multicolinearity) then the interaction(s) may help to improve the model
- **Model interpretability** Interactions can be also used just for the purpose of some better model interpretability (despite the fact that mostly interactions make the model interpretability more complex/challenging)

Interactions are not necessarily just between to explanatory covariates (so-called double interactions, or first-order interactions). In practice, we can technically use even higher-order interactions between three and more covariates – but they exponentially complicates the final model interpretability

Simple 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 quantify/estimate $(\beta_1 + \beta_3 X_2)$ which, however, depends on the value of X_2 – taking (hypotetically) infinitelly many values Which ones to use?
- For $X_2 = 2$ the effect of X_1 on $E[Y|X_1, X_2]$ only reduces to the quantification/estimation of β_1 Can we somehow achieve this?

Transformations of the covariates

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

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

  but different transformations (different choice 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**
  a very specific class of transformations that preserve most of the model qualitites are 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 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 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 structured model.

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

Model hierarchy

- **Advantages**
  - linear transformations of the covariates does not effect the model
  - different models are better comparable 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

- **Advantages**
  - linear transformations of the covariates does not effect the model
  - different models are better comparable 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

$\hookrightarrow$ 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**
  - the yhelp to overcome some issues with the covariates
  - the improve the overall flexibility of the model
  - interpretation of the model becomes more challenging
- **Linear transformations of the covariates**
  - they help with the model stability
  - when used wisely, they improve the interpretability of the model
  - they require a hierarchically well formulated model to work properly
- **Hierarchically well formulated model**
  - it has its specific advantages and disadvantages
  - inference in a hierarchical model is more straightfoward
  - some practical applications require a non-hierarchical model