

R workshop #2: multivariate regression analysis and factor interactions

Nicola Romandò

Introduction

Last year we have talked about linear models and their use to perform regression and analysis of variance (ANOVA). We only considered simple situations where we have one independent variable influencing our measured variable (one-way ANOVA) or two factors (two-ways ANOVA) that do not interact with each other. In the lectures we have now talked about interactions and how they change our interpretation of linear models. In this workshop we will have a look at how to deal with interactions in R.

Learning objectives

After completing this workshop you will be able to:

- Use linear models to perform multiple regression
- Use linear models to perform Analysis of Variance with multiple factors
- Correctly interpret the results of your analysis in the presence of interactions

Section 1 - A refresher on linear models

We start this workshop with a little refresher of linear models. A linear model is a statistical model that relates the changes in a dependent variable (Y) with the changes in one or more independent variables (X_1, X_2, \dots, X_n).

The general equation for such model is:

$$Y = X_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

Where:

- Y is our measured variables
- X_1, \dots, X_n are the factors (or predictors) that influence Y
- β_1, \dots, β_n are scaling coefficients that indicate the importance of each predictor.
- ϵ is the error, or residual. It represents the effect of all the factors that we did not measure in our experimental setup.

In this simple formula each predictor acts independently from the others. In other words, if we have two predictors, X_1 and X_2 , the effect of X_1 on Y will always be the same, independently of the value of X_2 . As we have seen in the lecture this is not always the case.

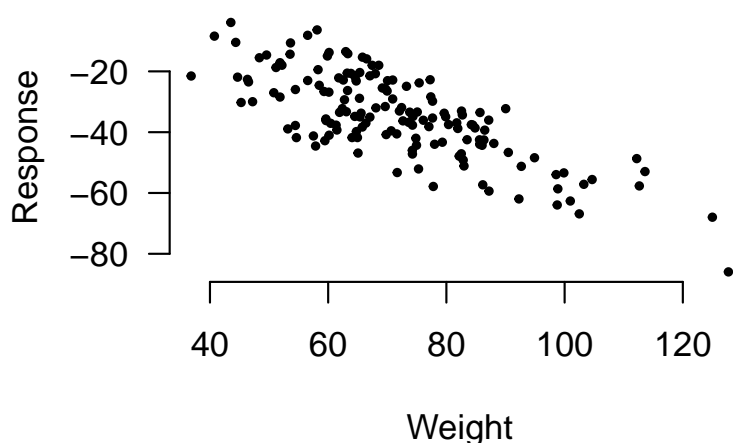
Simple regression

As a first example let's consider the dataset *pressure-workshop2.csv*. In this study the effect of a drug on reducing blood pressure has been investigated on 150 patients of different age, weight, and sex.

Start by familiarising with the data. How many men and women are there? What range of age and weight? Try and plot the various variables against each other and see if any particular patterns emerge.¹

Let's forget for a moment about the other variables and concentrate on the relation between Weight and response; it looks like the largest effect is seen in heavier patients.

¹ If you do not remember how to do that see Workshop 1.



We can use linear regression to test whether such a relation exists. Let's state our null hypothesis _____

Do you remember how to perform a linear regression in R? Try it, if you don't remember see the following page!

```
model <- lm(Response ~ Weight, pressure)
```

This generates the model

$$\text{Response} = \beta_0 + \beta_1 * \text{Weight} + \epsilon$$

What are the assumption of this model? Do you remember how to verify that they are satisfied? ²

Let's look at the output of the model³, what does it tell you?

² Let's discuss this in the forum!

³ I would say that the assumptions are generally satisfied, what do you think?

```
summary(model)

##
## Call:
## lm(formula = Response ~ Weight, data = pressure)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.190  -6.354   0.874   6.568  19.554
##
## Coefficients:
##              Estimate Std. Error t value
## (Intercept) 11.78143    3.32703   3.541
## Weight      -0.64852    0.04521 -14.346
##              Pr(>|t|)
## (Intercept) 0.000533 ***
## Weight      < 2e-16 ***
## ---
## Signif. codes:
##  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.133 on 148 degrees of freedom
## Multiple R-squared:  0.5817, Adjusted R-squared:  0.5789
## F-statistic: 205.8 on 1 and 148 DF,  p-value: < 2.2e-16
```

From the summary we get:

$$\beta_0 = 11.78 \text{ and } \beta_1 = -0.65$$

Therefore

$$\text{Response} = 11.78 - 0.65 * \text{Weight} + \epsilon$$

Simple regression (i.e. linear regression with a single predictor) is very simple to interpret. For any increase of 1 Kg in weight there is a decrease of 0.65 mmHg in blood pressure.