Session regression II: multiple linear regression

Learning outcomes

After this session, a student should be able to:

- visualize bivariate relationships in small datasets
- fit a linear regression model containing main and interaction effects
- assess the quality of the model fit
- determine if at least one predictor can explain the response
- determine which predictors explain the response
- asses the accuracy of predictions from the model

Visualizing the Advertising dataset

In this session we will use the Advertising dataset. This simple dataset consists of sales data for 200 products along with the amount of money spent on TV, radio, and newspaper ads. We would like to know how best to spend advertizing money to maximize sales.

To begin with, let us familiarize ourselves with the dataset. The data are stored in the `data` subdirectory of this session.

```
# load the data
ads = read.csv('./data/Advertising.csv')
```

First, we check to see what columns were imported

```
head(ads)
```

```
##   X    TV radio newspaper sales
## 1 1 230.1  37.8      69.2  22.1
## 2 2  44.5  39.3      45.1  10.4
## 3 3  17.2  45.9      69.3   9.3
## 4 4 151.5  41.3      58.5  18.5
## 5 5 180.8  10.8      58.4  12.9
## 6 6   8.7  48.9      75.0   7.2
```

It looks like a redundant column, `X`, has made it into the table. However, the other columns look like numbers. That's good.

Now, let us use the `pairs` function to get a quick overview of linear relationships within the dataset.

```
# visualize all pairwise relationships
pairs(ads)
```

Estimating the Regression Coefficients

Estimating coefficients

Relationship between the response and predictors

Model fit

Predictions

Qualitative predictors

Interaction terms

Non-linear transformation of the predictors

Potential problems: non-linearity, collinearity

Logistic regression