Practical 1: Review of the Normal Linear Model

The data we will use in this practical is the **mtcars** file contained within the R package. The data were extracted from the 1974 _Motor Trend_ US magazine and comprise fuel consumption and 10 aspects of car design for 32 cars (1973-74 models).

1. Attach the data file: `attach(mtcars)`
   View the data: `mtcars`

   The variable/column names and meanings are:

mpg:	**Miles/(US) gallon**
cyl:	**Number of cylinders (either 4, 6 or 8)**
disp:	Displacement (cu.in.)
hp:	Gross horsepower
drat:	Rear axle ratio
wt:	**Weight (lb/1000)**
qsec:	1/4 mile time
vs:	V/S
am:	Transmission (0 = automatic, 1 = manual)
gear:	Number of forward gears
carb:	Number of carburettors

2. We are going to fit a linear model to **m.p.g.** with **number of cylinders** (cyl) as a factor and **weight** (wt) as a covariate.

   We will formulate the model such that the cyl value of 4 (the lowest value) is the reference. We will need 2 dummy variables to represent cyl.

   $$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + e_i , \; e_i \sim NID(0, \sigma^2)$$

   Y = mpg
   X_1 = 1 if cyl = 6, 0 otherwise
   X_2 = 1 if cyl = 8, 0 otherwise

X_3 = wt
i = 1 to 32

3. Fit the linear model in R.

4. Interpret the parameter estimates for cyl:

5. Interpret the parameter estimate for wt:

6. We can view the model matrix constructed by R:

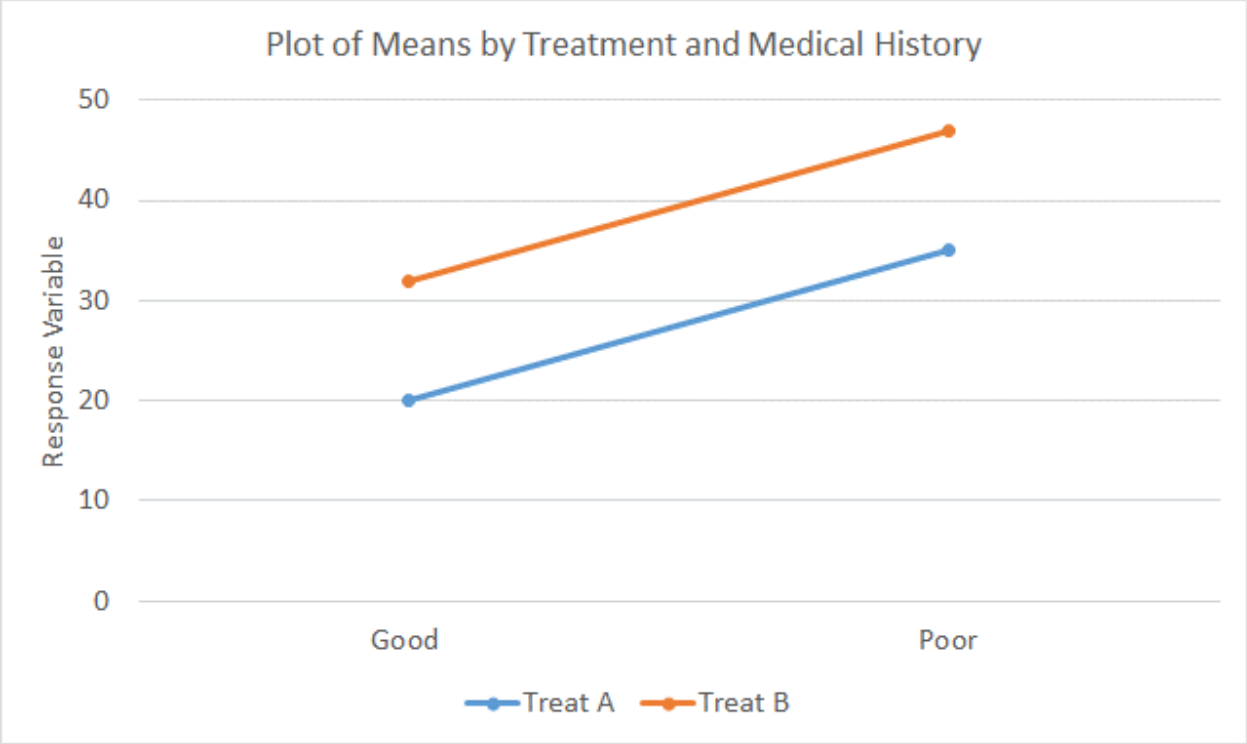
```
model.matrix(lm1)
```

7. In the above analyses, we treat cyl as a factor. Let us see what would happen if we treated cyl as a covariate. What is the model we are fitting? Interpret the parameter estimate for cyl.

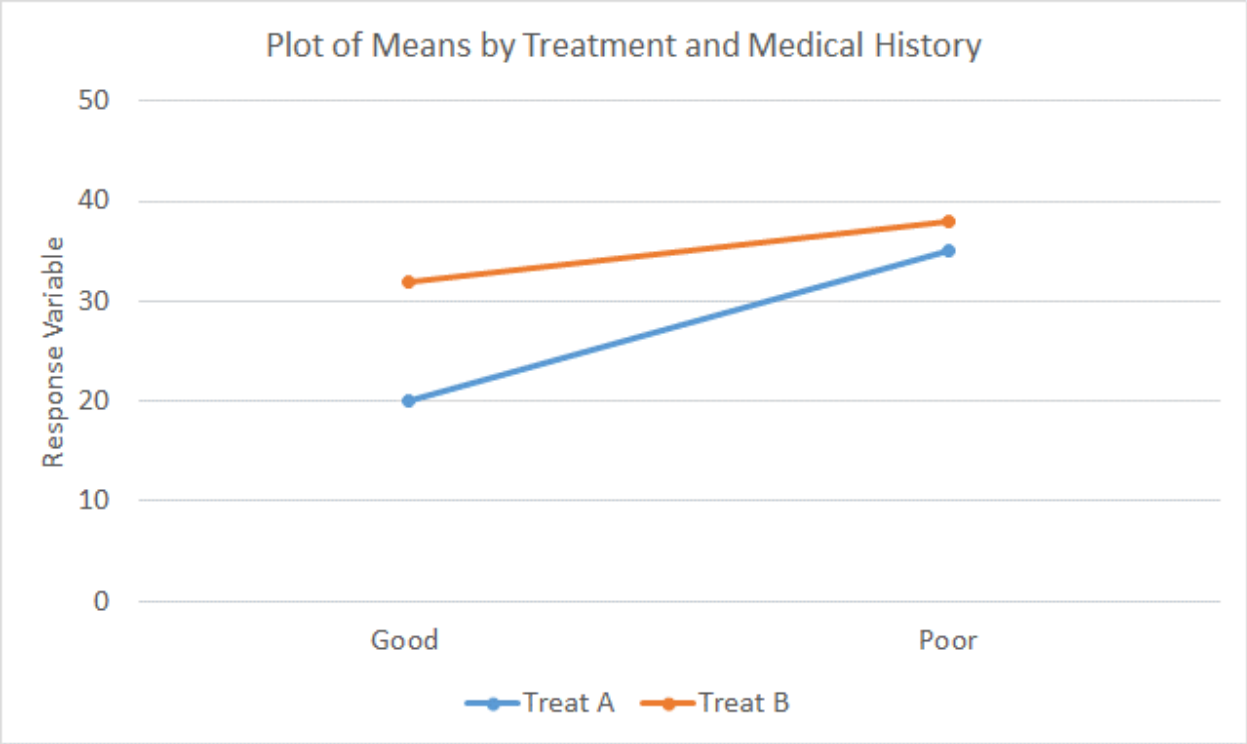
Concept of an Interaction

An interaction occurs when two or more explanatory variables do not act independently on the outcome/response variable. The combined effect of the explanatory variables is not just the sum of their separate effects.

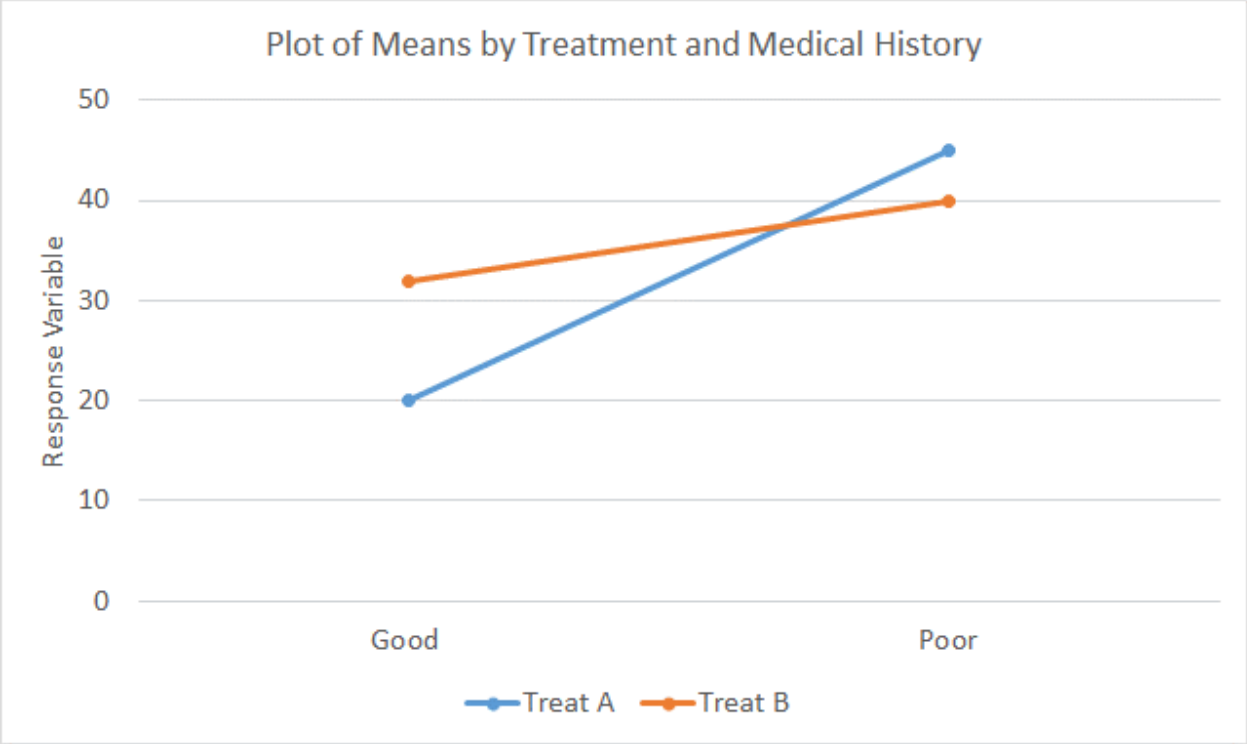
Illustration of an interaction between two factor explanatory variables:
Consider a clinical trial in which patients were randomly allocated to either treatment A or B. Each patient was also classified by their medical history (either good or poor).
The numeric outcome variable was recorded every patient and the means are plotted below by Gender and Treatment.



There is a difference in the outcome for treatments A and B. However this difference is the same for Good and Poor medical histories. No interaction exists between Treatment and Medical History.



There are differences in the outcome for treatments A and B.
The size of the difference is greater for Good than Poor medical history.



There are differences in the outcome for treatments A and B.
The size and sign of difference is different for medical history.

8. We will now add an interaction term to the 2nd model. What is the model we are fitting? What is the interpretation of the parameter estimates?