

Statistical Methods for Bioinformatics

II-5: Correlation vs Interaction

The problem of co-linearity

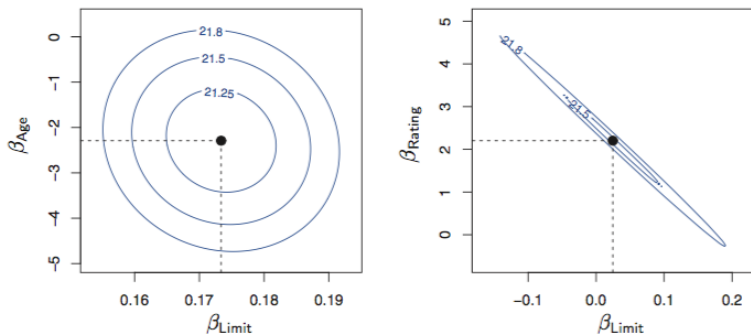
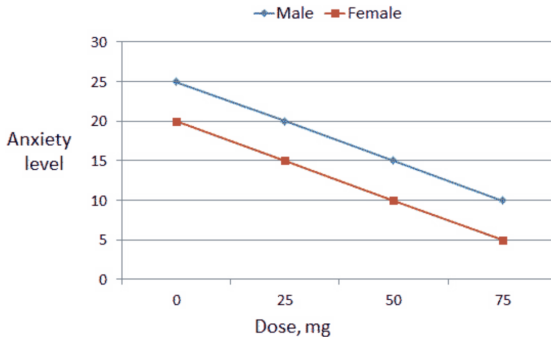


FIGURE 3.15. Contour plots for the RSS values as a function of the parameters β for various regressions involving the **Credit** data set. In each plot, the black dots represent the coefficient values corresponding to the minimum RSS. Left: A contour plot of RSS for the regression of **balance** onto **age** and **limit**. The minimum value is well defined. Right: A contour plot of RSS for the regression of **balance** onto **rating** and **limit**. Because of the collinearity, there are many pairs $(\beta_{\text{Limit}}, \beta_{\text{Rating}})$ with a similar value for RSS.

Interaction

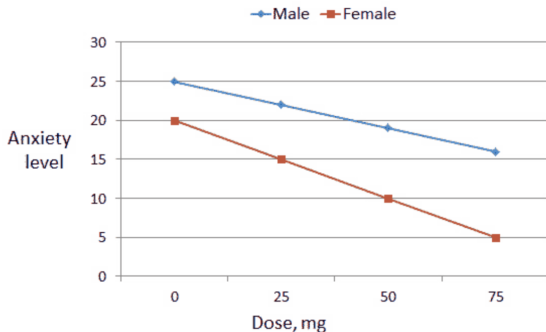
- Imagine an investigation of a drug on the treatment of anxiety. In a simple case you could have an influence of gender on the effect of the drug. Let's start with just independent linear effects:

$$y = \beta_0 + \beta_{drug}x_{drug} + \beta_{sex}x_{sex} + \varepsilon$$



Interactions

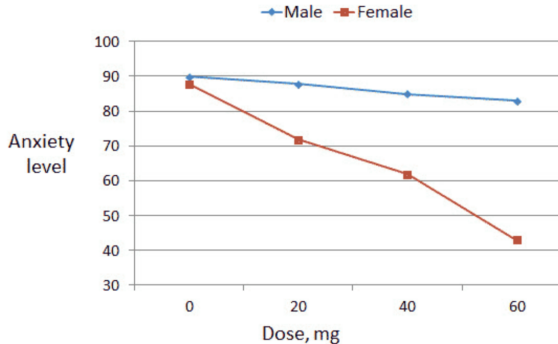
Now imagine that there are interactions:



You could model this like:

$$y = \beta_0 + \beta_{drug}x_{drug} + \beta_{sex}x_{sex} + \beta_{d*s}x_{sex}x_{drug} + \varepsilon$$

Challenges with interactions



You can not interpret the main effect of drug w/o taking into account the gender of the individual. Hence the idea underlying the formulation of additivity is violated. ¹

¹This example with images was taken from stattrek.com