

Linear models with categorical outcomes

Jeff Leek

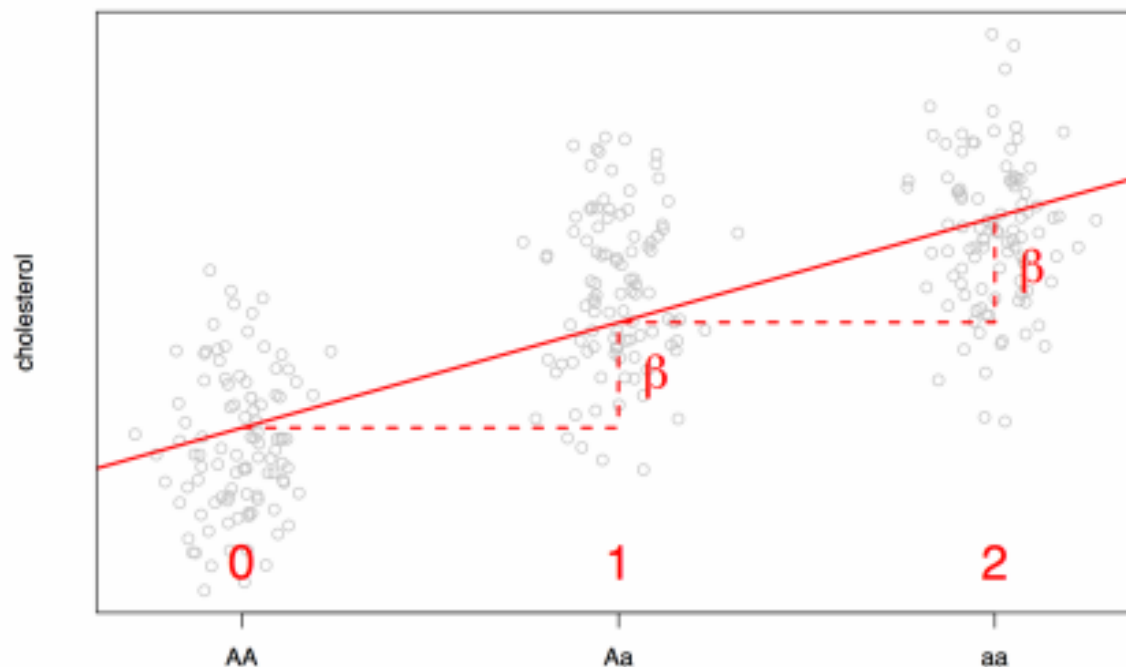
@jtleek

www.jtleek.com

Fitting lines = fitting means

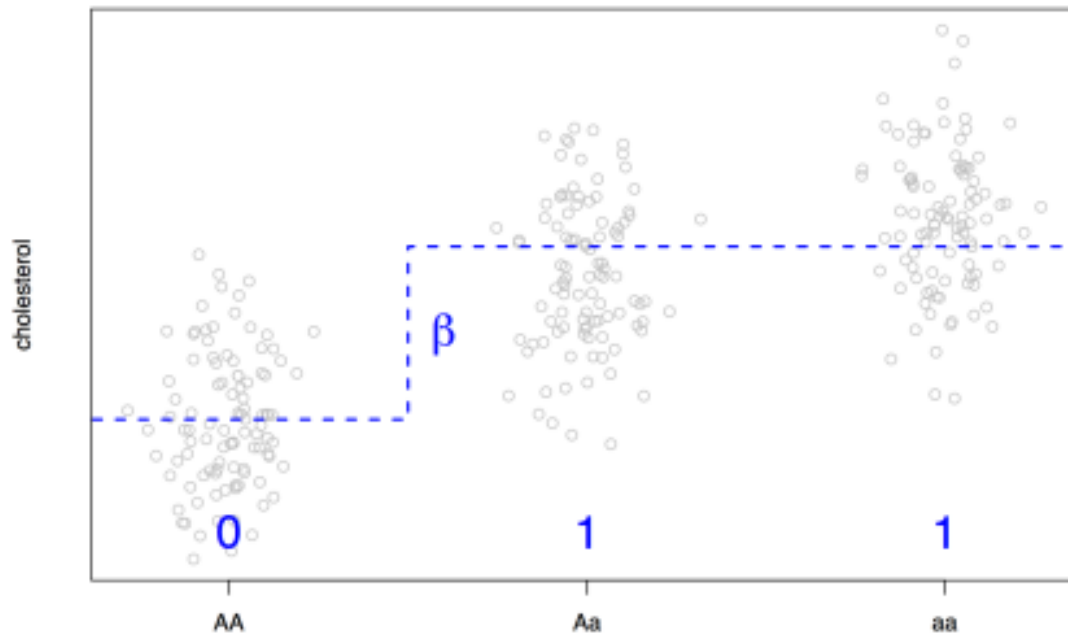
Many analyses fit the 'additive model'

$$y = \beta_0 + \beta \times \text{\#minor alleles}$$



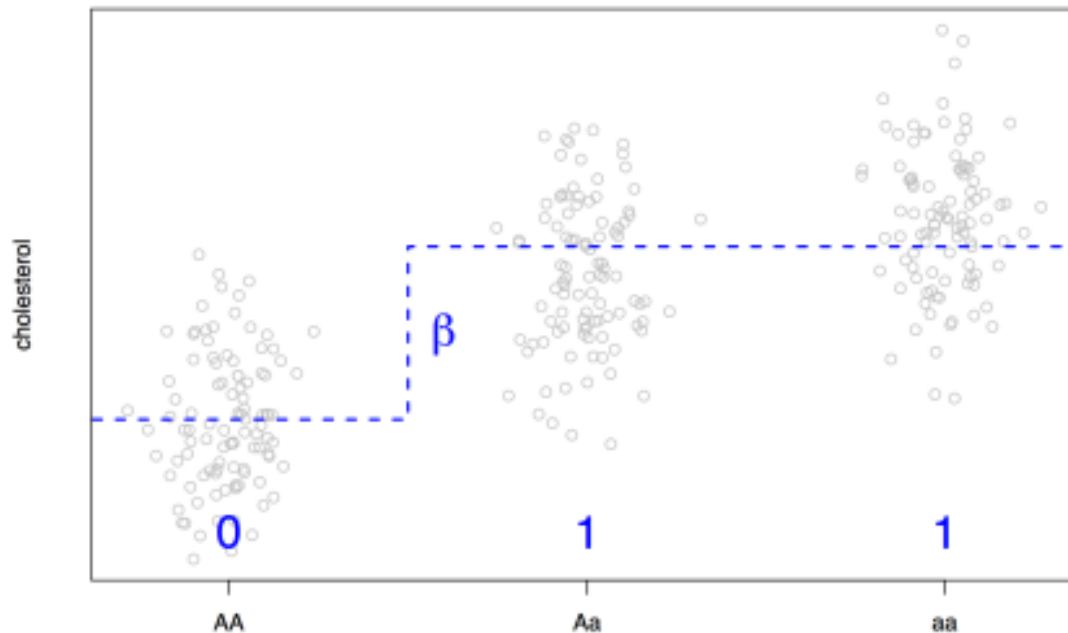
An alternative is the 'dominant model';

$$y = \beta_0 + \beta \times (G \neq AA)$$



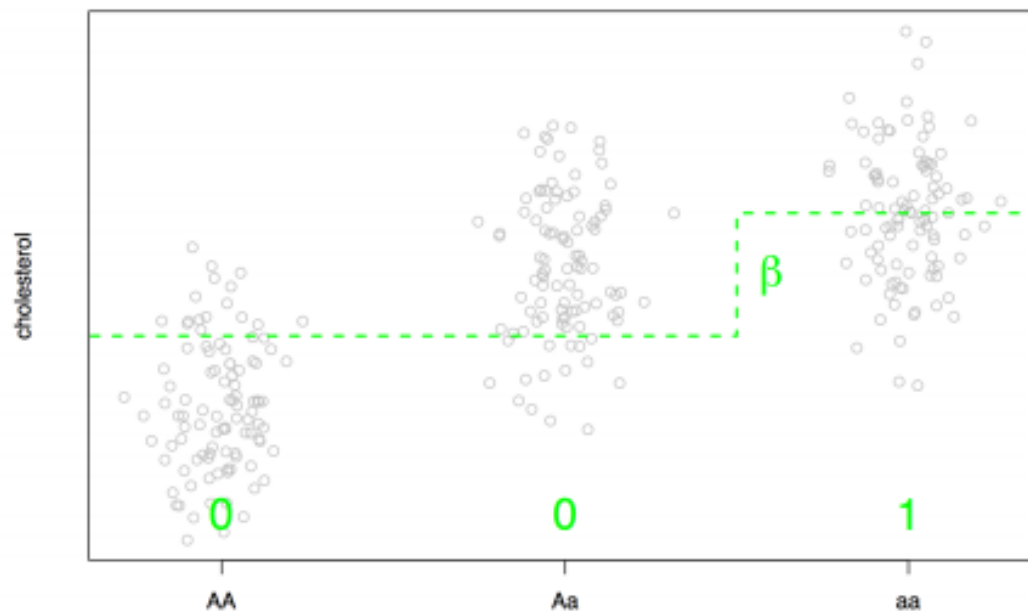
An alternative is the 'dominant model';

$$y = \beta_0 + \beta \times (G \neq AA)$$



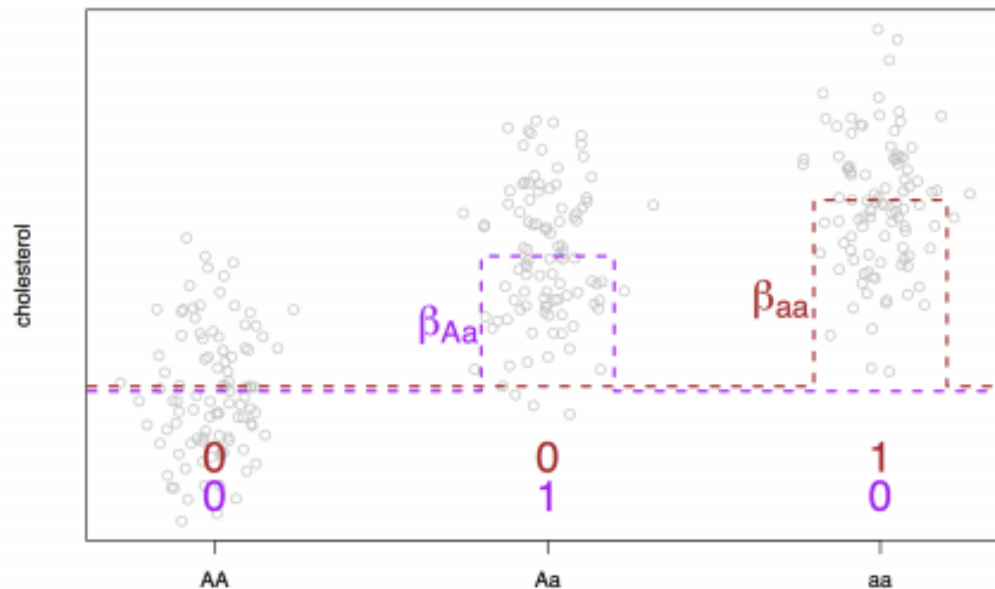
or the 'recessive model';

$$y = \beta_0 + \beta \times (G == AA)$$

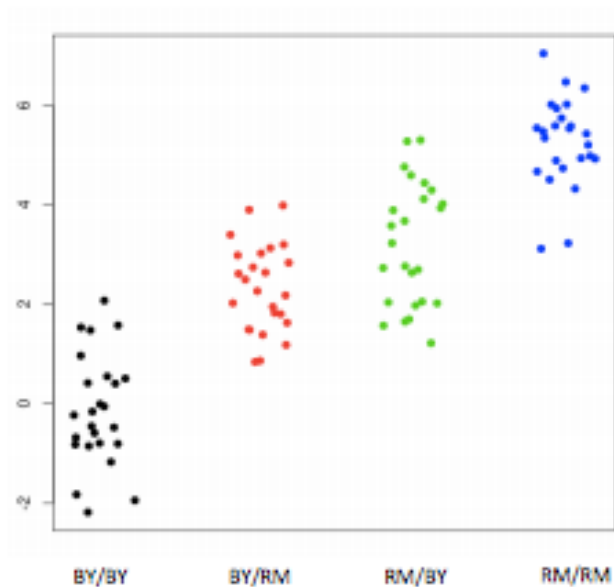


Finally, the 'two degrees of freedom model';

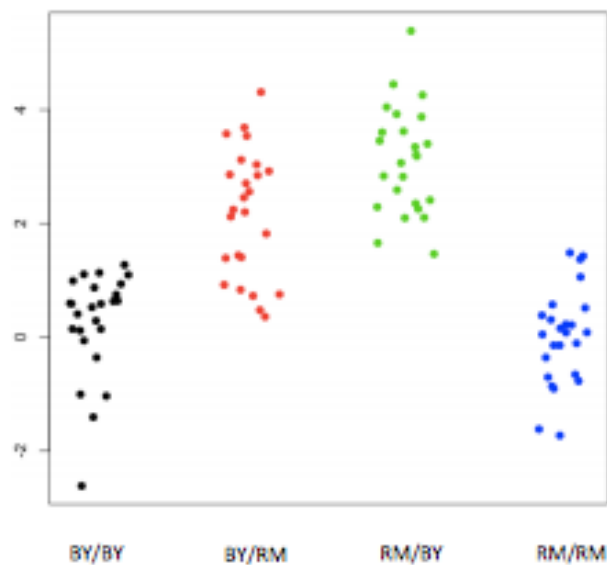
$$y = \beta_0 + \beta_{Aa} \times (G == Aa) + \beta_{aa} \times (G == aa)$$



Interaction terms



No interaction



Interaction

Expression = Baseline + RM Effect + BY Effect + (RM Effect * BY Effect) + Noise

Notes and further reading

- Linear models is a whole class (no joke): <https://www.coursera.org/course/regmods>
- Basic thing to keep in mind is how many levels do you want to fit? What makes sense biologically?
- Great additional notes in Chapter 2 here: <http://genomicsclass.github.io/book/>