Linear models with categorical covariates

Jeff Leek

@jtleek

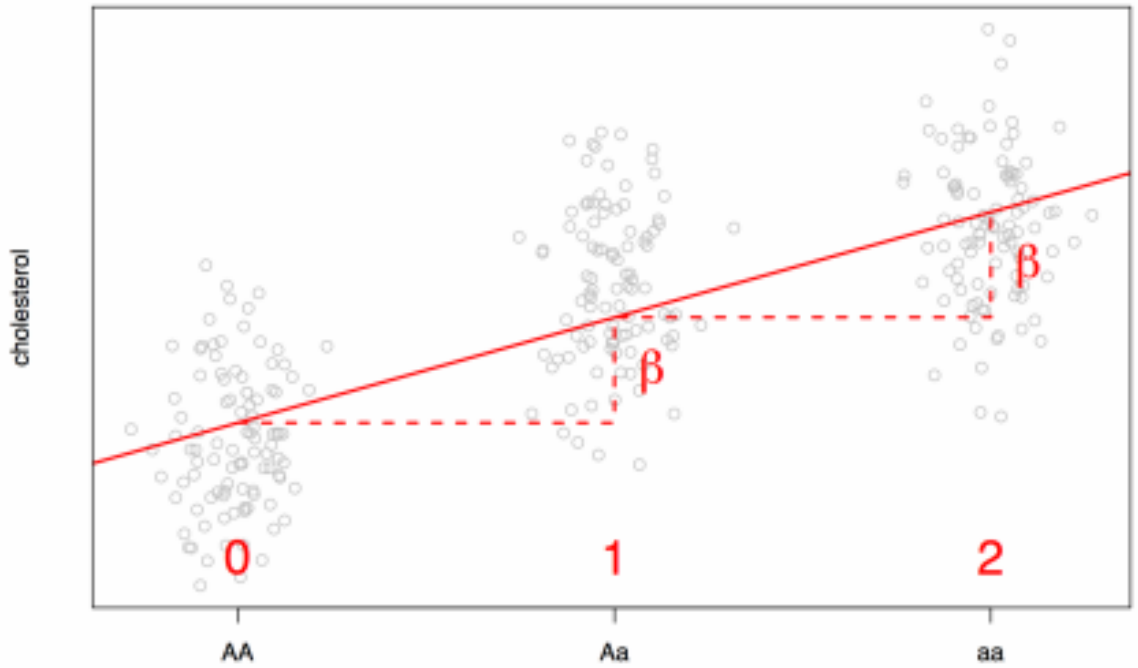
www.jtleek.com

Option 1:
Treat it as a continuous variable

Slide courtesy: Ken Rice and Thomas Lumley

Many analyses fit the 'additive model'

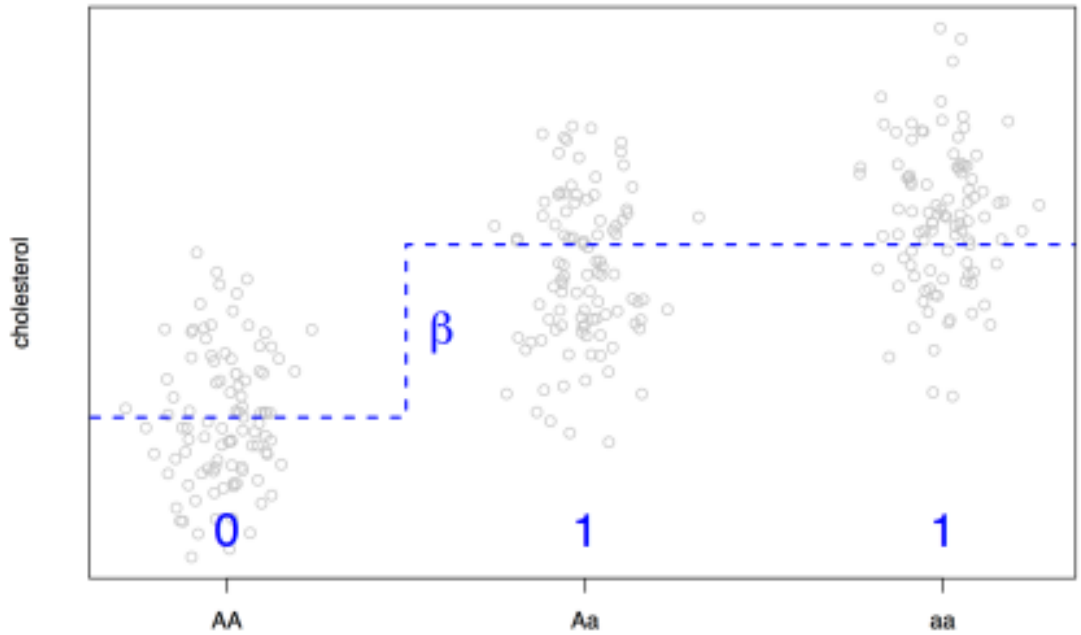
$$y = \beta_0 + \beta \times \#\text{minor alleles}$$



Option 2:
Use regression to fit means

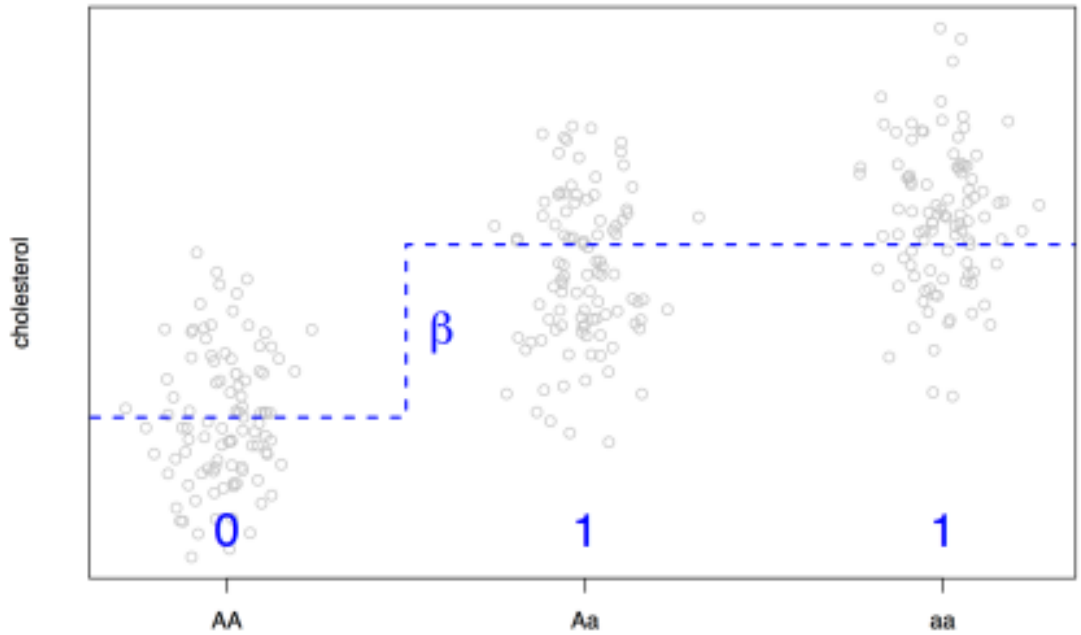
An alternative is the 'dominant model';

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



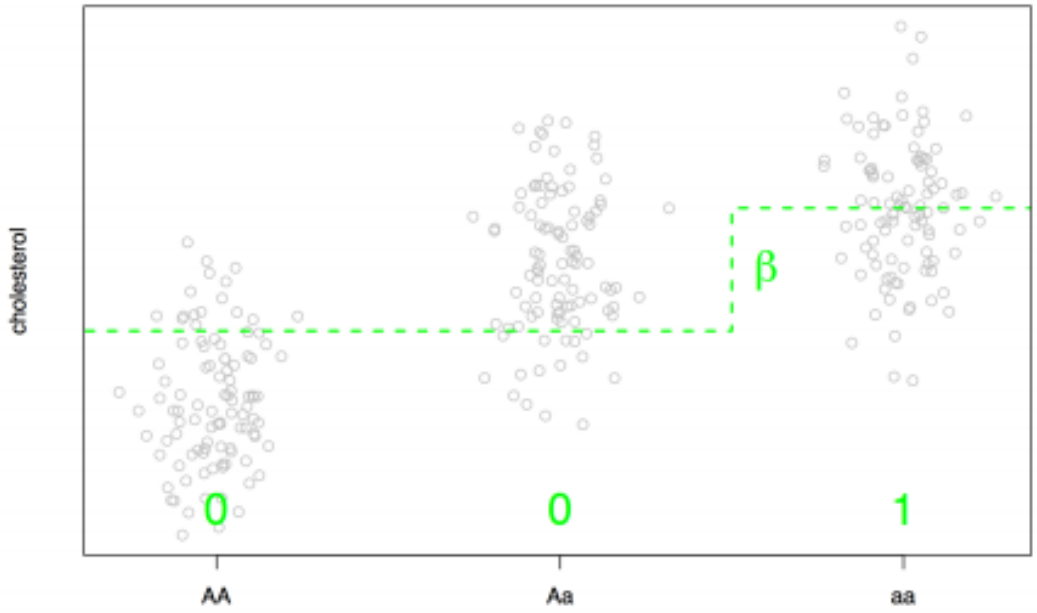
An alternative is the 'dominant model';

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



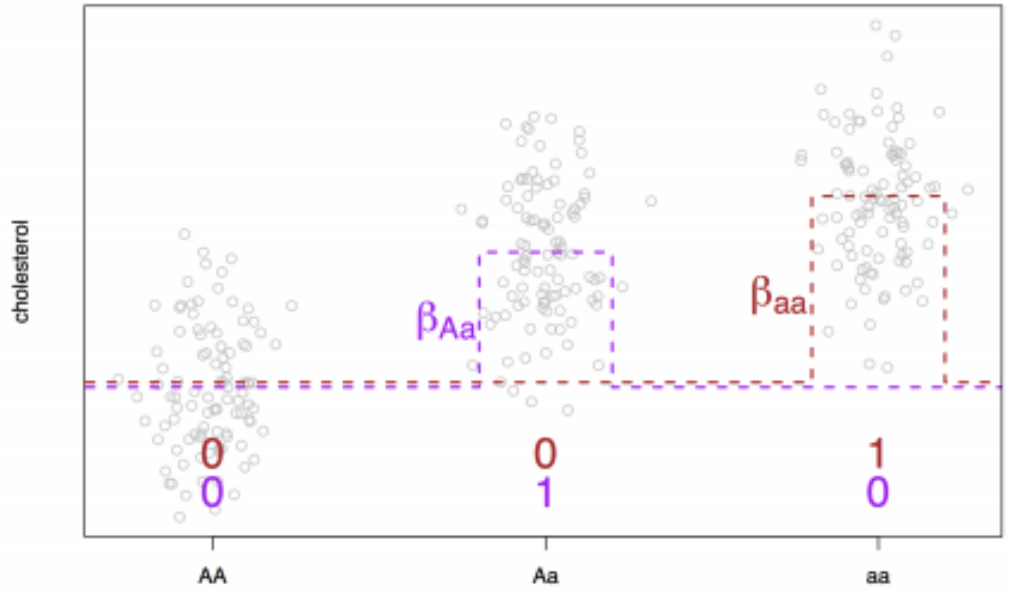
or the 'recessive model';

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



Finally, the 'two degrees of freedom model';

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



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/