

Comparing models

Jeff Leek

@jtleek

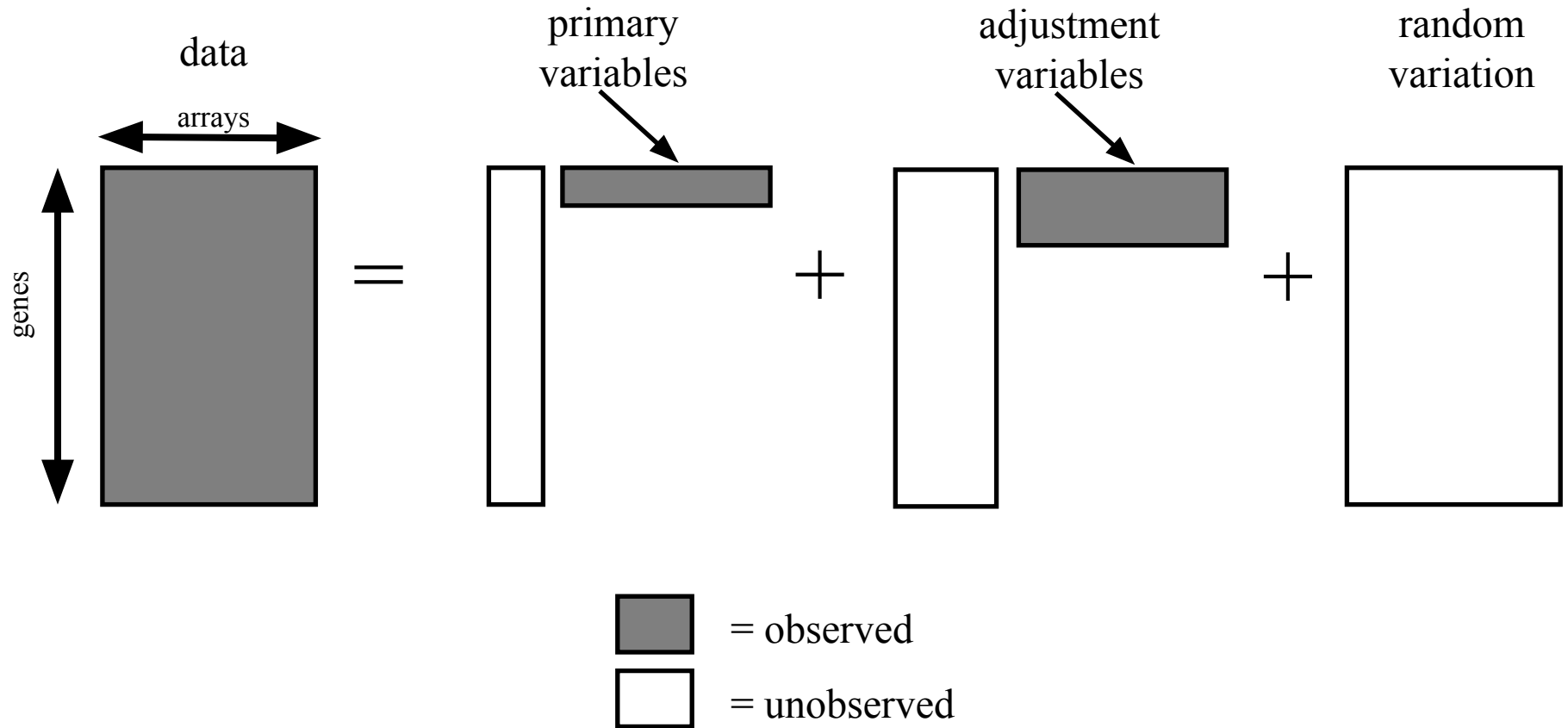
www.jtleek.com

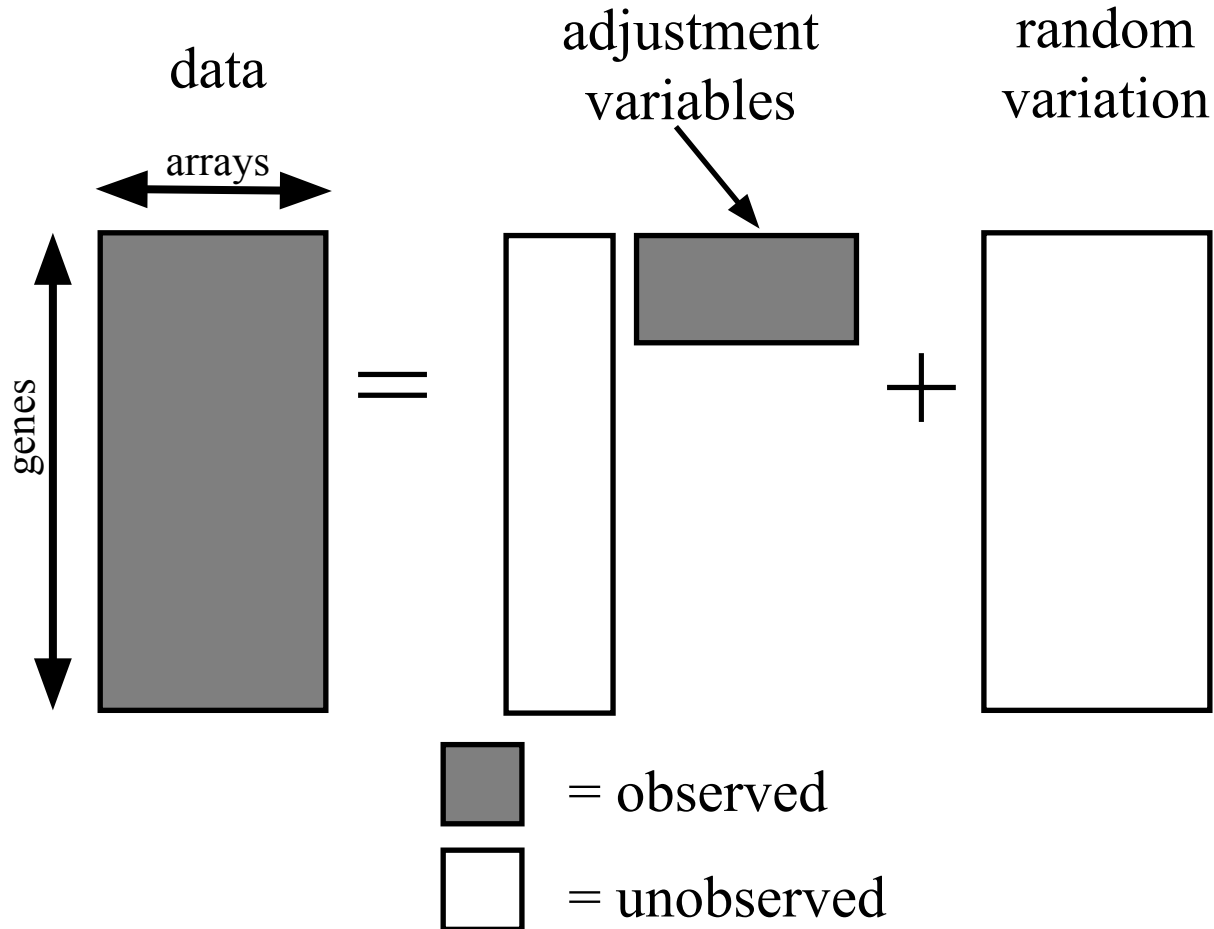
**Common goal: Identify associations
after adjusting for some confounders**

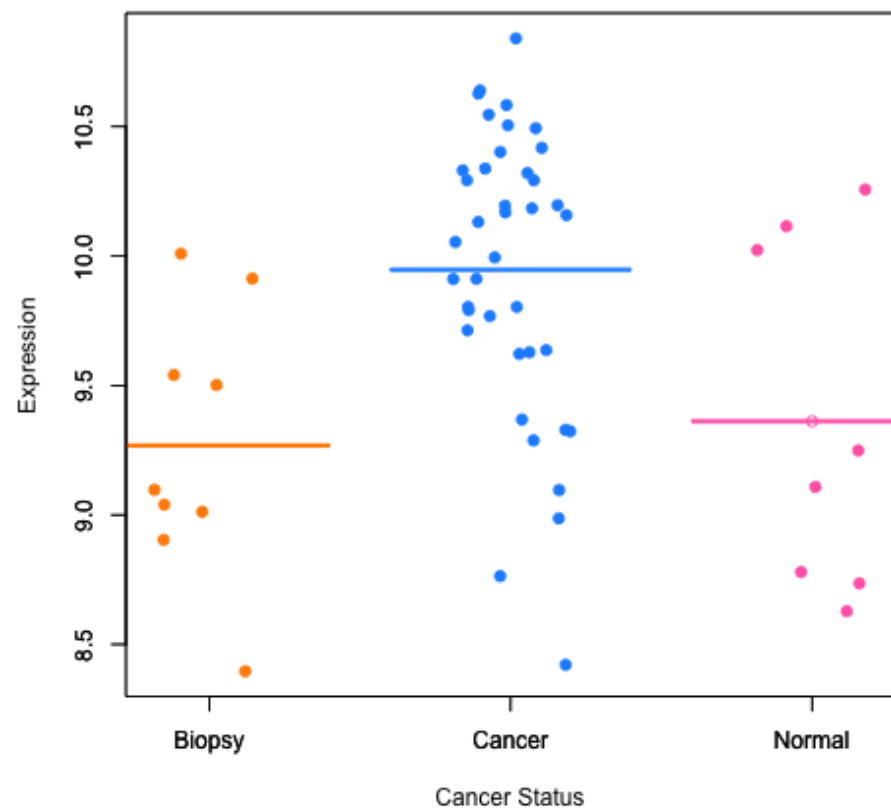
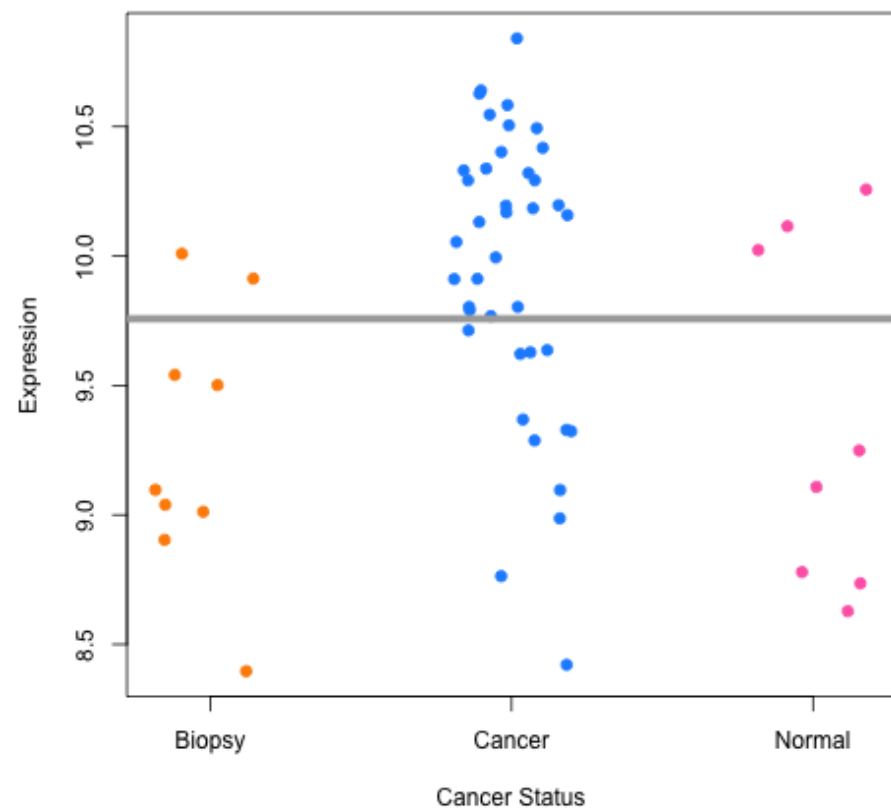
$$Y = b_0 + b_1 P + b_2 B + e$$

P = Phenotype you care about

B = Batch





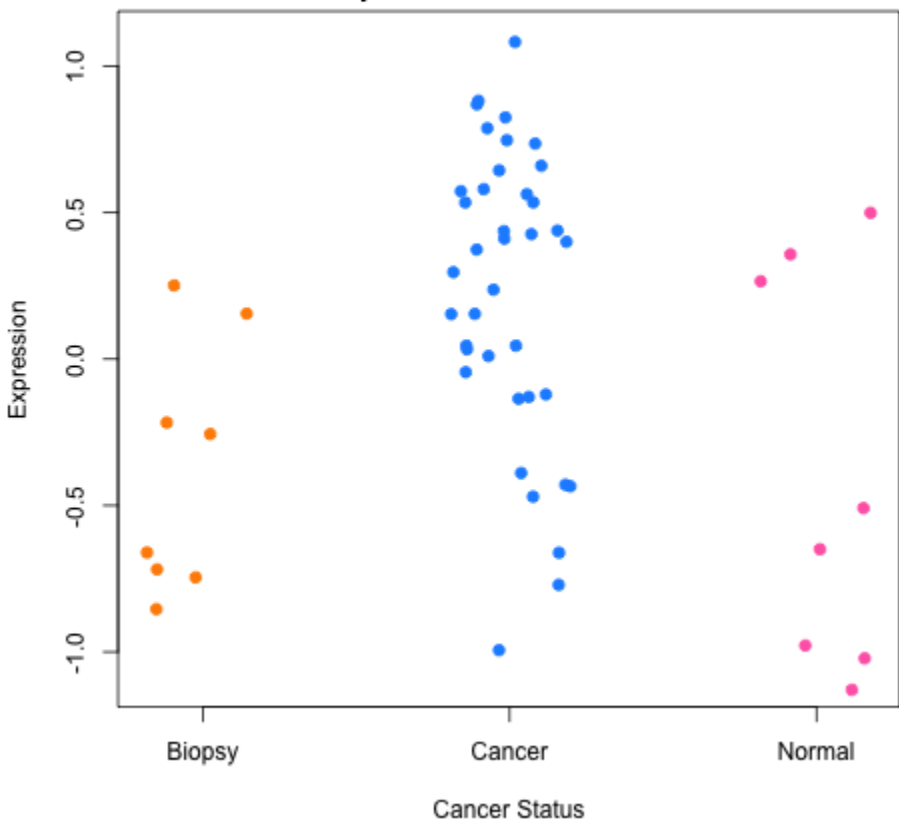


$$R = Y - b_0 - b_1 P - b_2 B$$

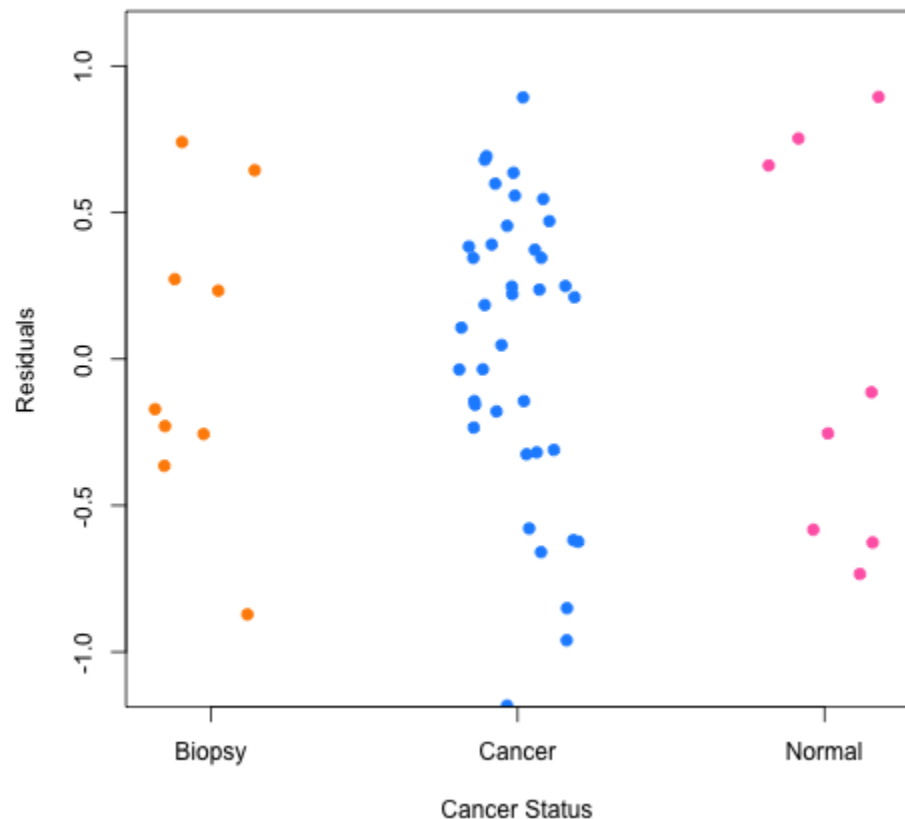
P = Phenotype you care about

B = Batch

$$RSS_0 = \sum_i (y_i - \hat{y}_i^0)^2 \approx 22.05$$



$$RSS_1 = \sum_i (y_i - \hat{y}_i^1)^2 \approx 17.21$$



$$\frac{n - p_1}{p_1 - p_0} \frac{RSS_0 - RSS_1}{RSS_1}$$

$$RSS_k = \sum_i (y_i - \hat{y}_i^k)^2$$

To account for the difference in the number of variables

$$\frac{n - p_1}{p_1 - p_0} \frac{RSS_0 - RSS_1}{RSS_1}$$

$$RSS_k = \sum_i (y_i - \hat{y}_i^k)^2$$

Notes and further reading

- This can be moderated, like the t-statistic
- Linear models for microarray data
 - <http://www.statsci.org/smyth/pubs/limma-biocbook-reprint.pdf>
- edge package vignette
 - <http://bioconductor.org/packages/release/bioc/vignettes/edge/inst/doc/edge.pdf>