Not entirely straightforward

Thomas Lumley

Two topics

- Interaction/effect modification
- Between/within cluster comparisons.

Interactions

- Effect modification describes how the association between X and Y differs for different values of Z

- Effect modification is scale dependent
  - mean vs geometric mean
  - risk difference vs relative risk
- If Z is associated with Y there must be effect modification on most scales, as a simple matter of arithmetic

Interactions vs confounding

- Confounding says that the overall association estimate is wrong, and that stratifying on the confounder is necessary to get the right answer
  - presence of confounding is a scientific issue not just a property of the data
  - confounding by unmeasured variables or by unmodelled variables gives the wrong answer

Interaction vs confounding

- Interaction says that the association is different in different strata
  - this is just a description of the data
  - ignoring interaction corresponds to averaging over the distribution of the effect modifier, which is valid and may even be desirable
  - Unmeasured effect modifiers always exist, but averaging over them is harmless.

Coefficients

- A model for blood pressure with age, dietary potassium (K), and treatment with thiazide diuretics (trt), and interaction

$$E[BP] = \alpha_0 + \alpha_1 \text{age} + \alpha_2 \text{trt} + \beta_1 K + \gamma \text{trt} \times K$$

- Model says that the age association doesn't depend on the other variables but that the association with potassium depends on whether the person takes diuretics (or *vice versa*)

Contrasts

- Difference in E[BP] per unit of K intake, in treated

$$E[BP] = \alpha_0 + \alpha_1 50 + \alpha_2 1 + \beta_1 1 + \gamma 1 \times 1$$

$$E[BP] = \alpha_0 + \alpha_1 50 + \alpha_2 1 + \beta_1 0.5 + \gamma 1 \times 0.5$$

$$\beta_1 0.5 + \gamma 0.5$$

- In untreated

$$E[BP] = \alpha_0 + \alpha_1 50 + \alpha_2 0 + \beta_1 1 + \gamma 0 \times 1$$

$$E[BP] = \alpha_0 + \alpha_1 50 + \alpha_2 0 + \beta_1 0.5 + \gamma 0 \times 0.5$$

$$\beta_1 0.5$$

Contrasts: for K as predictor of interest

- Coefficient of K (`main effect') is the slope (difference in E[BP] per unit K) in people with `trt=0` (untreated)
- Interaction coefficient is the difference in slope per unit of `trt`,
  - difference in slope between treated and untreated
  - difference between average difference in E[BP] per unit difference in K intake in treated and untreated

Contrasts: for trt as predictor of interest

- Coefficient of trt (`main effect') is the difference between treated and untreated at K=0,
  - ie, at zero potassium intake, not useful

- Interaction coefficient is the average difference in the treatment effect per unit difference in K intake
  - difference in coefficient per unit K intake
  - average difference in E[BP] between treated and untreated was this much higher for each unit higher K intake

Coding

- Since interpretation depends on coding, it may be useful to code the variable so that zero is a value inside the range of the data

- Important to look at how the variable is coded, to make sure you know where zero is.

Clusters

- Correlation in clusters can increase or decrease precision, depending on whether the comparison is between or within clusters
- Comparing BP for men and women based on 200 measurements
  - more information in 200 measurements on 200 separate people than in 2 measurements from each of 100 people
- Comparing BP before and after treatment
  - more information in before-after pairs on 100 people than 100 people before and 100 different people after

Separating components

- Can always separate between-cluster and within-cluster comparisons for a variable
  - average the variable over a cluster to get the between-cluster variable
  - subtract the cluster average from the individual value to get the within-cluster variable
- Between-cluster variable has the same value for all observations in a cluster
- Within-cluster variable has same mean (zero) for all clusters

Separating components

- Coefficient of between-cluster variable is the difference in mean outcome between clusters whose mean predictor differs by 1 unit
  - incorporates contextual effects
  - confounding by differences between clusters
- Coefficient of within-cluster variable is the difference in mean outcome between observations in the same cluster differing by 1 unit in the predictor.
  - differences between clusters cancel out

Separating components

- Useful when one component is likely to be more confounded
- Useful if contextual effects are interesting
- Useful to see if the associations are similar.

Data collection

- Mark-sense forms: one-dimensional summary statistics are made public
- Yellow sheets: returned only to me, not sent to anyone else.
- Remember to answer the TA questions

That's all, folks