Session 6

Coping with missing data

Matt Denwood

2021-06-30

Session 6: Coping with missing data

Outline

- When tests are missing at random -> easy (template_huiwalter)

- When tests are missing structurally -> does the test missingness correspond with expected prevalence?

- Removing data for cross-validation purposes

Types of missingness

- MCAR: Missing completely at random
  - There is absolutely no pattern to the missingness eg:
    - The animal was too aggressive to facilitate a blood sample
    - Somebody dropped the samples on the way to the PCR machine
  - This is the best kind

Types of missingness

- MCAR: Missing completely at random
  - There is absolutely no pattern to the missingness eg:
    - The animal was too aggressive to facilitate a blood sample
    - Somebody dropped the samples on the way to the PCR machine
  - This is the best kind

- MAR: Missing at random
  - There is a pattern to the missingness but we can explain it eg:
    - Test A was not done in population 1 because of financial limitations
    - Test B was only done if Test A was positive
  - We need to either include the explanation in the model, or ensure that it is not confounded with expected disease status

Types of missingness

- MCAR: Missing completely at random
    - There is absolutely no pattern to the missingness eg:
        - The animal was too aggressive to facilitate a blood sample
        - Somebody dropped the samples on the way to the PCR machine
    - This is the best kind

- MAR: Missing at random
    - There is a pattern to the missingness but we can explain it eg:
        - Test A was not done in population 1 because of financial limitations
        - Test B was only done if Test A was positive
    - We need to either include the explanation in the model, or ensure that it is not confounded with expected disease status

- MNAR: Missing not at random
    - There is an unknown (or unrecorded) pattern to the missingness

Practical session 6

Points to consider

1. How does missing-at-random data impact your results?

2. What about if you have data using confirmatory tests?

3. How can we use cross-validation as a method of checking assumptions?

Summary

- Observations that are MCAR are trivial to deal with using JAGS

- We can also treat MAR observations as if they are MCAR as long as the reason for missingness does not confound with expected prevalence, or we allow prevalence to differ between groups where the structural missingness differs

- MNAR is bad news

- Deliberately making observations missing is a good way to assess model assumptions