School of Mathematics and Statistics
Faculty of Science, Technology, Engineering and Mathematics

The Open University

2024 PhD Projects

Project title	A comprehensive approach to test for missing not at random
Principal supervisor	Prof Stefanie Biedermann
Second supervisor	Dr Karen Vines
Discipline	Statistics
Research area/keywords	Missing data, design of experiments, hypothesis testing
Suitable for	Full time applicants, Part time applicants

Project background and description

Missing data occur in many application areas, including sample surveys and clinical trials. Little and Rubin (2019) classify missingness into three categories: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). If the probability that a response is missing depends on the (unobserved) value of the response, we have MNAR missingness, and this can cause bias in the data analysis. For example, employees who earn high salaries are less likely to disclose their income in a survey, which means that inference on the income distribution may be biased substantially if standard analysis methods are applied.

It is therefore of interest to find out if missingness in a dataset is MNAR. However, using only the incomplete dataset, this is an untestable assumption. Carpenter and Kenward (2012) propose two statistical tests for MNAR missingness if some of the originally missing responses can be obtained, or recovered, for example through follow-up phone calls in a survey. Some exploratory work on investigating the properties of one of these tests has been done by Noonan et al. (2022).

In this project, we will build on the work by Noonan et al. (2022). In particular, we will extend the methodology to include a larger variety of models. As recovering missing responses can be expensive, the follow-up should be done in a cost-effective way while still guaranteeing high power of the test. In order to do this, we will apply and develop methodology from design of experiments (see, e.g., Atkinson et al. (2007)) to select our recovery sample.

This statistics project at the interface of missing data problems, design of experiments and hypothesis testing will require both theoretical and computational work. The candidate should have a strong knowledge of statistics with a suitable degree. Computational (e.g. R) and programming experience would be useful.

Background reading/references

- Atkinson, A., Donev, A. and Tobias, R. (2007). Optimum experimental designs, with SAS. Oxford University Press.
- Carpenter, J. and Kenward, M. (2012). Multiple imputation and its application. John Wiley & Sons.
- Little, R. and Rubin, D. (2019) Statistical analysis with missing data. 3rd edition. John Wiley & Sons.
- Noonan, J., Adediran, A., Mitra, R. and Biedermann, S. (2022). An integrated approach to test for missing not at random. https://arxiv.org/abs/2208.07813