Offer #2023-07002

Post-Doctoral Research Visit F/M Handling missing values in multiblock data

**Contract type**: Fixed-term contract  
**Level of qualifications required**: PhD or equivalent  
**Fonction**: Post-Doctoral Research Visit

**About the research centre or Inria department**

The Inria centre at Université Côte d'Azur includes 37 research teams and 8 support services. The centre's staff (about 500 people) is made up of scientists of different nationalities, engineers, technicians and administrative staff. The teams are mainly located on the university campuses of Sophia Antipolis and Nice as well as Montpellier, in close collaboration with research and higher education laboratories and establishments (Université Côte d'Azur, CNRS, INRAE, INSERM ...), but also with the region's economic players.

With a presence in the fields of computational neuroscience and biology, data science and modeling, software engineering and certification, as well as collaborative robotics, the Inria Centre at Université Côte d'Azur is a major player in terms of scientific excellence through its results and collaborations at both European and international levels.

**Context**

Within the framework of the PEPR, DIGPHAT Digital Pharmacological Twins where the aim is to model multi-scale and longitudinal data in pharmacology, the candidate will develop methods to handle missing values.

**Assignment**

Assignments:

With the help of Julie Josse, the recruited person will conduct research on missing values.

For a better knowledge of the proposed research subject:

A state of the art, bibliography and scientific references are available at the following URL, do not hesitate to log in:

Website with resources on missing values [https://rmisstastic.netlify.app/](https://rmisstastic.netlify.app/)

Team website [https://team.inria.fr/premedical/](https://team.inria.fr/premedical/)

**Collaboration**:

The recruited person will be in connection with Jean-Baptiste Woillard who is a key interlocutor for the PEPR DIGPATH

**Main activities**
The project will use data sets of different types, sources and cohorts, standardization procedures will be carried out to ensure the quality of the data and the removal of experimental technical biases. In addition, missing values are systematically encountered in clinical, biological and pharmacological data collections and can lead to significant loss of data that cannot be taken into account by computational models. Statistical and machine learning approaches for the imputation of missing values will be implemented to take this issue into account and generate optimal databases for the development of pharmacological models.

Naive approaches such as complete case analysis which can lead to important bias, cannot be applied in high-dimensional settings when almost all data can be deleted. There exists an abundant literature on the topic and many methods are available either to estimate some parameters (EM, multiple imputation) or to do supervised learning with missing values. However, missing data structured by block of variables represent a new field of the statistical analysis of missing data, which has not been studied until now. We will consider two approaches: 1) imputing the data so that it can be analysed by any analyst with any statistical methods and 2) modifying the statistical algorithm so that it can handle missing values. For the former we will consider imputation of missing values with low-rank methods for multi-block data to take into account the relationships between variables of different groups and the relationships between observations while reducing the dimensionality of the problems. Such methods can also be considered in a multiple imputation setting to reflect the variability due to missing values and assess which confidence should be given to an analysis performed from an incomplete data set. Handling missing values with multi-source data (either structured by block of variables of different nature or by blocks of observations) will be an innovation. From a methodological point of view, the goals are to develop a framework to handle missing values in data integration and to account for the variability due to missing values. Many projects require complex, heterogeneous, and large databases to be analyzed by multiple teams for which data repositories are created. Such data are highly incomplete so that the imputation methods developed in this project could become standard in this context and would avoid important biases due to bad management of missing data.

Skills

Technical skills and level required:

Languages:

Relational skills:

Other valued appreciated:

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave +10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Remuneration

Gross Salary: 2788 € per month

General Information

- Theme/Domain: Computational Neuroscience and Medicine
- Statistics (Big data) (BAP E)
- Town/city: Montpellier
- Inria Center: Centre Inria d'Université Côte d'Azur
- Starting date: 2024-02-01
- Duration of contract: 1 year
- Deadline to apply: 2024-02-29

Contacts

- Inria Team: PREMEDICAL
- Recruiter: Josse Julie / julie.josse@inria.fr
About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

The keys to success

There you can provide a “broad outline” of the collaborator you are looking for what you consider to be necessary and sufficient, and which may combine:

- tastes and appetencies,
- area of excellence,
- personality or character traits,
- cross-disciplinary knowledge and expertise...

This section enables the more formal list of skills to be completed and ‘lightened’ (reduced):

- "Essential qualities in order to fulfil this assignment are feeling at ease in an environment of scientific dynamics and wanting to learn and listen."
- "Passionate about innovation, with expertise in Ruby on Rails development and strong influencing skills. A thesis in the field of **** is a real asset."

Warning: you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

Instruction to apply

Applications must be submitted online on the Inria website. Collecting applications by other channels is not guaranteed.

The position is open to:
- Inria internal mobility, remuneration according to statutory conditions
- mobility from other public body, by posting for a period of three years, renewable, remuneration according to statutory conditions
- in short term contract from service fixed-term

Defence Security:

This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy:

As part of its diversity policy, all Inria positions are accessible to people with disabilities.