J.H. Ginn - 7 November 2019 - 15:06:33

(1.1) **Title**

MSc Methodology & Statistics for the Behavioral, Biomedical and Social Sciences

(1.2) **Study type**

Single study/research project

(1.3) **Division**

Methodology & Statistics

(1.4) **Start date**

01 September 2019

(1.5) **End date**

10 May 2020

Basic study information

J.H. Ginn - 20 January 2020 - 14:30:14

(1) Name(s), position(s) and division(s) of the responsible researcher(s):

Name Position Division E-mail

Emmeke Aarts

Assistant professor

Methodology and Statistics

e.aarts@uu.nl

(2) Name(s), position(s) and division(s) of the executive researcher(s):

Name Position Division E-mail

Jasper Ginn

Postdoctoral researcher

Methodology and Statistics

j.h.ginn@students.uu.nl

(3) Research area/discipline:

Methodology & Statistics

(4) What is the study's main objective (hypothesis)?

The thesis uses a Monte Carlo simulation design to determine the statistical power of multilevel hidden markov models ('the model'). The study may include an application of the model to real, previously collected and open data.

(5) Primary funder of the study:

N/A

(6) Does the study concern a multi-center project, e.g. in collaboration with other universities, a GGZ mental health care institution, or a university medical center?

○ Yes
◉ No

(7) Where will the study (data collection) be conducted? If this is abroad, please note that you have to be sure of the local ethical codes of conducts and permissions

No new data will be collected for the purposes of this study

Study details (I)

J.H. Ginn - 20 January 2020 - 14:30:14

(1) Does your study exclusively concern the analysis of existing data, document or records? Where can the data be found?

◉ Yes
○ No

The data is described in the following article: B Kemp, AH Zwinderman, B Tuk, HAC Kamphuisen, JJL Oberyé. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. IEEE-BME 47(9):1185-1194 (2000). It is available for download at: https://physionet.org/content/sleep-edfx/1.0.0/

(2) Are the sources of the existing data, documents or records publicly available?

◉ Yes
○ No
○ Not applicable

(3) Will the data be processed by the principal investigator in such a manner that participants can be identified either directly or indirectly (through identifiers (such as a code) linked to them)?

○ Yes
◉ No

(4) If the study uses de-identified (or pseudonymized) data, does the responsible or executive researchers have access to the key to the code permitting re-identification of the person whose data are being studied?

○ Yes
◉ No
○ Not applicable

(5) The research will involve only the use of anonymous survey procedures, interview procedures or the observation of public behavior

◉ Yes
○ No

(6) Will participants be asked to report their own or others' sexual experiences, alcohol or drug use, or suicidal thoughts, and will their identities be known to you?

○ Yes
◉ No

Study details (II)

J.H. Ginn - 20 January 2020 - 14:30:14

(1) Will participants that are recruited be >16 years?

◉ Yes
○ No

(2) Will participants that are recruited be mentally competent (wilsbekwaam in Dutch)?

◉ Yes
○ No

(3) Will participants that are recruited provide active informed consent?

◉ Yes
○ No

(4) Will the probability and magnitude of possible harm or discomfort anticipated in the research be greater than ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests?

○ Yes
◉ No

(5) Does the participant population contain vulnerable persons? (e.g., incapacitated, children, mentally challenged, traumatized, pregnant), or a taboo subject (e.g., own or others' sexual activity, hard drug use, suicide thoughts, religious belief, political preference)

○ Yes
◉ No

(6) Will participants be subjected to:

	Yes	No
Inquiries into their sexual behavior or orientation	○	◉
Inquiries into drug use (also alcohol, smoking, soft drugs)	○	◉
Assessment of delinquency	○	◉
Inquiries relating to religious or philosophical belief	○	◉
Inquiries relating to political opinions	○	◉
Inquiries into ethnic origin	○	◉
Inquiries into trade Union membership	○	◉
Inquiries into violent experiences	○	◉
Inqueries into personal health	◉	○
Inquiries into criminal convictions and offences	○	◉
Shocking images/videos	○	◉
Deception (information letter does not state real study objective)	○	◉
Physical pain (electrical/ thermal shocks, noise)	○	◉
Following orders behaviorally (by force, or outside the context of the lab with possible harmful consequences for the participant or his/her social environment?)	○	◉
A new technique for data collection?	○	◉

(7) Will data of the following categories be processed:

	Yes	No
Photo data	○	◉

	Yes	No
Video data	○	◉
Biological material (buccal, blood, hair)	○	◉
Genetic data	○	◉
Biometric data (fingerprint, iris or retinal scan, voice recognition and face scan)	○	◉
Directly identifying data (name, address, date of birth or a combination of those items)	○	◉

Study details (III)

J.H. Ginn - 20 January 2020 - 14:30:14

(1) What is the study's theoretical and practical relevance? (500 words max.):

Hidden Markov Models are often used to model data that is thought to be the result of latent states. For example, we know that, during sleep, the brain emits different activity patterns. These patterns are the direct result of the sleep state we are in (REM/ deep/ light sleep etc.). Recent extensions of the HMM allow us to model such dynamics for multiple subjects using multilevel framework. This model is called a Multilevel Hidden Markov Model (mHMM) or Mixed Markov Model (MxMM). The model that has been developed by Dr. Emmeke Aarts uses Bayesian Bayesian methods, which allows the model to converge faster and more reliably. Because of the novelty of this model, there is a need to examine the accuracy and the stability of parameter estimates when varying the sample size and the number of time points for each person. This, in turn, will help applied researchers in designing their experiments when using mHMMs. In my thesis I develop a Monte Carlo simulation to assess the number of subjects and the number of data points needed to obtain reliable estimates.

(2) What are the central hypotheses?

The purpose of the thesis is to examine the required sample size of persons, as well length of the time-series data available for each of the subjects. In particular, I will address the following research question: What number of subjects are needed and how long should each subject's time-series data be for Multilevel Hidden Markov Models to obtain reliable estimates of model parameters using multivariate outcome data?

(3) What is the study's design and procedure? (500 words max.):

I will conduct a simulation study to answer the research question using simulated data based on existing research about sleep stages. In the thesis, I will vary the sample size of persons and the number of observations for each person and estimate the mHMM using the R statistical language and the R library mHMMBayes. To this end, I will create several simulation scenarios that cover situations in which the model should succeed in estimating the parameters to situations where it should fail. The outcomes of the scenarios will be assessed

using common metrics for accuracy, bias and coverage. All model estimates will be stored. I will make the data and R code used in the thesis available on GitHub. It is common to base the simulated data on observed datasets. In this study, I use publicly available EEG measurements and sleep state annotations. The EEG datasets measure brain activity. The annotations describe what state the subject is likely in at any one time. After basic preprocessing to normalize and clean up the EEG data, I will fit an mHMM on the data to obtain parameter estimates that can be used to simulate new datasets. This will be done using the package 'mHMMBayes'. The study may include a section in which I analyze the EEG data using the mHMM. This serves as a practical example for other researchers.

(4) Optional attachments:

No files have been uploaded yet

(5) What data collection instruments, stimuli and/or manipulations will be used?

N/A

(6) Optional attachments:

No files have been uploaded yet

(7) Please state which statistical procedures will be used.

- Monte Carlo Simulations - Multilevel Hidden Markov Models

(8) Will a method be used that may, by coincidence, lead to findings of which the participant should be informed? If so, what actions will be taken in the case of a coincidental finding?:

○ Yes
◉ No

Participants

J.H. Ginn - 20 January 2020 - 14:30:14

(1) Please provide reasons to justify why this particular group of participants should be subjected to these conditions.

N/A

(2) What possible risks could participating in the study hold for participants?

N/A

(3) What measures are implemented to minimize risk for participants?

N/A

(4) How does the burden on the participants compare to the study's potential scientific contribution (theory formation, practical usability)?:

N/A

(5) What is the number of participants? Provide a power analysis and/or motivation for the number of participants. The current convention is a power of 0.80. If the study deviates from this power, the FERB would like you to justify why this is necessary :

N/A

(6) Age category of the participants

N/A

(7) How will the participants be recruited

N/A

(8) Please state any specific in- and exclusion criteria and how these are tested.

N/A

(9) How much time will the prospective participants have to decide as to whether they will indeed participate in the study?:

N/A

(10) Are the participants fully free to participate and terminate their participation whenever they want and without stating their grounds for doing so?:

◉ Yes
○ No

(11) Will the participants be in a dependent relationship with the researcher?:

○ Yes
◉ No

(12) Is there an independent contact person or a general email address of a complaint officer to whom the participant can contact?

○ Yes
◉ No

(13) What time investment and effort will be requested from participants?

N/A

(14) Will the participants be compensated for their efforts? How (financial reimbursement, travelling expenses, otherwise). What is the amount?

○ Yes
◉ No

(15) Will this compensation depend on certain conditions, such as the completion of the study?

○ Yes
◉ No

Datamanagement

J.H. Ginn - 20 January 2020 - 14:30:14

(1) Who will be responsible for managing access to the data?

Jasper Ginn

(2) What type of data will you collect or create? Please provide a brief description of the data, including the type, volume (if known), format and content.

The data generated by the study are as follows: - Simulated data generated based on the sleep data set described in the previous steps. These datasets will contain e.g. simulated observed EEG, EKG and heart rate data based on average values found in the sleep data set. - Results of applying the multilevel hidden markov model to the simulated data. These concern estimates of the parameters. They will be stored as RDS (compressed comma-delimited) files.

(3) Will you be exchanging (personal) data with organizations/research partners outside the UU?

○ Yes
◉ No

(4) If so, will a data processing agreement be made up?

○ Yes
○ No
◉ Not applicable

(5) Will standard minimum and maximum retention periods apply to the data?

◉ Yes
○ No, namely:

(6) Will data be collected and stored according to FSBS protocol? For the current version of the protocol, please see this page: https://ferb.sites.uu.nl/relevant-documents. Please explain how data collection and storage will be organized.

○ Yes
◉ No

(7) Is secondary use of your data is intended or foreseeable?

○ Yes
◉ No

(8) If so, where will you make your data available?

(9) If so, what access and usage conditions will apply?

Attachments

J.H. Ginn - 20 January 2020 - 14:30:14

(1) Text (advert) for the recruitment of participants (optional)

placeholder.pdf deleted J.H. Ginn 7 November 2019 - 16:25:24

(2) Information letter for participant *(required)*

placeholder.pdf deletedadded J.H. Ginn 20 January 2020 - 14:30:09

(3) Consent form for participants *(required)*

placeholder.pdf deleted J.H. Ginn 7 November 2019 - 16:25:24

(4) Written or oral feedback information (debriefing text) (optional)

No files have been uploaded yet

(5) (Descriptions of) questionnaires (optional)

No files have been uploaded yet

(6) (Descriptions of) measurement instruments (optional)

No files have been uploaded yet

Thank you

J.H. Ginn - 20 January 2020 - 14:30:14

(1) Thank you for completing your project description. Please click the button below to submit your proposal to the FERB and to return to the start page.