Pitfalls and promises of linear regression approaches for modeling affect dynamics: pinpointing challenges and forwarding guidelines and solutions

A Data Management Plan created using DMPonline.be

Creator: Sigert Ariens

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Principal Investigator: Sigert Ariens

Data Manager: Sigert Ariens

Project Administrator: Sigert Ariens

Grant number / URL: PDMT2/23/018

ID: 203469

Start date: 01-10-2023

End date: 01-10-2024

Project abstract:

Affective science has seen a sharp increase in studies that collect intensive longitudinal data (ILD) to investigate how affective processes evolve over time. These designs are typically used to target three types of research question: Researchers are often interested in contextual influences on the affective processes, the dynamics in the affective processes, or how context influences the dynamics. There are some unique challenges associated with addressing these research questions, with important implications for the reliability of modeling results. In this project, I syste matically investigate these challenges and possible solutions in the form of modeling strategies. I hope to communicate to the applied literature which issues to look out for, how the reliability of results can be impacted by failing to approach a challenge adequately, and how researchers can use empirical data to verify the reliability of their results.

Last modified: 22-12-2023

Pitfalls and promises of linear regression approaches for modeling affect dynamics: pinpointing challenges and forwarding guidelines and solutions

Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		Indicate: ***N****(ew data) or* ***E****(xisting data)*	Indicate: **D**(igital) or **P**(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
EMOTE database	The EMOTE (Everyday Measures of Temporal Emotions) Database is an open-access, searchable, and cumulative database of experience sampling data on daily emotional functioning. Experience sampling methods (also known as ecological momentary assessment) involve sampling human experiences in real time (or close to it) across multiple measurement occasions. The EMOTE database contains data from 23 ESM studies which will be used as part of a secondary analysis project	E	D	N		< 100GB	
Consistency data	Data from 1398 participants who participated in an online gambling experiment, recruited through prolific. For reference to the original study, see PPW OKPIV / Researchers / Niels Vanhasbroeck / Consistency of Affective Measures · GitLab (kuleuven.be)	E	D	N		< 100GB	
Preregistrations	The empirical and simulation studies will be preregistered	N	D	T		<1GB	
Simulation data	Data from simulation studies	N	D	N		<100GB	
Analysis syntax	Syntax associated with simulations or empirical analyses	N	D	N		<1GB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

The EMOTE study consists of 26 datasets, 23 of which are suitable for reanalysis in the context of this project. The names associated with the datasets in the EMOTE database are (see https://emotedatabase.com/datasets/):

50 beeps per day

ACU emotions in daily life

Depressive symtoms diary

Emotional Events in daily life

Everyday emotion regulation

Exam results study

FEEL Study 1

Leuven 3-wave longitudinal study

Leuven BPD stud

Leuven clinical study

Leuven couples study 2014

Leuven couples study 2016
Leuven emotion dynamics 2017
Leuven emotion dynamics study
Leuven emotions in daily life 2008
Leuven emotions in daily life 2011
Leuven emotions in daily life 2012
Objectification in women's daily lives Study 1
Objectification in women's daily lives Study 2
Objectification in women's daily lives Study 3
RESS-EMA study
Trier social stress study
William & Mary emotion diary

The consistency data is available on the KU Leuven Gitlab servers. The project has the identifier: affectiveconsistency 1.1.0. The project and data are maintained by Dr. Niels Vanhasbroeck.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- Yes, human subject data (Provide SMEC or EC approval number below)

An ethical application for processing the EMOTE database is currently under review (G-2023-7445). An ethical application for processing the consistency data will be submitted at a later date.

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- Yes (Provide PRET G-number or EC S-number below)

An ethical application for processing the EMOTE database is currently under review (G-2023-7445). An ethical application for processing the consistency data will be submitted at a later date.

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or Data transfer agreements, Research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- Yes

By using data from the EMOTE database, I agree to their terms and conditions. I do not own the data and do not have the right to share the data with third parties, though I may communicate the results of processing the data in scientific publications. I will comply with the data retention policy of the KU Leuven (10 years), this is allowed under the agreement with the EMOTE database.

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, codebook.tsv etc. where this information is recorded).

We will create separate folders for the EMOTE and consistency data. Both datasets already have documentation of the information in the dataset (e.g. variables, which questionnaires were used in the study). For the purposes of the processing in this study, we will create preregistration documents in the same folders, and document the analyses (including data preprocessing) using R-markdown files. Separate folders will be made for simulation studies. These will contain readme documents detailing the structure of the simulation studies, simulation data, and metadata about software and package versions. Simulation data and syntax will be stored in sub-folders of the simulation study folders.

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- No

I do not own the data. The analyses and simulations will be documented in preregistration files, readme files, and use a self-explanatory file structure.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Shared network drive (J-drive)

Digital data will be stored in a restricted network share on the J: drive, which can only be accessed by the involved researchers. All drives are managed by KU Leuven personnel, bound by the KU Leuven general and ICT codes of conduct. All empirical data is pseudonymised. Both the EMOTE database and the KU Leuven have measures in place which ensure that participants can not be re-identified. For the EMOTE database, researchers uploading data must ensure that re-identification is not possible, and in addition the EMOTE team provides random identifiers to participants when data is uploaded. The Consistency data is managed by dr. Niels Vanhasbroeck. I do not have access to the raw data, only the pseudonymised data.

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

All ICT solutions at KU Leuven are subject to the university-wide ICT information security standards. The faculty's ICT service organizes the raw network storage it procures from central ICT services in such a way that access permissions are limited, fixed, delegated to and audited by data managers who do not need to have an IT background.
Digital data will be stored in a restricted network share on the J: drive, which can only be accessed by the involved researchers.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

The yearly cost for network share storage on the J: drive is 800 euro per TB, including backup and unused space. The necessary funding for storage and backup for the contracted service has been foreseen.

Data Preservation after the end of the Research Project

Which data will be retained for 10 years (or longer, in agreement with other retention policies that are applicable) after the end of the project?

In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

- All data will be preserved for 10 years according to KU Leuven RDM policy

Where will these data be archived (stored and curated for the long-term)?

- Other (specify below)

Digital data will be stored at a restricted area of the K: archive drive, which can only be accessed by the involved researchers and the unit's data manager. All drives are managed by KU Leuven personnel, bound by the KU Leuven general and ICT codes of conduct.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

The yearly cost for network share storage on the K: archive drive is 200 euro per TB. The funding for data archiving has been foreseen.

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?
Please explain per dataset or data type which data will be made available.

- No (closed access)
- Yes, as open data

I do not own the empirical datasets and as such can not provide access to the data to researchers not involved in this project. The preregistrations, simulation data, and syntax will be made publicly available on the open science framework.

If access is restricted, please specify who will be able to access the data and under what conditions.

See above

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

Please explain per dataset or data type where appropriate.

- Yes, intellectual property rights

See above

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- Other (specify below)

Simulation data, syntax, and preprints will be made publicly available on the open science framework.

When will the data be made available?

- Other (specify below)

Upon publication of the associated results

Which data usage licenses are you going to provide?

If none, please explain why.

- CC-BY 4.0 (data)
- Other (specify below)

Simulation data, syntax, and preprints will be made publicly available on the open science framework. When doing so, a CC-BY 4.0 license is provided. For the empirical data I do not have the right to provide licenses.

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- Yes, a PID will be added upon deposit in a data repository

The open science framework provies a DOI per project.

What are the expected costs for data sharing? How will these costs be covered?

I do not anticipate costs associated with data sharing.

Responsibilities

Who will manage data documentation and metadata during the research project?

Sigert Ariens

Who will manage data storage and backup during the research project?

Sigert Ariens

Who will manage data preservation and sharing?

Sigert Ariens

Who will update and implement this DMP?

Sigert Ariens