Validation of the Sleep Acceptance Scale

Marwin Carmo, University of California, Davis

2023-12-14

Table of contents

Preface

This page was developed as the final project of Dr. Emorie Beck's course Data Cleaning and Management, taught at University of California, Davis in the Fall quarter of 2023.

Background

This project aims to create a comprehensive pipeline for validating the Sleep Acceptance Scale (SAS), an instrument designed to evaluate people's acceptance of sleep problems. The data for this project come from adults between 18 and 59 living in Brazil who were surveyed online using REDCap.

Here, I do not discuss the results of the analyses in-depth as they are expected to be developed in a future publication in a scientific journal. What will be shown in this project are the analyses' code and outputs that usually do not fit into a traditional scientific publication.

Motivation

There are two main goals that I wish to achieve with this project. First, it is to make the data analysis phase of this research project transparent and accessible to the reader. Further, it is aimed to be a "living" document where the readers can make suggestions or critiques to improve the quality of the work. I also hope that by sharing the code and outputs in an online book format, researchers with limited or no knowledge of R programming can

Although this document was designed to be accessed independently of the formal journal publication, it can also serve as supplementary material to the manuscript. Creating a fully reproducible manuscript would be challenging because the co-authors of this research project do not work with R. With this document, at least the data analysis can be shared in a reproducible format.

Approach

Thanks to a suggestion from Dr. Emorie, this project was redesigned to the format of an HTML Quarto book. This format allows us to easily combine multiple `.qmd` documents into a single manuscript. Each "chapter" in this book corresponds to a different step of the analysis. This way, the scripts can be executed independently or bundled in a single file.

To create this product, I started a new RStudio project using the Quarto book template. Then, different `.qmd` files were created for each step of the analysis and saved in the project's general folder. Additionally, a subfolder for the data was created, containing the raw data, the cleaned data from step 1, and intermediate outputs saved to optimize the rendering of the document.

The project was also made as a Github repository, available publicly. Once finished, the book was hosted as a website on Github Pages, available at https://marwincarmo.github.io/sas/

Reflection

Several layers of concepts taught in the course were used to create this project. Because the ultimate goal was to create a reproducible analysis script, it needed to be constructed using a project-based approach, organizing the files in a hierarchical folder structure. To have the data set ready for the analyses, it had to undergo some steps such as transforming date columns, creating new variables based on existing ones, re-coding, treating missing values, pivoting, manipulating strings, and others. Functions for grouping, selecting and filtering were also used frequently throughout the project.

Although the end product has accomplished its goal, this project has limitations that were not addressed during its development and room for improvement in the future. For instance, it was not pre-registered. Although the current analysis plan has changed in some aspects compared to the initial proposition, documenting these differences would facilitate the inspection of those changes by reviewers and interested readers. Another fundamental limitation is the lack of a data dictionary and a codebook. These files are important to ease the re-use of the data and should be developed shortly. This project also lacks comments on the code chunks. Adding comments will help future re-analyses and is instructive for people interested in using pieces of the code in their work.

This was the first project done aiming to adopt a fully open and reproducible framework. Coding not only for myself but for others who might want to reproduce the same analyses was challenging but rewarding. Working on this project proved to be an exciting opportunity to put into practice many of the ideas discussed in class. It has also helped me build my intuitions in creating projects integrating human and machine readability.

1 Introduction

Insomnia is a disorder related to dissatisfaction with the duration or quality of sleep. It can be a source of distress and impairment by decreasing productivity and lowering energy to engage in social activities (American Psychiatric Association, 2013).

Cognitive and behavioral models of insomnia emphasize the role of sleep-related cognitions as maintainers of insomnia (Espie et al., 2006; Harvey, 2002; Lundh, 2005; Morin et al., 1993; Ong et al., 2012; Perlis et al., 1997). This disorder is linked to a stringent attachment to sleep needs and expectations, which feeds worry, thought suppression, and rumination (Lundh, 2005; Ong et al., 2012). There is compelling evidence that lower levels of psychological flexibility – from which acceptance is a process – are linked to symptoms of depression and anxiety (Bai et al., 2020; Bluett et al., 2014; Ruiz, 2010; Twohig & Levin, 2017). Additionally, psychological inflexibility correlates with higher levels of sleep difficulty even after controlling for the effect of depressive symptoms (Kato, 2016).

The Sleep Problem Acceptance Questionnaire (SPAQ) (Bothelius et al., 2015) is the only measure of acceptance of sleep difficulties developed so far. Because its items were developed with a focus on people with severe sleep problems, this questionnaire has limited application with good sleepers or with those with light to mild sleep problems. Additionally, general tools for measuring acceptance do not address sleep-related behaviors. For such reasons, we developed a new measure of sleep acceptance, titled **Sleep Acceptance Scale (SAS)**.

1.1 Methods

1.1.1 Participants

Participants we primarily found through online advertisements, specifically on the social media accounts of the General Hospital of the University of São Paulo, School of Medicine (HC-FMUSP). The data collection period was from May 2021 to July 2022, and we aimed to have both participants with and without insomnia complaints. The first group consisted of individuals already enrolled in an experimental behavioral treatment for insomnia, which this study is a part of. To include participants without insomnia, we requested volunteers who believed they had no sleeping issues. All the participants were natural from Brazil. To be included in the study, participants must be between 18 and 59 years of age and have no reported difficulties reading or writing in Portuguese–SAS was first developed in Portuguese and then translated do English.

1.1.2 Measures

1. *Sleep Acceptance Scale (SAS)*: Developed as a 6-item self-report questionnaire rated using a scale ranging from 1 (never) to 7 (always), such that high scores indicate lower sleep acceptance. Its items are:
2. *Insomnia Severity Index (ISI)* (Bastien et al., 2001; Morin et al., 2011): An 7-item questionnaire to assess insomnia severity and its impact on the patient's life. Raters use a 5-point scale ranging from 0 (no problem) to 4 (very severe problem).
3. *The Hospital Anxiety and Depression Scale (HADS)* (Zigmond & Snaith, 1983): This scale assesses psychological distress in non-psychiatric patients. It contains a two-factor structure with seven items assessing Anxiety plus seven other items measuring Depression.
4. *Sleep Problem Acceptance Questionnaire (SPAQ)* (Bothelius et al., 2015): An 8-itens instrument that measures the acceptance of sleep difficulties. The SPAQ contains the subscales "Activity Engagement" and "Willingness," with four items composing each factor; respondents rate them on a 7-point scale, where 0 means "Disagree," and 6 indicates "Completely agree."

1.1.3 Statistical Analyses

1. Data screening and item evaluation
2. Assessment of the underlying structure
3. Measurement invariance
4. Reliability and validity

2 Data cleaning

2.1 Loading the required packages

```
library(dplyr)
library(lubridate)
```

2.2 Importing the raw data

First we will import the raw data set downloaded from the REDCap repository where this project is stored. To be able to share this data, we removed sensitive information from participants such as name, telephone number, email address, and CPF (the Brazilian SSN).

```
raw_data <- readr::read_csv("data/raw_data/TerapiaDeAceitaoECom_DATA_2023-08-16_1318.csv")
head(raw_data)
```

```
# A tibble: 6 x 588
  record_id redcap_event_name   redcap_survey_identifier identificao_timestamp
      <dbl> <chr>               <lgl>                    <chr>
1         1 recrutamento_arm_1  NA                       3/15/2021 21:24
2         1 elegibilidade_arm_1 NA                       <NA>
3         1 reteste_arm_1       NA                       <NA>
4         2 recrutamento_arm_1  NA                       3/15/2021 23:45
5         2 elegibilidade_arm_1 NA                       <NA>
6         2 reteste_arm_1       NA                       <NA>
# i 584 more variables: data_preenchimento <chr>, data_nascimento <chr>,
#   idade <dbl>, latitude <dbl>, longitude <dbl>, forma_divulgacao <dbl>,
#   identificao_complete <dbl>, consentimento_tcle_timestamp <chr>,
#   consent_sign <chr>, consentimento_tcle_complete <dbl>,
#   recrutamento_e_seleo_4b68e8_timestamp <chr>, dsm_1 <dbl>, dsm_2 <dbl>,
#   dsm_3 <dbl>, dsm_4 <dbl>, dsm_5 <dbl>, triagem_disponibilidade_1 <dbl>,
#   triagem_disponibilidade_2 <dbl>, triagem_disponibilidade_3 <dbl>, ...
```

2.3 Data transformation and selection

This dataset contains 588 variables and a few of them are relevant to the analyses in this project. Therefore, we will transform some of the variables of interest and then select those that we need. We want to have a clean dataset containing the columns of age, event, sex, race, marital status, education, work situation, location, sleep habits, insomnia severity index, anxiety, depression, dysfunctional beliefs and attitudes about sleep, sleep problem acceptance and sleep acceptance.

```
clean_data <- raw_data |>
  dplyr::with_groups(record_id,
                     tidyr::fill,
                     c(idade, dsm_1:dsm_5), .direction = "down") |>
  dplyr::mutate(identificao_timestamp = dplyr::na_if(identificao_timestamp, "[not complete
                identificao_timestamp = lubridate::mdy_hm(identificao_timestamp),
                data_nascimento = lubridate::mdy(data_nascimento),
                idade_new = lubridate::time_length(difftime(identificao_timestamp, data_na
                age = dplyr::coalesce(idade, idade_new)) |>
  dplyr::select(record_id, redcap_event_name, sexo, age, etnia, estado_civil, escolaridade
                dplyr::starts_with("dsm_"), dplyr::starts_with("igi_"), dplyr::starts_with
                dplyr::matches("aaq_[0-9]"), dplyr::matches("dbas_[0-9]+$"), dplyr::contai
                dplyr::starts_with("ebas_")) |>
  dplyr::filter(redcap_event_name %in% c("elegibilidade_arm_1", "reteste_arm_1"),
                age >= 18 & age < 60,
                !record_id %in% c(2562, 1766, 2972, 2681, 430, 2559, 3053, 540 , 522 , 369
                )
```

2.4 Translation of column names

This project was developed in Portuguese and the variables were named in the same language. To make them accessible to non-Portuguese-speaking researchers, we should rename the columns translating them to English.

```
namesEN <- c("record_id", "redcap_event_name", "sex", "age", "race", "marital_status", "ed

names(clean_data) <- namesEN
```

2.5 Saving the clean dataset

```
readr::write_csv(clean_data, "data/clean_data/clean_data.csv")
```

3 Data screening

In this phase we examined response frequency and item statistics to assess item variation, distribution, and data entry. We also investigated inter-item correlations and searched for unusual response patterns by identifying multivariate outliers using Mahalanobis distance.

3.1 Loading the required packages and data

```
library(kableExtra)
library(dplyr)
library(ggplot2)
library(modelsummary)
library(lavaan)
library(corrplot)
```

```
mydata <- readr::read_csv("data/clean_data/clean_data.csv")

arm1 <- mydata |>
  dplyr::filter(redcap_event_name == "elegibilidade_arm_1",
                !dplyr::if_all(dplyr::starts_with("sas_"), ~ is.na(.))) |>
  dplyr::select(record_id, paste0("sas_", 1:6))
```

Sample size: 1345.

3.2 Descriptives

3.2.1 Response frequecy

```
sas_transposed <- arm1[,-1] |>
  tidyr::pivot_longer(cols = dplyr::everything(), names_to = "item", values_to = "score") |>
  dplyr::mutate(score = factor(score, levels=1:7),
                item = factor(item, levels=paste0("sas_", 1:6)))
```