Analysing the Effect of Age and Narrative Identity on Well-being

Dr Mark Andrews, Associate Professor of Statistical Methods, Nottingham Trent University, UK

The dataset

In this problem set, our case study is based on research by Lind et al. (2024), published in the journal *Psychological Science* in 2024, which was about the effects of age, personality traits, and narrative identity on depression and well-being in middle-aged people. In the dataset that we will use here, we will use a subset of their original data and focus specifically on the effect of age and narrative identity on well-being. In general, narrative identity refers to the internalized and evolving stories that people construct about their lives to make meaning of themselves and their experiences over time. In Lind et al. (2024), they were particularly focused on narrative identity that related to the fulfillment of motivational themes, specifically what they refer to as agency motivations (motivations toward empowerment and self-determination) and communion motivations (motivations toward connection with others). In the data we will use, each participant has a measure of well-being each year over nine years. They also have a score measuring the average extent, over the nine years, to which they had narrative identities involving fulfilled agency or communion-based motivations. The ultimate focus of our analysis will be to analyse how the participants' age and narrative identity score predict their well-being over time.

Activities

Using a data set named `ylc.csv`, which is a subset of the original data collected by Lind et al. (2024), we want to address the following activities:

1. Load up the data from the file `ylc.csv` and look at the data frame to see how many columns (variables), how many rows (observations), and try to assess what each of the variables mean.
2. Calculate summary statistics for all the variables, and in particular, look closely at the variables describing the demographic characteristics of the sample.
3. Visualize the main trends in the data. Specifically,
    1. How does well-being vary by age?
    2. How does well-being vary by narrative identity?
    3. How does well-being vary simultaneously by changes in age and narrative identity?
4. Conduct statistical analyses to test the following hypotheses:
    1. Does well-being vary by age?
    2. Does well-being vary by narrative identity?

3. Is there an interaction between age and narrative identity? In particular, use the following two separate statistical methods to address these hypotheses:
4. Repeated measures ANOVA
5. Linear mixed effect regression

Install and load packages

Before we begin, there are a number of R packages that we need to do this analysis:

- `tidyverse`: A widely used collection of R packages for efficient general data processing in R.
- `skimr`: A package for providing quick statistical summaries of data frames
- `afex`: A package for doing factorial ANOVA, including repeated measures ANOVA.
- `lmerTest`: A package for doing hypothesis testing in linear and generalized linear mixed effects models. The `lmerTest` is dependent on, and will therefore pre-install, the `lme4` packages, which is the most widely used package for linear and generalized linear mixed effects modelling in R.
- `emmeans`: A package that provides easy framework for post-hoc comparisons and related inferences from various regression and ANOVA models.

You can test whether any given package is installed a few different ways. For example, the following code tests if I have `tidyverse` installed in my R session, which it is:

```
# change "tidyverse" to any other package name to
# test if that package is installed
"tidyverse" %in% rownames(installed.packages())
```

```
[1] TRUE
```

If a package is not installed, there any different ways you can install it. For example, the following command will install`tidyverse`:

```
install.packages('tidyverse')
```

If you want to install multiple packages at the same time, use the `c()` (short for “combine”) function to combine all the names of all packages you want to install into a vector of names. For example, this command installs `tidyverse`, `skimr`, `afex`, `lmerTest`, `emmeans`:

```
install.packages(c('tidyverse', 'skimr', 'afex', 'lmerTest', 'emmeans'))
```

After you have installed the required packages, they should be loaded into your R session using the `library()` command as follows:

```
library(tidyverse) # this loads 9 packages
library(skimr)
library(afex)
library(lmerTest)
library(emmeans)
```

Appendix

The following code shows how the `ylc.csv` data file was obtained from processing the data accompanying Lind et al. (2024), which is available at https://osf.io/zqe4x.

```
# original raw data
# from https://osf.io/zqe4x/files/osfstorage
ylc <- read_csv("YLC processed data R2 OSF 7.16.24.csv")

ylc_id <- ylc |>
  # only participants from whom there is narrative identity information
  drop_na(ac_val) |>
  # and only participants from whom we have more than data point (data from
more than one year)
  add_count(ID) |>
  filter(n > 1) |>
  select(ID) |>
  distinct()

# get demographic variables
ylc_demographics <-
  ylc |>
  filter(ID %in% ylc_id$ID, year == 1) |>
  select(ID,
    ethnicity = `race/ethnicity`,
    sex = `t1: sex of respondent`,
    dob = `t1: date of birth (dd-mmm-year)`,
    interview_date = `date completed interview (dd-mmm-year)`,
    education = `highest level of education completed`
  ) |>
  mutate(
    age = time_length(interval(dob, interview_date), "years"),
    sex = if_else(sex == 0, "male", "female"),
    sex = factor(sex, levels = c("male", "female")),
    education = case_when(
      education == 1 ~ "high-school",
      education == 2 ~ "some-undergrad",
      education == 3 ~ "completed-undergrad",
      education == 4 ~ "some-postgrad"
    ),
    education = factor(education, levels = c("high-school", "some-undergrad",
"completed-undergrad", "some-postgrad")),
    ethnicity = case_when(
      ethnicity == 1 ~ "white",
      ethnicity == 2 ~ "black",
      ethnicity == 5 ~ "inter-racial",
      ethnicity == 6 ~ "other"
    ),
    ethnicity = factor(ethnicity, levels = c("white", "black", "inter-
```

```
racial", "other")),
  ) |>
  select(ID, sex, age, ethnicity, education)

# join to original
ylc2 <- ylc_demographics |>
  left_join(select(ylc, ID, wellbeing = wb_comp, year, ac_value =
ac_val_pmean), by = "ID") |>
  relocate(year, .before = wellbeing) |>
  arrange(ID, year)


saveCSV(ylc2, file = 'ylc.csv')
```

References

Lind, Majse, Sebnem Ture, Dan P McAdams, and Henry R Cowan. 2024. “Narrative Identity, Traits, and Trajectories of Depression and Well-Being: A 9-Year Longitudinal Study.” *Psychological Science* 35 (12): 1325–39.