Dear all,

As there have been quite a few questions about how to perform the required analyses, following a few tips. Furthermore, there might have been some confusion due to already given suggestion.

Please consider this “white paper” as the actual and “right” suggestion for your data analysis!!!

Procedure for data analysis:
1. identify your dependent variable
2. identify your independent variables
3. identify what type of data each is (e.g. binary/unordered factor, ordinal scale, continuous).

4. identify the design (between parcitipants or within-pariticiants/repeated-measures)

Once you have done so, it makes things a lot easier to model.

For example, if I was analysing the following situations (real datasets may have a combination of all of these situations):

Situation 1 (causal effect)

Predicting *weight* from *height* and *country.*
<u>Variables</u>
dependent variable = weight (continuous)
independent variables = height (continuous), country (unordered factor)

tibble_1

Subject_Id *factor*	**Weight** *numeric*	**Height** *numeric*	**Program** *factor*
1	43	115	Germany
2	65	180	Austria
3	97	145	Australia
…	…	…	…

Once I made sure the columns in my data frame were of the correct types, I could then use brms like so:

<u>Model:</u>
model1 <-brm(weight ~ height * university_program, data = tibble_1, family = "gaussian")

I use gaussian here because the dependent variable is continuous.

Situation 2 (group comparison)
income from graduate of program x differs from *income of graduates from program y*

<u>Variables</u>
dependent variable = income (continuous)
group variable = university program (unordered factor)

tibble_2

Subject_Id	**income**	**Program**

factor	*numeric*	*factor*
1	400	Germany
2	600	Austria
3	400	Australia
…	…	…

Once I made sure the columns in my data frame were of the correct types, I could then use brms like so:

Model:
model2 <- brm(income ~ Program, data = tibble_2, family = "gaussian")

Situation 3 (dependent variable is binary or ordinal)

- if the dependent variable is either **binary or ordinal**, I would need to specify a different 'family'. "bernoulli" for binary data, and "cumulative" for ordinal data.

- If one of the predictors should be treated as ordinal (e.g. Likert-scales), I can specify that predictor as 'monotonic' using mo(predictor) in the formula.

Graduates rate on “liking” of degree and “happiness” (7-point-scale).

We assume that liking affects happiness: Happiness ← Liking

tibble_3

Subject_Id	**Liking**	**Happiness**	**…**
1	4	3	…
2	6	3	…
3	4	2	…
…	…	…	…

Our model would be:

model3 <- brm(formula= Happiness ~ mo(Liking), data = tibble_3, family= “cumulative”)

Situation 4 (repeated-measures)

4.1.
If there are multiple rows per person (e.g. the same question was asked at multiple times in people’s lives), I would need to add a by-participant random intercept in order to control for individual differences.

E.g. the first situation as a repeated-measures design:

Model4 <- brm(weight ~ height * country + (1|subject_id), family = “gaussian”)

4.2.
If you would like to compare between two conditions whereas the participants took part in both conditions (equivalent to a dependent t-test) then you would also need the by-participant random intercept

Rating on a Likert-scale for both conditions.

Model5 <- brm(Rating ~ condition + (1|subject_id), family = “cumulative”)