Modeling Repeated Measures on Product Innovation

Alex Winchester (wincheser@email.arizona.edu)
Zhuo Chen (zchen1@email.arizona.edu)
Elias Ramirez-Moreno (eliasrm22@email.arizona.edu)

1 Executive Summary

After examining existing literature on mediation analysis with repeated measures, we have identified two options to proceed with analysis.

The first option would require reducing the problem to comparing only two products (or groups of products). The MLMED and MEMORE SPSS packages have support for this problem when there are only two groups for comparison. This would allow for a "proper" mediation analysis where the total, direct and indirect effects can be recovered. This report includes example code for MLMED.

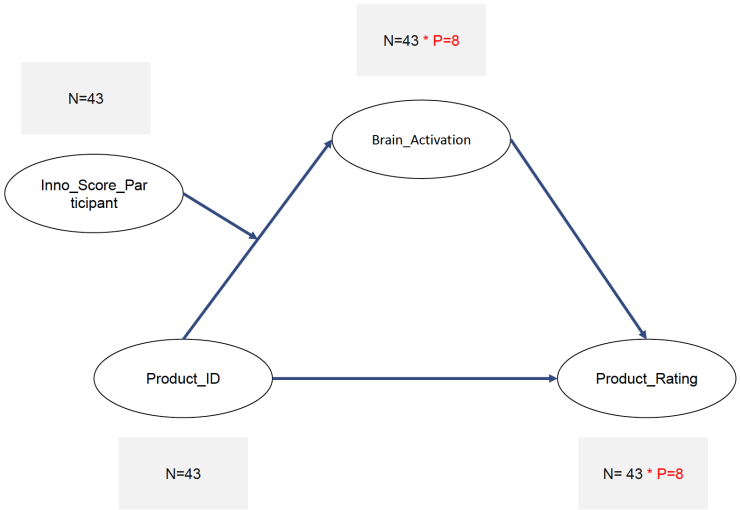
The second option is to model the brain activation outcome and the product rating outcome separately. While we lose the ability to separate the direct and indirect effects, we would be able to include all 8 of the products in the analysis. This can be modeled using standard mixed models implementations. This report will demonstrate using lme4 in R.

2 Background Information

An experiment was designed and data collected to determine if the activation of the anterior insula is related to how innovative a person views a product. 43 subjects were shown 8 objects and asked to fill out a survey.

This survey asked participants to rate the innovativeness of the products on a sliding scale (several questions were combined to form one score), included demographic questions, and asked participants to self-report their innovativeness on a sliding scale.

Next, participants were shown images of the products inside an fMRI machine and the activation of the anterior insula was measured. The hypothesized model is



3 Modeling Considerations

In order to choose an appropriate model to fit, it is worthwhile to examine the structure of the data. Each participant self-reports their innovativeness, this does not vary between products, but does vary between participants. The brain activation and product rating would be expected to vary both between participants and between objects.

Because of this structure, we think of the products as repeated measures on individuals. In the notation of multilevel mediation models:

- The level 1 unit is the product (the repeated measure on individuals)
- The level 2 unit is the individual (the unit on which we apply repeated measures)
- The predictor is the product ID and is categorical and level 1 (varies at product level)
- A continuous level 2 covariate needs to be included (self-reported innovation score, varies at the individual level)
- Brain activation is a continuous level 1 mediator (varies at the product level)
- The outcome is product rating and is continuous and level 1 (varies at the product level)

This design is called 1-1-1 (perhaps 2-1-1 because of the additional covariate), and is well-studied. The complication is being brought in by the fact that the predictor (the object shown to participants) is categorical, but not binary. Existing software implementations do not seem to support this special case.

4 Suggested Approaches

4.1 Binary Multilevel Mediation Analysis

The first option to modeling these data is to reduce the number of products (or groups of products) to compare down to 2. Both the MLMED and MEMORE SPSS packages are capable of performing analysis in this case. Some additional subtleties are needed when using MLMED:

- The binary predictor (or product ID) will need to be coded as -1, 1. If the number of products are the same for each participant, this will group-mean-center the product_ID variable.
- The covariate (self-reported innovation score) and mediator (brain activation) may need to be group-mean-centered prior to analysis. This can either be done by hand, or using bmlm in R. The MLMED documentation claims that this is done under the hood automatically, but it is unclear to me if this is true. Group-mean-centering prior to analysis should not be problematic in either case.
- This 2020 paper by Xiao Hu[1] uses the MLMED package on data with similar structure to this. For group-mean-centering, they used the bmlm package in R.

I expect the correct code for the most basic model to look something like this:

```
MLmed data = <data_name>
  /x = Product_ID
  /m1 = Brain_Activation
  /y = Product_Rating
  /cluster = Participant_ID
  /L2cov1 = Inno_Score_Participant
  /folder = <file_path>
```

By default, all intercepts in the model will be treated as random and slopes treated as fixed. I think this is a good place to start model fitting. The argument `randYint=0` and `randMint=0` can be used to remove the random intercepts. Specifying random slopes is more complex. The argument for the predictor is `randx` and you must specify a list of binaries where each binary refers to a path in the model. 1 indicates that the slope should be random on this path and 0 indicates fixed.

[1] `https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00637/full`

4.2 Separate Mixed Effects Models

If instead of mediation analysis we use separate mixed models, we can model all 8 products. I would suggest fitting the following 3 mixed effects models:

1. Brain Activation as a function of Product ID and Participant Innovation Score:

   For this model, I think that the best place to start would be to treat everything in the model as fixed except for random intercepts for each participant. Using the package lme4[2] in R, this looks like

   ```
   data$Participant_ID = as.factor(data$Participant_ID)
   data$Product_ID = as.factor(data$Product_ID)
   library(lme4)
   fit <- lmer(Brain_Activation ~  Product_ID + Inno_Score_Participant
               + (1|Participant_ID), data=data)
   summary(fit)
   ```

2. Product Rating as a function of Brain Activation:

   Next, we should model product rating using only Brain Activation, with random intercepts for each participant. The R code would be:

   ```
   fit <- lmer(Product_Rating ~ Brain_Activation + (1|Participant_ID), data=data)
   summary(fit)
   ```

3. Product Rating as a function of Brain Activation and Product ID:

   Finally, a full model including both Brain Activation and Product ID, with random intercepts for each participant.

   ```
   fit <- lmer(Product_Rating ~ Brain_Activation + Product_ID
               + (1|Participant_ID), data=data)
   summary(fit)
   ```

By fitting these three models, we first model the relationship between the predictors (Product ID and Participant Innovation Scores) and the mediator (Brain Activation). The next two models should be compared to determine if the mediator alone is sufficient to explain the variation in Product Rating, or if the addition of Product ID improves the model.

5 Conclusion

The main problem preventing existing software implementations of multilevel mediation analysis to be used on these data was identified: Product ID is categorical, but not binary. To proceed with modeling and analysis, we recommend either reducing the number of products to 2 and using the MLMED package in SPSS or using separate mixed effects models.

Both approaches give slightly different perspectives on the data. Reducing the number of products will give greater insight into causal knowledge by computing direct and indirect effects. Using separate mixed effects models, we do not get the same level of causal knowledge, but can examine the effects of a greater number of objects.

[2] https://cran.r-project.org/web/packages/lme4/index.html