fdsfda

Rong Guang

2023-02-13

Contents

0.1 R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

0.2 Including Plots

You can also embed plots, for example:
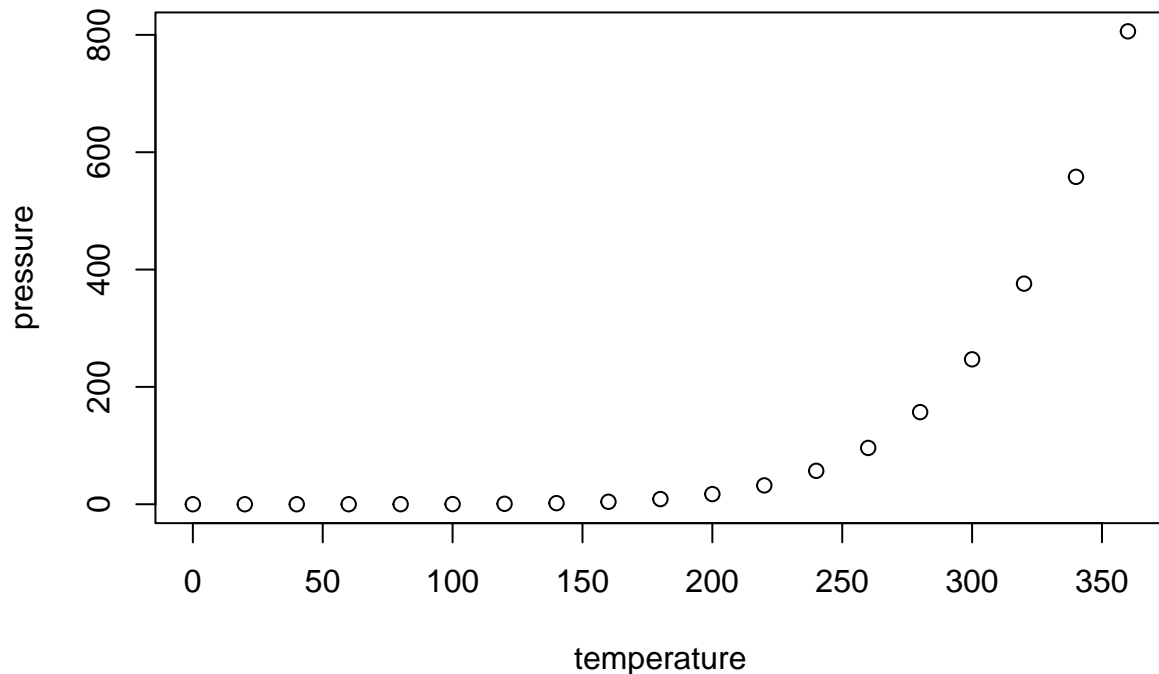
Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

dfsdajklfd

dfdasasffads

dfadsfdfsa

0.3 I will explore two intertwined questions:

If we keep doing post-hoc analyses step by step for 20 rounds (20 models), how would the model fits change over these rounds?

If we keep doing the above post-hoc analyses in a judicious sense-making manner, will the model fits really out-perform those by sheer statistical procedure?

0.4 To answer these questions, the plan is:

The goodness of model fit will be measured by fit indices (and other indicators) from a test data set (in contrast to a train data set. 50/50 splits of MBI data);

The error of model fit will be captured by a procedure bearing similarity with bootstrapping. Specifically, the train/test will be randomly split for 60 times. For each train/test set, the SEM model will be estimated for 20 modifications, respectively. However, the regression path to set free in each step for both train and test datasets will be determined singly by the MI of train set.

0.5 This will be done in an automated way:

All the procedure above will be executed twice. They are same in every aspect except for the procedures by which the regression paths to be set free are selected. One selects paths by sheer statistical numbers (Always select the path with largest MI); the other selects the paths in a rule-based way. See next paragraph.

The rule-based path selection will also be automated. Yet it will, to my best efforts, approximate the implicit logics that Byrne used for the example in our exercise. They are:

(1)each candidate regression path will be examined in two aspects: a. Logics and b. study focus.

(2)To measure logics, all the 11 factors will be assigned as positive or negative perception according to what Byrne had explicitly defined in the chapter. The reasoning is a.When a factor leads to another factor with the same positive/negative direction with it, the EPC should >0; b. When a factor leads to another factor with different positive/negative direction with it, the EPC should <0. All the parameters obey these rules will be seen as logical; all the parameters violate these rules will be taken as illogical.

(3)Based on Byrne, a regression path that satisfied the study focus could be defined as a logical path in which the dependent variable is one of the three indicators from BMI inventory (EE, DP, PA).

(4) In the automation, R first loop through all the rows of MI table to find regression path that are both logical and of study focus. The first parameter that meets the criteria will be chosen for modification; If no one parameter is found to be competent until the last row, R will loop through all the rows again for the parameter that is logical (but couldn't be of study focus), and the first one that meets this criterion will be chosen for modification.

(5) Note that the sheer statistical selecting procedure also had a very minimal logic that excludes all the paths that lead from the BMI indicators to other indicators outside BMI.