Vitamin E and IQ - exploratory study

Jana Hozzová

Task

This document will guide you through the process of exploratory analysis. Although it is much less (or not at all) restricted compared to the confirmatory study, it is still essential to write everything down. It's a part of exercise in Experiments seminar dedicated to quantitative evaluation of experiments. We will work with made-up example of influence of vitamin E on IQ.

First, answer the questions. Obviously, you will come up with more hypotheses you want to examine and analyses you want to do and conditions you want to look into as you will explore the data. But write down your starting points and then also everything you have come up with. Some options are pre-filled for you, in others let your imagination go wild.

When you have your initial ideas written down, examine the folder with data files. There is one data file for the control group and multiple numbered data files for treatment groups. Start with the treatment data file with the number of month you were born in (e.g. 6 for June).

Please explicitly note that you are doing exploratory analysis. They play around and do your analyses. Finally, write down all hypotheses that seem reasonable given your data and you want to examine further in confirmatory study.

Text in italics are instructions and examples, normal text is what is pre-filled for you.

This document is written in R markdown, for details see [here] (https://blog.rstudio.com/2014/08/01/the-r-markdown-cheat-sheet/).

Initial ideas

What are the questions being asked?

Decide for yourself. Some examples of hypotheses:

- *Vitamin E influences IQ.*
- *Vitamin E does not influence IQ.*
- *Vitamin E increases IQ.*
- *Vitamin E does not increase IQ.*
- *Vitamin E increases IQ by at least two IQ points.*
- *Vitamin E decreases IQ by at least two IQ points.*
- *Vitamin E increases IQ, but less than two points.*
- *Vitamin E increases IQ by five points exactly.*

Describe the key dependent variable(s) specifying how they will be measured.

Dependent variable is IQ, measured by a Wechsler Adult Intelligence Scale test in controlled environment.

How many and which conditions will participants be assigned to?

Choose one of two following options or think up your own:

Two conditions: control group and treatment group (100% of recommended daily intake take in form of a daily pill from company GoodVitamins).

Three conditions: control group, treatment group with low doses of vitamin E (100% of recommended daily intake taken in form of a daily pill from company GoodVitamins for 30 days) and treatment group with high doses of vitamin E (300% of recommended daily intake, taken in form of three daily pills from GoodVitamins for 30 days).

Specify exactly which analyses you will conduct to examine the main question/hypothesis.

You are welcome to use Welch's t-test for basic testing, but feel free to use something else if you find it more suitable to test your hypothesis. In any case specify the statistical test you want to use and its parameters, e.g. direction of effect, α levels, etc.

Any secondary hypotheses and analyses?

Go for it, come up with another hypothesis that you would find interesting to examine. For example:

- *Vitamin E influences IQ more with higher dose.*
- *Vitamin E does not influence IQ with lower dose, but it does with higher dose.*

How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

Due to funding restrictions, we are able to obtain only 100 participants per group in our exploratory study.

Will you use sequential analysis (gradual acquisition of data) or test the data in multiple ways?

Test the data multiple times in some way. As we are doing exploratory study, we are not worried about Type 1 error rate, any hypotheses we come up with must be confirmed later anyway. Some examples:

- *You can imagine you are getting the data gradually, e.g. because it takes long to complete the process with participants. So take only the first 50 participants per group and test them. And only after that take the other 50 and do the test with 100 datapoints per group.*
- *Consider the effect of some confounding variables in your analysis. Divide both groups in two parts according to the following: age (below and above 40), socio-economic status (income below 150000 Kč per year per person in household and income above that), achieved education (university degree of some sort or not). For each division, compare corresponding control and treatment groups (i.e. control below 40 with treatment below 40, control above 40 and treatment above 40). You can take a dataset with a different number for different value of confounding variable.*

Analysis

Set the sample size and load data.

```
samplesize <- 100
# for reading csv file, use read.csv function
#   type ?read.csv into console for help
control
treatment
```

Do the statistical tests.

```
# for t-test, you don't need any special library,
#     just type ?t.test into console for help
# for testing "bounded" hypotheses, i.e. vitamin E increases IQ by more than two points,
#     you can use package TOSTER
# for Bonferroni correction, just divide your alpha
#     a_i <- alpha/number_of_tests and use a_i in tests
# for FDR, there is library(astsa) and its function FDR,
#     look into real-world example notebook for details
# for other tests and corrections,
#     use google, R has packages for almost everything
```

Hm, I wonder if . . . *Fill in your next hypothesis building on what you see in data.*

Do the other analyses.

```
# just do what you promised in pre-registration
```

Conclusions

What would be your next steps:

- *hypotheses you have come up with*
- *estimated effect size we can see in your data*
- *tests to conduct*
- *inspect more dependent variables*
- *include and collect more independent variables*
- *modifications to the future study setup (e.g. including possible confounding variables)*

And now you can go and do the confirmatory study.