TMA4267: Design of Experiments

Sara Elise Wøllo (477759)
Aurora Hofman (999947)
Camilla Idina Elvebakken (477809)

October 3, 2019

Introduction

In this project we want to study what affects our ability to remember. As students, being able to learn information quickly and keep the information stored is a vital part of our daily life. Knowing what affects our memory could therefore be beneficial.

There are many variables that affects the human brain, and as a consequence, its ability to remember information. There are already many studies on the connection between for example sleep and memory. What kind of environment is best for learning is especially interesting in our university setting. In this experiment we want to check to what degree certain factors affects our ability to memorize random numbers. This is tested by letting a test person memorize random numbers between 1-99.

Selection of the factors and levels

We have chosen to do an 2^3 factorial design experiment. This means that we have chosen to test three independent factors, each with two different levels. The three factors we are testing are study time, waiting time and memorization aids.

Study time is the amount of time the test person has to look at the numbers. The two levels chosen for this factor are 1 minute vs. 3 minutes. We made sure the levels differed quite a lot (200% increase in time) to be able to see a difference.

Waiting time is the time from the test person stops looking at the numbers until it has to reproduce as many as it can. During this waiting time we will talk to the test person to try to distract them. For waiting time, one level is no waiting, 0 minutes, and the other is 3 minutes. The interest in this factor is seeing weather or not it matters with a distraction period the same length as the actual study time. We also found some documentation that stated that the short term memory only lasts for about 30 seconds [1]. We therefore chose a waiting time longer than this.

Memorization aid involves being able to use pen and paper to maybe memorize the numbers better during the study time. The first level is no aids, and the second is with pen and paper.

Since our factors either have a time dependent level or a "with-or-without" level, it is fairly easy to control that the factors are on the desired levels. The motivation and explanation behind the choice of factors where many. We used the same person during the entire experiment to avoid that different peoples ability to memorize would affect the results. We did however think that the

time allowed for the person to memorize the numbers were a factor to consider. In classes, many students also write down notes even if the lecture notes are available to study after class. This is because many believe that we learn better when writing things down. We therefore also wanted to consider the factor of a "cheat sheet", where you could take notes while studying the numbers. Lastly, if you learn a number or another fact, you tend to forget it after sometime if you don't continue practicing. We therefore also wanted to test the factor of waiting time. Does it matter if you firstly memorize, then you are distracted by outside sources, and then you have to reproduce the numbers?

We do not expect an interaction between any of these factors, as they are focused on three different aspects of remembering data. We do expect that when certain levels of these factors occur at once, the result may be that the ability to remember the numbers are better or worse than others. We don't expect this to be because of interaction between the factors, but more that one of the levels of the different factors are expected to give better result than the other, and when more of these appear at the same time, we expect to see it in the results.

Selection of the response variable

The response variable we have chosen is the amount of numbers remembered by the test person. We gave the test person 1 minute to write down all the numbers they could remember. We did not see another natural response variable, as this is a measurable and easy method of knowing how many of the numbers were remembered correctly. We counted the amount of numbers correctly remembered from the sequence of random numbers. The response variable is discrete, as you are not allowed to remember half numbers. However since there is quite a lot of variety in the amount of numbers remembered we count this as semi continuous and use linear modelling. The measurements are accurate, as they are easy to count and collect.

Choice of design

As stated before, we chose a 2^3 factorial design for the experiment. This was done because we wanted to test three different factors, when they were both in their "high" and "low" levels. We did the experiment in duplicate to get more variance in our results and to avoid over-fitting when developing a regression model that could later be used for prediction. We did not find it necessary to implement blocking in our experiment design. As we ran our experiments fairly quickly, such that nothing about our surroundings changed significantly, and we used the same test person and made sure she was focused and ready for each run, we did not see it necessary to arrange our experimental units into similar blocks. We did however do each run of the experiment in a random order. The use for a fractional factorial design becomes evident when the number of factor in the design is large or the resources needed to preform the experiments are limited, but as we encountered neither of these approaches were not necessary for our experiment.

Implementation of the experiment

We used Rs FrF2 to randomize the order of our experiments. The experiment was preformed in one sitting. The conditions where equal during the entire experiment.

One potential problem with doing all the experiments in one go is that the test person could get tired and therefore remember worse at the end than at the beginning. On the other hand, the test person could in fact get better at remembering simply by having done it more often.

In each experiment we used a new sequence of numbers generated by Python's random function. We did not allow multiple equal numbers in one sequence. The order in which the numbers were remembered did not matter in our experiment, only the amount of correct numbers remembered.

This means that one could argue that it is not a genuine replicate of the experiment since the sequence of numbers can sometimes by coincident be easier. For example if there is a pattern within the numbers, or the amount of one-digit numbers are large. This however we evaluated to be the best way to replicate our experiment as genuine as possible. Each experiment was done in the same manner.

Table 1: Table of the experimental data of the experiment.

Experimental data							
Study time	Waiting time	Aids	SW	SA	WA	SWA	Response
1	-1	1	-1	1	-1	-1	18
-1	1	1	-1	-1	1	-1	9
-1	-1	1	1	-1	-1	1	11
-1	1	-1	-1	1	-1	1	7
1	-1	-1	-1	-1	1	1	16
1	1	1	1	1	1	1	15
1	1	-1	1	-1	-1	-1	17
-1	-1	-1	1	1	1	-1	8
-1	1	1	-1	-1	1	-1	8
-1	-1	-1	1	1	1	-1	11
-1	-1	1	1	-1	-1	1	7
1	1	1	1	1	1	1	16
1	-1	1	-1	1	-1	-1	15
-1	1	-1	-1	1	-1	1	6
1	1	-1	1	-1	-1	-1	4
1	-1	-1	-1	-1	1	1	16

The result of the experiment can be seen in table 1.

Analysis of the data

Firstly, we can see from our interaction plot in figure 2, that the factors study time and waiting time are fairly parallel, the same with factors study time and aids. That indicates that there is no interaction between the factors. We do however see that in the plots with waiting time and aids, the lines are not parallel. That indicates that there is an interaction between the factors. However when looking at the summary 1 one can see that this interaction is not significant.

```
Call:
lm.default(formula = respons ~ .^3, data = df)

Residuals:
   Min     1Q Median     3Q    Max
 -6.50  -0.75   0.00   0.75   6.50

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     11.500      0.897  12.820 1.29e-06 ***
tid              3.125      0.897   3.484  0.00827 **
vent            -1.250      0.897  -1.393  0.20097
hjelp            0.875      0.897   0.975  0.35792
tid:vent        -0.375      0.897  -0.418  0.68691
tid:hjelp        0.500      0.897   0.557  0.59251
vent:hjelp       0.875      0.897   0.975  0.35792
tid:vent:hjelp   0.250      0.897   0.279  0.78755
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.588 on 8 degrees of freedom
Multiple R-squared:  0.6741,    Adjusted R-squared:  0.3888
F-statistic: 2.363 on 7 and 8 DF,  p-value: 0.1257
```

Figure 1: Summary from R from linear model with three factors. The summary shows which factors are significant under the null hypothesis that all estimates are zero. Tid = study time, vent = waiting time and hjelp = aids.

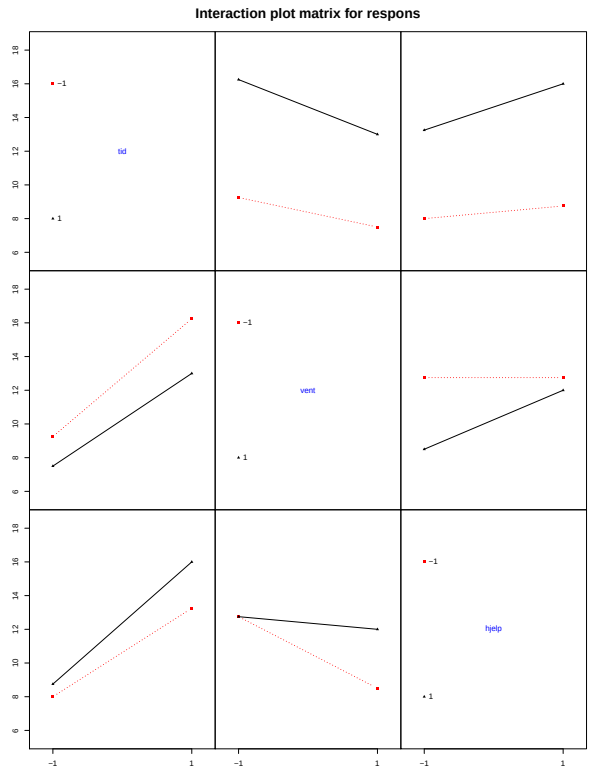

Figure 2: Here the interaction between the different factor are shown. When the lines are parallel or almost parallel there is no interaction

Conclusion and recommendations

When looking at the main effect plot in figure 3, we see a clear difference between the time-factor and the others. It is significantly steeper. The other two factors are not parallel to the x-axis either, which indicates that there was a difference in the results between the factor's two levels, but that it is not as prominent at the difference in the time-factor between the two levels.

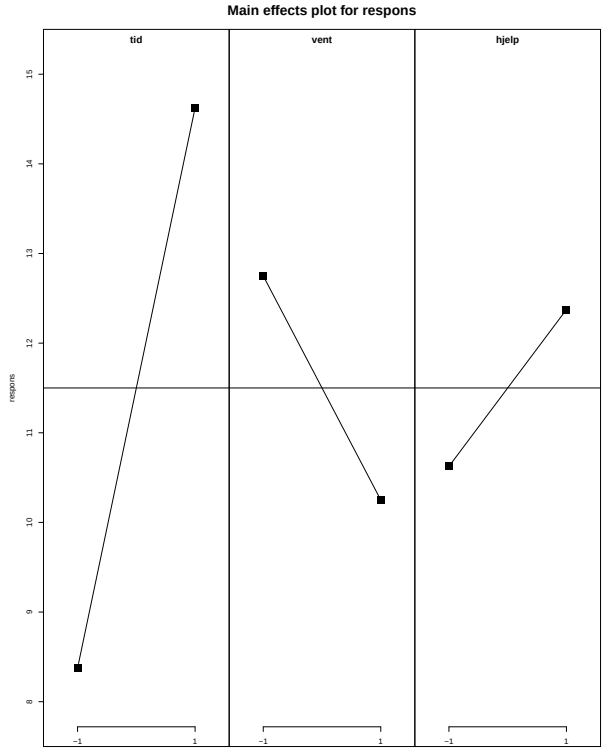

Figure 3: The figure shows the main effects of each factors. Meaning $\sum \beta_{factor_i} \cdot response\frac{1}{n}$ or in other words the difference between when a factor is in the high level and the low level. Tid = study time, vent = waiting time and hjelp = aids.

When looking at our QQ-plot in figure 4, we see that there is a clear linearity in the points if one do not look at the ends. These are a bit off. However when performing the Anderson Darling test we get a p-value = 0.2093 so the null hypotheses that our data is normal distributed can at least not be rejected. We also see from our residual plot in figure 5, that there is no clear pattern in the points other than them being symmetric around the x axis. This might be because we had to responses for each combination of factors. From this, it is reasonable to assume that the error has a normal distribution.

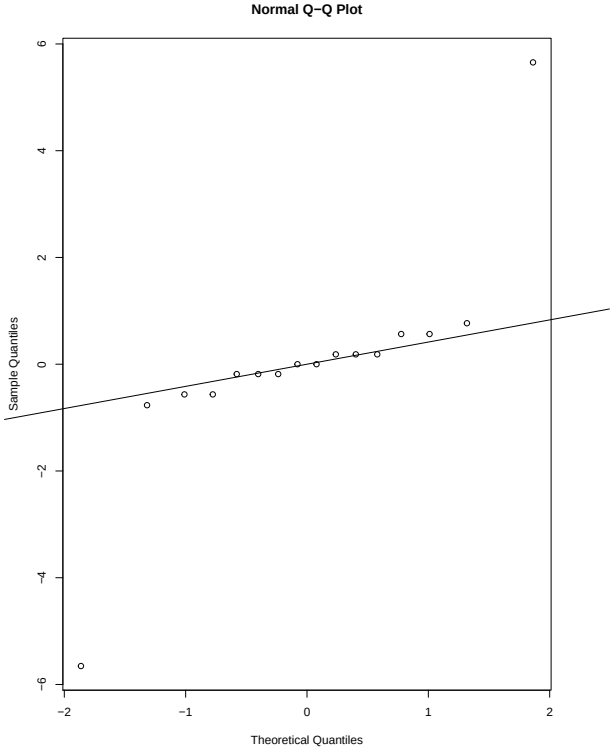

Figure 4: Normal QQ-plot

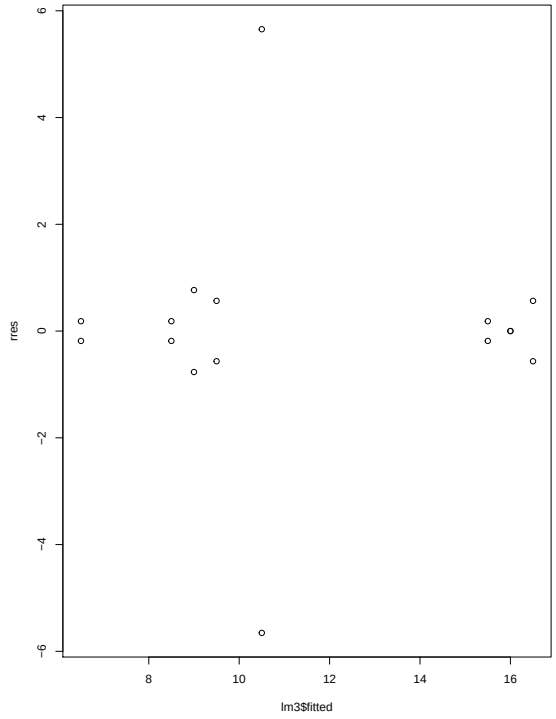

Figure 5: This is the Studentized Residual Plot, which is the predicted values plotted against the residuals divided by an estimate of the standard deviation

We can also see from figure 1, that the study time-factor is the most significant. The P-value is far lower in time than in any other factor and the only one which is significant with level 0.05.

This also tells us that even though other factors may have looked significant when looking at figure 3, we here see that it is only the time-factor which truly is significant in our experiment.

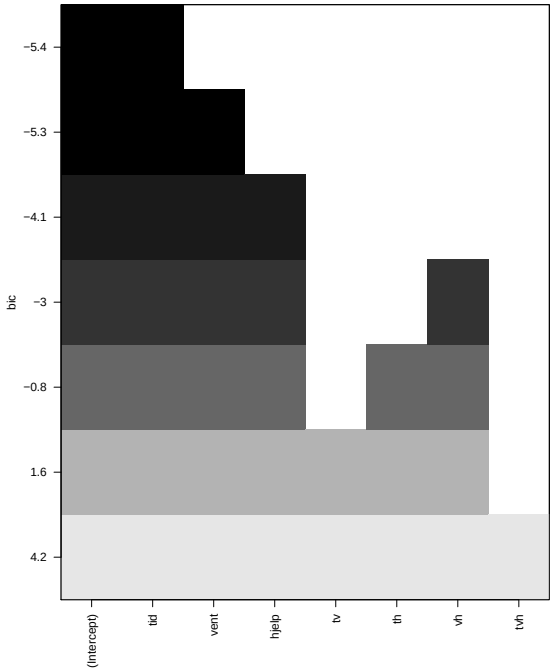

Figure 6: The figure shows a visualization of a best subset analysis using *bic*.

To sum it all up one can see from the summery in figure 1, time is the only significant factor. One can see a similar trend in figure 3 where the difference in low and high level for time is a lot larger than for waiting time and memorization aid. When looking at the interaction plots most of the lines in the plots are parallel or almost parallel. One could maybe consider the interaction between the waiting time and the memorization aid but when looking at the summary 1 this is not significant.

When doing a best subset analysis we get the same result as suspected from the summary 1 which can be seen in figure 6. The best model is the one with only time as a factor. The plot uses *bic* to compare the models with different number of factors. Also when using other methods than *bic* such as *cp* and rss they show that the best model is the one with only study time as a factor.

From this experiment we can therefore conclude that only the difference in time had a significant impact on the memory performance of our test person.

References

[1] *Korttidsminne*. URL: `https://www.tempolex.no/laerevansker/korttidsminne`.