**Final report**

**Business Analytics   
Course: Project Big data (X\_400645).**

**Students:**

Yannick Hogebrug – 2626424 - y.r.hogebrug@student.vu.nl  
Jesse Schouten – 2621562 - [j7.schouten@student.vu.nl](mailto:j7.schouten@student.vu.nl)

Submission date: 28-06-2018



**Introduction**

This study was focused on the data analysis of a bedtime procrastination study. Bedtime procrastination is related to sleep problems, to which severe outcomes including memory and health problems have been related by several studies (Kroese et al., 2014). By analyzing personal sleeping data, as well as poststudy questionnaire data filled out by the participant, the following question was attempted to be answered:

* *Can bedtime procrastination be significantly influenced by experiment?*
* *How well can bedtime procrastination be predicted?*

These questions will be answered during this report. Furthermore, a few visualizations will provide extra insight into the data.

**Data description and exploration**

After merging the two datafiles, the final data has multiple variables for each participant in the study. These variables are shown below including a short explanation if necessary:

- Gender: Male (=1) or female (=2)

- Age:

- Chronotype: 7 point scale if you are more a morning person (1) or an evening person (7)

- Bp\_scale: Bed procrastination scale; the higher, the more procrastination

- Motivation: Going to bed on time each night (1 = not motivated, 7 = very motivated)

- Daytime\_sleepiness: 4-point scale from 0-3; 8 questions, values summed

- Self\_reported\_effectiveness: do you feel more rested since the intervention (range 0-7)

- Group: Control- (0) or experimental group (1)

- Delay\_nights: Number of nights a participant delayed their bedtime (range 0-12)

- Delay\_time: Mean time in seconds a participant delayed their bedtime

- Sleep\_time: The mean bedtime in seconds

Because procrastination is defined as: “voluntary delay of an intended course of action despite expecting to be worse off for the delay”, delay\_time is considered the most valuable variable to answer the earlier stated research question (Kroese et al., 2014).

To get an impression of all variables, we calculated some descriptive statistics, which are shown in Table 1.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **count** | **mean** | **standard deviation** | **median** | **min** | **max** |
| **gender** | 42 | 1,5714 | 0,4949 | 2 | 1 | 2 |
| **age** | 42 | 31,7381 | 12,1500 | 27 | 18 | 61 |
| **chronotype** | 42 | 4,9762 | 1,8450 | 5 | 1 | 7 |
| **bp\_scale** | 42 | 5,0690 | 0,9051 | 5.165 | 2,67 | 6,67 |
| **motivation** | 42 | 4,4524 | 1,1170 | 5 | 1 | 6 |
| **daytime\_sleepiness** | 42 | 16,0476 | 3,8480 | 16 | 8 | 26 |
| **self\_reported\_effectiveness** | 42 | 2,6190 | 1,3619 | 2 | 1 | 6 |
| **group** | 42 | 0,4524 | 0,4977 | 0 | 0 | 1 |
| **delay\_nights** | 42 | 7,2143 | 3,2699 | 8 | 0 | 12 |
| **delay\_time** | 38 | 2354,6842 | 1438,2945 | 1974 | 0 | 5482 |
| **sleep\_time** | 38 | 28822,3158 | 2848,8786 | 29190 | 21929 | 34644 |

Table 1: Measurements per participant of variables in dataset

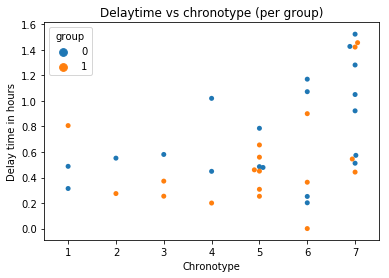
Table 1 shows all appropriate measures to introduce the dataset as a whole. The count stands out as some variables have empty cells, according to the facts not all counts of the variables are the same. For the analysis of the variable with an empty cell, it was decided to remove the corresponding row. So, for some tests more data might have been used then others, as it depends on the selected variables having empty cells. Furthermore, there is not much worth mentioning as there doesn’t seem to be particular strange values or outliers in the data. This might be explained because some filtering had been done before the creation of merged data frame, and some variables having a pre-defined finite scale.

Because the delay time is of importance is this research, we further investigated this variable with the help of a plot. In this research we also look for differences between the two groups of the study, namely the control and the experimental group. The only difference between the groups is the fact that for the experimental group, the lights automatically dim at the intended bedtime. Because of this, we expect to see less delay time for the experimental than for the control group. Figure 1 shows the distribution of the delay time for both the control as the experimental group.



Figure 1: Delay time per participant group

Some difference is spotted, but it is not clear whether this is significant due to lack of observations.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **mean** | **SD** | **median** |
| **Delay nights** | *Control* | 7.0435 | 3.4445 | 7.0000 |
|  | *Experimental* | 7.4211 | 3.0317 | 8.0000 |
| **Sleep time** | *Control* | 28731.8000 | 2914.5239 | 29643.0000 |
|  | *Experimental* | 28922.8889 | 2770.6521 | 29001.0000 |
| **Delay time** | *Control* | 2724.4500 | 1396.4360 | 2828.0000 |
|  | *Experimental* | 1943.8334 | 1371.0637 | 1620.0000 |

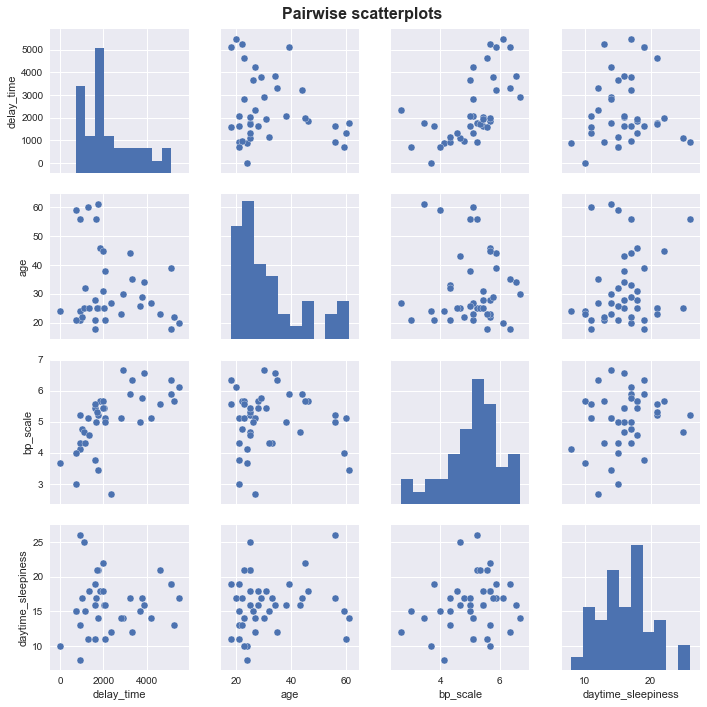
**Data analyses**

Table 2: Results of Wilcoxon signed rank test (difference in group)

|  |  |  |  |
| --- | --- | --- | --- |
| **‘Mean delay time’ vs.:** | **‘Bed procrastination scale’ (pearson)** | **‘Age’ (kendall)** | **‘Daytime sleepiness‘ (pearson)** |
| **Correlation coefficient** | 0.6118 | -0.02746 | 0.08328 |
| **p-value** | 4.4724e-05 | 0.8107 | 0.6191 |

Table 2: Correlation tests for ‘mean delay time’ vs. variables

|  |  |  |
| --- | --- | --- |
|  | Statistic | P-value |
| Delay nights | -0.3285 | 0.7425 |
| Sleep time | -0.1462 | 0.8838 |
| Delay time | 1.9295 | 0.05367 |



**Discussion**

**Conclusion**

**Literature**:

Kroese, F. M., De Ridder, D. T., Evers, C., & Adriaanse, M. A. (2014). Bedtime procrastination: introducing a new area of procrastination. Frontiers in psychology, 5, 611.

