Dragos Pop & Jimmy Gijsel

Assignment 3 – Report

1. Correlations

In the beginning, two correlation tests were used to evaluate the association between three pairs of variables, all including the mean time spent delaying bedtime:

a) The Pearson correlation test between bed procrastination scale and the mean time delaying bedtime shows a reasonable correlation between the variables, namely 0.611, while the p-value (0.00004) suggests that this is significant.
b) Kendall rank correlation test indicates roughly no correlation between age and average time delaying bedtime, 0.02, and the p-value (0.81) confirms it since it is larger than the significance level.
c) Opposite to one's expectations, the correlation between mean time spent delaying bedtime and daytime sleepiness according to Pearson correlation test is not significant as supported by the low coefficient (0.08) and the p-value (0.61).

2. Significant differences

In order to determine if there are significant differences between the experimental and the control groups, the data sets were first tested for normality. In case both are considered to be normally distributed according to the Shapiro-Wilk test (p-value larger than significance level suggest normality), a t-test will be performed. Otherwise, the Wilcoxon rank-sum test will be used. A significance level of 0.05 was chosen.

a) As the experimental group in terms of the number of nights participants delayed their bedtime is not normal, the Wilcoxon rank-sum test was carried out. Its results (W = 0.07, p-value = 0.94) lead to a failure in rejecting the null hypothesis, meaning that there are no significant differences.
b) The t-test performed on the mean sleeping time of the participants provided the same conclusion, that there are no substantial distinctions since its p-value (0.71) is considerably larger than the significance level.
c) Again the experimental group was not regarded as normal distributed and the Wilcoxon rank-sum test was used. This displayed a p-value of 0.053, indicating that the null hypothesis is again not rejected and there are no significant differences between the groups. However, if the significance level would have been chosen higher with only 0.01, the result would have been opposite (rejecting the null hypothesis)

3. Factors that would best predict the delay time

From point 1, it is already known that there is a good association between the delay time and the bp scale, and that age and daytime sleepiness are almost not correlated at all. In addition, we consider motivation a factor that has an effect on delaying the bedtime because if one is not motivated to go to bed at the intended time there is no reason to do so. The sleep time could be a good predictor on the assumption that if one slept for a long time, he/she went to bed later, delaying the bedtime. However, this could be a dangerous assumption as people can also sleep more if they go to bed earlier. In conclusion, our hypothesis is that bed procrastination scale and motivation are the best factors to predict the delay time.

4. Regression model

In the picture attached below, one can see the result of the regression model that uses bp_scale and motivation as factors to predict the mean delay time. The adjusted R-squared of 0.353 reflects the quite low fit of the model and the p-value of motivation (0.3878) suggests that the factor is not significant since it is higher than 0.05. In order to determine if a better result can be obtained, the step up method was performed. Adding each factor one at a time and taking the largest resulted R-squared if the element is significant according to its p-value concluded that the best model is only composed of bp_scale, having a fit of 0.357.

OLS Regression Results

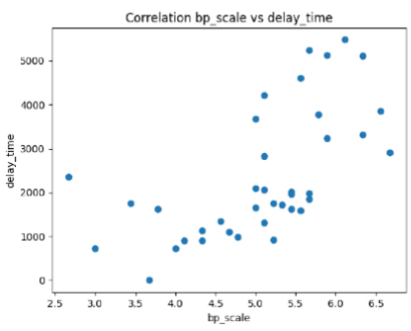
Dep. Variable:	delay_time	R-squared:	0.388
Model:	OLS	Adj. R-squared:	0.353
Method:	Least Squares	F-statistic:	11.11
Date:	Wed, 19 Jun 2019	Prob (F-statistic):	0.000184
Time:	10:52:03	Log-Likelihood:	-320.89
No. Observations:	38	AIC:	647.8
Df Residuals:	35	BIC:	652.7
Df Model:	2		
Covariance Type:	nonrobust		

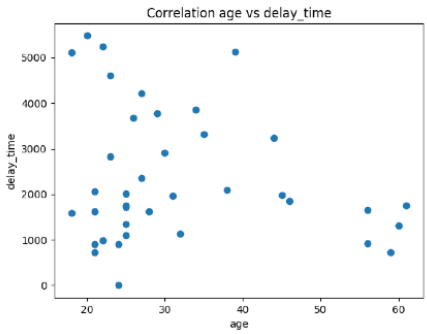
	coef	std err	t	P>|t|	[0.025	0.975]
const	-3198.8333	1357.308	-2.357	0.024	-5954.314	-443.353
bp_scale	959.6480	204.788	4.686	0.000	543.906	1375.390
motivation	156.2134	174.812	0.894	0.378	-198.673	511.100

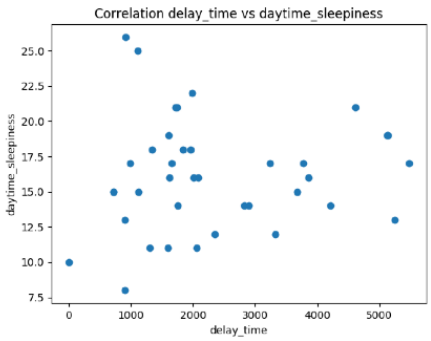
Omnibus:	5.215	Durbin-Watson:	2.011
Prob(Omnibus):	0.074	Jarque-Bera (JB):	4.722
Skew:	0.792	Prob(JB):	0.0943
Kurtosis:	2.311	Cond. No.	49.5

5. Visualisations

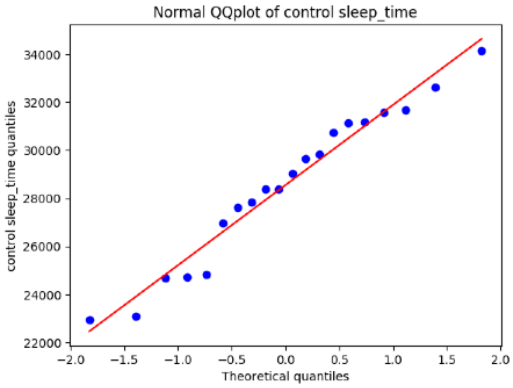
The scatter plots from below represent graphically the correlations from 1. While the first plot seems to roughly have the points arranged together in a straight line, the other two have the dots spread, supporting the conclusion of the first exercise, namely that bp_scale and delay time are roughly correlated.

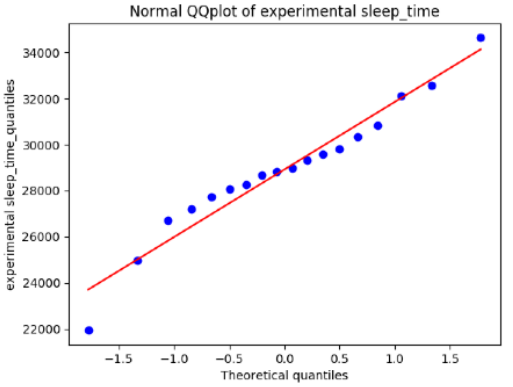






Another method to test for normality, other than the Shapiro-Wilk test, is looking for a straight line in the qqplot of a sample against normal distribution. From the picture on the right, it is visible that both, the control and experimental group of sleep_time follow a straight line in the Normal QQplot, indicating the same evidence that they originate from a normal distribution as Shapiro-Wilk test suggested at exercise 2 b).





In addition to the Wilcoxon rank-sum test, one can check for significant differences between two samples by looking at their boxplots one next to the other. The pictures below suggest the same conclusion such as the one of exercise 2.