

# Take Home Assignment 1

Marc Pabst

5/3/22

In this note, we will use data obtained from a public repository to conduct an a priori power analysis. The objective of the prospective study is to investigate the following research question:

1. Is there a correlation between social mindfulness and life satisfaction?
2. Is there a correlation between social mindfulness and prosociality?
3. Is there a correlation between social mindfulness and impulsivity?
4. Is there a correlation between life satisfaction and prosociality?
5. Is there difference in life satisfaction between people living in the north and the south of the Netherlands?

## Descriptive Statistics

To obtain the final sample for analysis, we remove 7 participants from the dataset because they had missing or invalid values in one or more of the variables of interest (1 missing or invalid data point in “social mindfulness”, 2 missing or invalid data points in “prosociality”, no missing or invalid data points in “impulsivity”, and 4 missing or invalid data points in “life satisfaction”, no missing or invalid data points in “location”). We did not find any outliers in our data (for our purposes, we define outliers as values that are more than 3 standard deviations away from the mean). The final sample therefore consisted of 63 participants. We depict mean, standard deviation, minimum, and maximum for the continuous variables in Table 1. Frequencies and for the binary variable “Location” are shown in Table 2.

Table 1: Descriptive statistics for the continuous variables of interest in the final sample.

Variable	Minimum	Maximum	Mean	SD	N
Impulsivity	1.00	200.00	85.67	44.86	63.00
Life satisfaction	7.00	100.00	51.05	20.69	63.00
Social mindfulness	38.00	83.00	60.43	9.55	63.00
Prosociality	9.00	93.00	49.92	18.25	63.00

Table 2: Frequencies for the the variable “Location” in the final sample as well as conditional means and corresponding 95% confidence intervals for life satisfaction in the north and south of the Netherlands.

Location	N	Mean LS	95% CI
north	35	54.31	47.18 - 61.45
south	28	46.96	39.86 - 54.07

We plot distributions of the continuous variables of interest in Figure 1. Clearly, all variables are approximately normally distributed and are therefore suitable for parametric testing. Scatter plots of the continuous variables of interest are shown in Figure 2. We observe that there is a positive correlation for all four pairs of variables as defined by our first four research questions (compare blue lines in Figure 2).

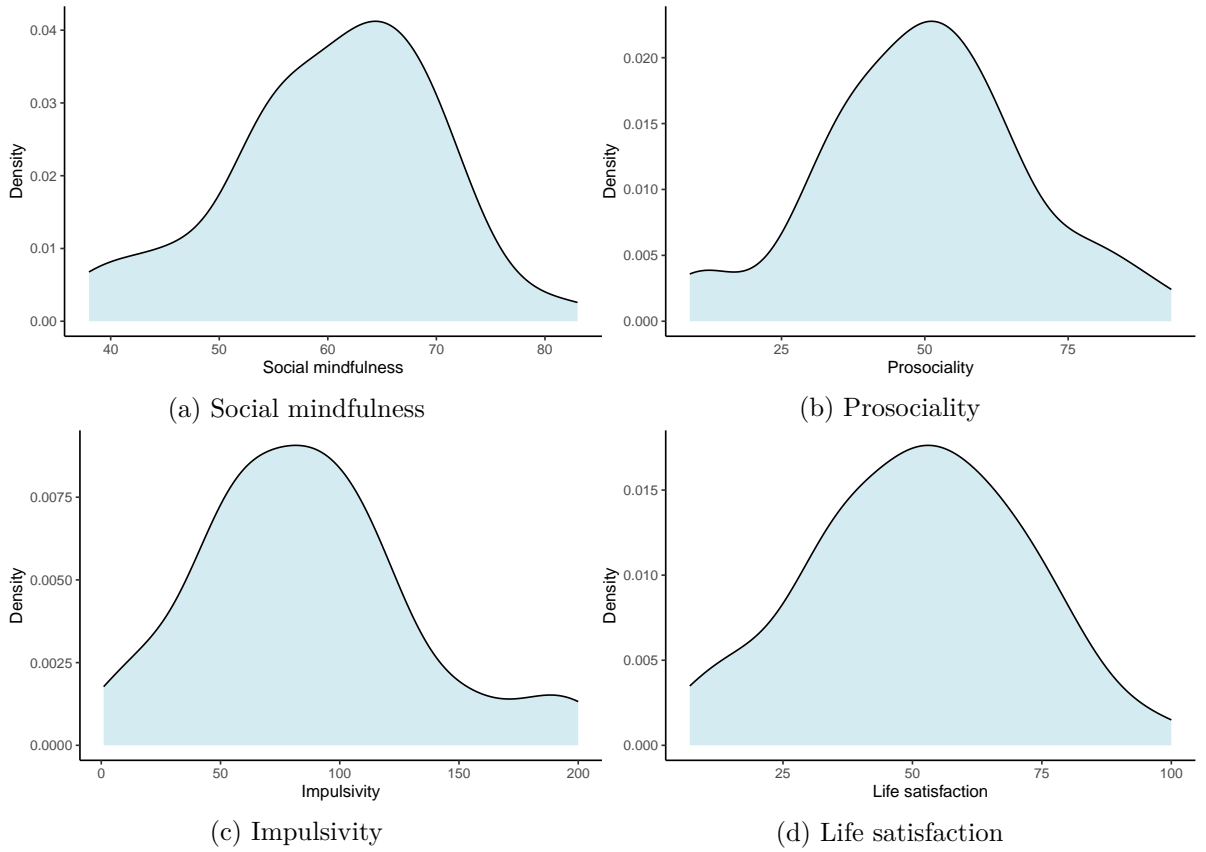


Figure 1: Density plots for the continuous variables of interest in the final sample.

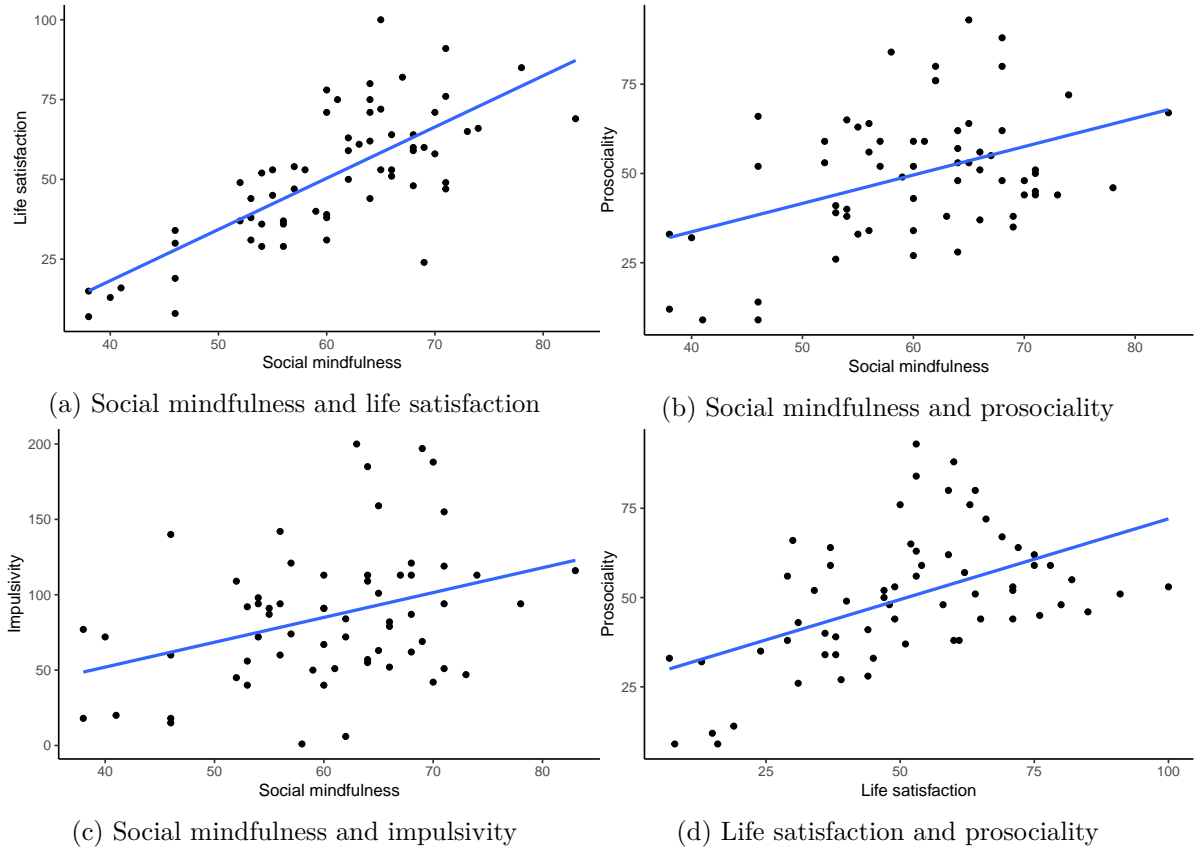


Figure 2: Scatter plots of the continuous variables of interest in the final sample. Blue lines linear relationship between the variables.

## Statistical Analysis

We analyze the final sample of 63 data points that we obtained as described above. To examine the research questions as outlined earlier, we calculate correlations between the continuous variables of interest and group differences in life satisfaction between the north and south of the Netherlands. Effect sizes therefore reflect Pearson’s  $r$  for correlations (first four research questions) and the unstandardized group difference for the fifth research question. All  $p$ -values are reported based on a two-tailed  $t$ -test. We depict the results of the  $t$ -tests for all research questions in Table 3. For the fifth research question, we report a  $p$ -value based on a  $t$ -test for independent samples with no correction of variance.  $P$ -values are corrected for multiple comparisons using the method proposed by Bonferroni (1966).

Table 3: Observed Pearson correlations, corresponding confidence intervals,  $t$ -statistics, and  $p$ -values) for the five research questions (Mind = social mindfulness, Pro = prosociality, Imp = impulsivity, LS = life satisfaction, Loc = location).  $p$ -values are corrected for multiple comparisons using the method proposed by Bonferroni (1966). Effect sizes for the four correlations between the continuous variables are reported as Pearson’s  $r$ , while an (unstandardized) effect size for the difference in life satisfaction for locations is reported as the mean difference between the two groups. Note that the conditional means for the correlation (and corresponding confidence intervals) between location and life satisfaction are reported in Table 2. For the fifth research question, we also report Hedges’  $g$  as a measure of effect size.

Research Question	Eff. Size	95% CI	$t(61)$	$p$	Hedge’s $g$
Mind ~ Pro	0.42	0.19 - 0.60	3.58	0.003	
Mind ~ Imp	0.35	0.11 - 0.55	2.93	0.024	
Mind ~ LS	0.74	0.60 - 0.84	8.62	< 0.001	
LS ~ Pro	0.51	0.30 - 0.67	4.66	< 0.001	
LS ~ Loc	7.35	-3.06 - 17.76	1.41	0.815	0.35

We find that people in the sample that are more prosocial are also more socially mindful ( $r = 0.42$ ,  $t(61) = 3.58$ ,  $p = 0.003$ ). We also find that people in the sample that are more socially mindful are also more impulsive ( $r = 0.35$ ,  $t(61) = 2.93$ ,  $p = 0.024$ ) and more satisfied with their lives ( $r = 0.74$ ,  $t(61) = 8.62$ ,  $p < 0.001$ ). Also, we find that people in the sample that are more satisfied with their lives are also more prosocial ( $r = 0.51$ ,  $t(61) = 4.66$ ,  $p < 0.001$ ). Finally, we could not find any evidence that people in the north of the Netherlands are more satisfied with their lives than people in the south of the Netherlands ( $r = 0.18$ ,  $t(61) = 1.41$ ,  $p = 0.815$ ).

## Power Analysis

We perform a power analysis to determine the number of participants needed to achieve a power of 0.95 for a subsequent study for each of the five research questions. Thus, we will use the observed effect sizes from the present dataset. Here, we use the `pwr` package to perform the power analysis. The power analysis is performed for a two-tailed  $t$ -test with a Bonferroni-corrected alpha level of  $0.05/5 = 0.01$ . The results are shown in Table 4. Here, we also report the expected power for each individual research question based on the number of participants required to achieve a power of 0.95 or more for all five research questions. Note that all reported sample sizes are rounded to the nearest integer and always concern the total number of participants in the study. We find that a total sample of 560 participants is needed to achieve a power of 0.95 or more for all five research questions. We also calculate the power for all research questions as a function of sample size in Figure 3. Here, we observe that the power increases with the sample size for all research questions.

Table 4: Power analysis for the five research questions (Mind = social mindfulness, Pro = prosociality, Imp = impulsivity, LS = life satisfaction, Loc = location). The power analysis is performed for a two-tailed test with an alpha level of 0.05 and a target power of 0.95. The effect sizes are calculated as the correlation coefficient for the first four research questions and as the difference in means for the fifth research question. Values in the column “Target N” indicate the number of participants needed to achieve a power of 0.95.

Research Question	Observed Effect Size	Target N	Expected Power using $n = 560$
Mind ~ Pro	0.42	92	1.0000000
Mind ~ Imp	0.35	136	1.0000000
Mind ~ LS	0.74	23	1.0000000
LS ~ Pro	0.51	59	1.0000000
LS ~ Loc	0.18	560	0.9571518

## Discussion

In the present note, we used a sample obtained from an open repository to conduct a power analysis for a prospective study on the relationship between social mindfulness, prosociality, impulsivity, life satisfaction, and location. We found that a sample of 560 participants would be sufficient to achieve a power of 0.95 or more for all five research questions. However, the needed sample size for the fifth research question is much larger than the needed sample size for the other four research questions. Thus, the fifth research question requires a larger sample size to achieve a power of 0.95. Because recruiting participants for a study is costly and time-consuming, researchers should consider if investigating the fifth research question is feasible

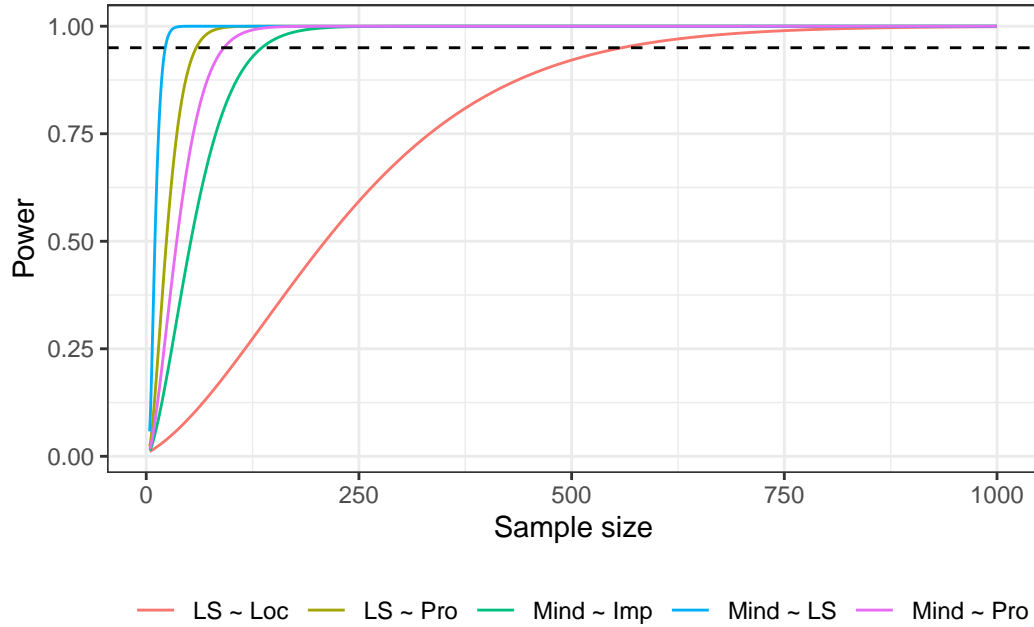


Figure 3: Power analysis for the five research questions as a function of the sample size. The power analysis is performed for a two-tailed  $t$ -test with a Bonferroni-corrected alpha level of 0.01. Figure shows the power for every individual research question as a function of the total sample size. Note that for the fifth research question, the sample size therefore concerns the total number of participants in the study, i.e., the number of participants in the north and the south group combined.

with the available resources. Dropping the fifth research question would reduce the needed sample size to 136 participants.

A prospective study should also consider using a more liberal correction for multiple comparisons. Using a Bonferroni-corrected alpha level of 0.01 is known to be more conservative than necessary to properly control the family-wise error rate and account for multiple comparisons, resulting in reduced power. A more powerful alternative like the Benjamini-Hochberg correction might be more appropriate in this case.