

Assignment Statistics 6

Gustavo Villca Ponce¹, MohammadHossein Haqiqatkah², & Sigert Ariens³

¹ student number:r0292033

² student number:r0607671

³ student number:

Assignment Statistics 6

Introduction

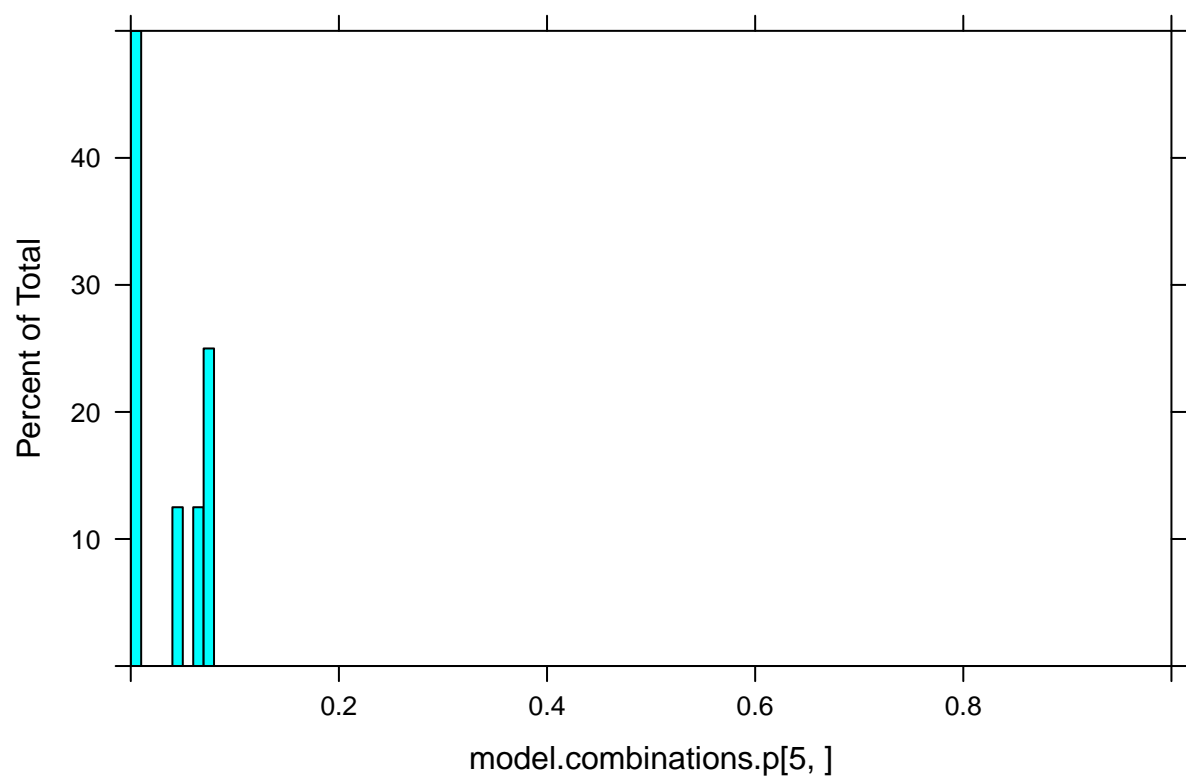
This is a statistical report, based on (add citation) for the class of *Statistics VI (Seminar on statistical analyses of psychological research data)* [P0Q01a]. As required by the guidelines of this project, this report will consist of three main parts, in which we will try to 1)Check the reproducibility status of the published results , 2)Check the robustness status of the confirmatory analyses and 3) Check the pre-registration status of the published results by comparing the pre-registered protocol to the published paper.

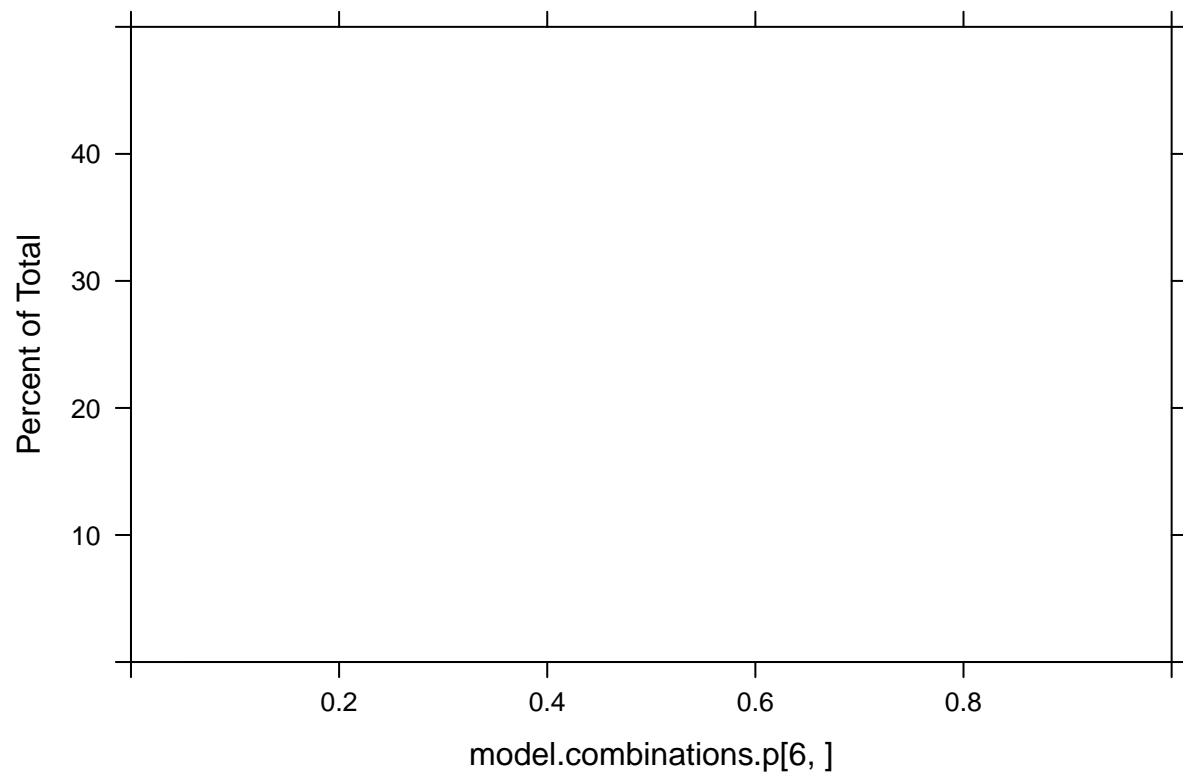
Reproducibility

Exploratory analysis. To begin the replication portion of this report, we start by exploring the possibility of a monotonic relationship between digital screen-time and mental well-being as described by Przybylski and Weinstein (2017), We achieve this by making use of the Bayesian Information Criterion (BIC) and comparing the simple linear models of all variables concerning digital screen-time with their simple and quadratic counterparts. We found that for the models that controlled for confounding variables, the fit was greater for the models with both a linear and quadratic component. For the unadjusted models, we found one model of which the BIC index favoured a linear model alone. Specifically, for smartphone use during weekdays (see, Przybylski & Weinstein (2017), figure 1).

Confirmatory analysis. Following the steps described by the authors we start the exploratory data analysis by creating quadratic models of all four types of digital activities consisting of both linear and of non-linear components, next we extracted all the important value(SD , P -values, β , Confidence intervals and Cohen's d) out of the models and created two tables , the first table contains the outcome of the models without taking into account the control variables described in the paper, namely gender ,ethnicity and Socio Economical Status. The second table contains the outcomes of the models with the control variables (See tables below)

Reproducibility analysis. Although we were able to extract all the important statistics from the raw data without too many issues, notice that some of our values are different from those reported in Przybylski and Weinstein (2017). Specifically, we noticed two types of differences in both tables, the first type are small one decimal differences, for example, in our replication of their analysis we obtain a β value of $\beta = 0.99$ for the linear component of the variable “watching films and tv programs in weekdays”, whereas in their paper the authors reports a value of $\beta = .98$, we see this occur not only for β but for other values too, such as the standard deviation of the linear component of the variable “time spent playing games” (our $SD = 0.12$ vs their $SD = .11$), in the same model we obtain a *cohens'd* value of 0.20, compared to their *d* value of .19. These small one decimal differences can be found in both tables, a potential explanation would be a differences in the rounding of the number. This explanation becomes less likely, however, once we take onto account the second type of difference we encountered. We found decimal differences exceeding one decimal, for example, in table 2 we observe a $\beta = 0.27$ for the linear component of “playing games in the weekdays” vs $\beta = .21$ reported in the paper. These multi-decimal differences can’t be explain fully by a rounding difference of the decimals. In order to find the origin of the different outputs, we looked into the data used by the authors for their SPSS analysis. We noticed a difference in the amount of missing values(NA) between the raw data and the data used in their SPSS analysis, which will be elaborated on in the preregistration section. It is likely that the researcher handled the missing values in a way that wasn’t reported in the paper, making it hard for us to fully replicate the results without any differences. Furthermore, the way the variables were coded in an ambiguous manner, making it difficult to determine which specific variables were used in the models reported by the researchers. Overall, there was a lack of clarity in crucial data processing steps such as missing values and variable identification.





```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
## [1] 9
## [1] 10
## [1] 11
## [1] 12
## [1] 13
```

```
## [1] 14
## [1] 15
## [1] 16
## [1] 17
## [1] 18
## [1] 19
## [1] 20
## [1] 21
## [1] 22
## [1] 23
## [1] 24
## [1] 25
## [1] 26
## [1] 27
## [1] 28
## [1] 29
## [1] 30
## [1] 31
## [1] 32
## [1] 33
## [1] 34
## [1] 35
## [1] 36
## [1] 37
## [1] 38
## [1] 39
## [1] 40
```

```
## [1] 41
```

```
## [1] 42
```

```
## [1] 43
```

```
## [1] 44
```

```
## [1] 45
```

```
## [1] 46
```

```
## [1] 47
```

```
## [1] 48
```

```
## [1] 49
```

```
## [1] 50
```

```
## [1] 51
```

```
## [1] 52
```

```
## [1] 53
```

```
## [1] 54
```

```
## [1] 55
```

```
## [1] 56
```

```
## [1] 57
```

```
## [1] 58
```

```
## [1] 59
```

```
## [1] 60
```

```
## [1] 61
```

```
## [1] 62
```

```
## [1] 63
```

```
## [1] 64
```

```
## 39.08 sec elapsed
```

Multiverse Analysis.

Model Driven Multiverse. We found that, by manipulating the variables included in the models in 8 different ways (including/excluding each control variable “gender”, “ethnicity”, and “SES”. We found that this multiverse of models only resulted in different p values for the linear and linear+quadratic models of weekday game usage, and the only variable underlying this difference was “gender”. Including the control variable of gender, regardless of any other combination of variables, shifted the average p value of the slope of the linear model from $p = 0$ to `p.values.avg.5.g`.

Data Driven Multiverse. We took into account different combinations of possible coding strategies that could have been used by the researchers, in an approach akin to Steegen, Tuerlinckx, Gelman & Vanpaemel (2016). For the recoding of the categorical variable “ethnicity”, we allowed each level but “white” to be recoded into the binary variable “minority”. The same was done for the categorical variable “deprived”, although similarly to the previous recoding, the first level was maintained due to reasonability concerns. This resulted in a multiverse consisting of $2^6 = 64$ possible codings of the variable. We then examined the distribution of the p and d values resulting from applying the full model, thus including all control variables, to the different data sets. Two models had diverse distributions of p values. Specifically, the distributions of p values of the purely linear model of weekday game usage resulted in the histogram in figure 1

The distributions of p values of the linear+quadratic model of weekday game usage resulted in the histogram in figure 2

It is interesting to see the multiverse-frequency of p values below the significance threshold of $\alpha = .001$ given by the researchers. We find that, for this model, the multiverse encapsulating all possible researcher choices in variable coding only results in a significant value for this model in 1% of the cases. For all other cases, the p values of the models seem reasonable. However, we obtain a different p value than the authors when using the full model on the subset of the data multiverse used by the authors. Specifically, we find a p value of $p = 0.08$ for the linear model describing weekday game usage, compared to the

$p = 0.059$ authors. These differences, along with the difference in regression parameters reported above, can likely be ascribed to the ambiguities mentioned throughout the paper.

Preregistration. The data were acquired according to the specifications made by the authors in the preregistration document. However, in the technical report on their OSF page, the authors said to use a 3% margin of error at the 95%CI to estimate sample size. In the published paper, the authors report a 0.3% margin of error, arriving at the same estimate of sample size ($N = 298,080$). Furthermore, The authors reported a total n of 120,115 participants with usable data. When we attempted to replicate their analyses, we met with a further reduction of n to 98278. This is not reported anywhere in the published article. Finally, the two data documents provided by the authors differ in the amount of NA data they contain. Where the .csv file contains 21837 rows with missing values, the .sav file contains 44642 rows with missing values. None of these inconsistencies were reported by the authors in the final paper, and it is unclear which data set the authors ultimately used in their analysis. The fact that the data are ambiguous is a major obstacle for replication analyses.

The preregistration document stated that testing the displacement hypothesis was to be done by linear regression modelling, predicting mental well-being through composite scores of screen time. The authors did not conduct these analyses. They explain that “Interocular” tests were sufficient to exclude this hypothesis. Although we only found one unadjusted model for which the BIC criterum favours a purely linear model, this does allow one to question why the authors refrained from the formal hypothesis test detailed in the preregistration document.

The authors ignored the measure of summed screen time which they included in their preregistration document. The authors reported this accordingly, although our above analysis allows the questionability of their deviation on this point. Finally, the authors did not concretely specify what particular coding they intended to use for the control variables of “whether living in a deprived area”, and “whether black and minority ethnicity”. While

the conditionality of the term “whether” implies binary coding, the subsequent reference to the specific questions would also allow one to assume that the authors used the values from those questions. This is relevant because other codings of these variables are also present in the data. Although this is a minor issue, a more clear issue is also present. The authors noted in their deviations from analysis plan that the preregistered control variables of “whether parents married” and “whether native born”, they did not include the omission of these variables in the published article.

References

Table 1

(#tab:table making)Uncontrolled variables

	b	SE	CI(2.5%)	CI(97.5%)	p	d
Watch Weekday Linear	0.99	0.10	0.79	1.20	0.00	0.06
Watch Weekday Quadratic	-0.14	0.01	-0.16	-0.12	0.00	0.09
Watch Weekend Linear	1.55	0.10	1.36	1.74	0.00	0.10
Watch Weekend Quadratic	-0.17	0.01	-0.18	-0.15	0.00	0.13
Play Weekday Linear	3.71	0.12	3.47	3.95	0.00	0.20
Play Weekday Quadratic	-0.34	0.01	-0.37	-0.32	0.00	0.18
Play Weekend Linear	3.20	0.09	3.03	3.38	0.00	0.22
Play Weekend Quadratic	-0.27	0.01	-0.28	-0.25	0.00	0.20
Computer Weekday Linear	1.32	0.10	1.11	1.52	0.00	0.08
Computer Weekday Quadratic	-0.17	0.01	-0.19	-0.15	0.00	0.11
Computer Weekend Linear	1.60	0.09	1.42	1.78	0.00	0.11
Computer Weekend Quadratic	-0.18	0.01	-0.19	-0.16	0.00	0.14
Smatphone Weekday Linear	-0.56	0.09	-0.73	-0.40	0.00	0.04
Smatphone Weekday Quadratic	-0.01	0.01	-0.03	0.00	0.11	0.01
Smatphone Weekend Linear	0.46	0.08	0.29	0.62	0.00	0.03
Smatphone Weekend Quadratic	-0.10	0.01	-0.11	-0.08	0.00	0.08

Table captions

Table 1. (#tab:table making)Uncontrolled variables