[Re] Marriage and happiness: Providing evidence against a relationship between inequality and happiness in Oishi, Kesebir, and Diener (2011)

Lauren Heintz, Tara Wilson, Ben Mathieu, Will Wright

22 February, 2020

A reference implementation of:

Grunberg, Rebecca L., Hyejun Kim, and Minjae Kim. 2014a. “Marriage and Happiness Grunberg Kim Kim.pdf.” Harvard Dataverse. <https://doi.org/10.7910/DVN/25655/MB980L>

Which is a reference implementation of:

Oishi, S., Kesebir, S., & Diener, E. (2011). Income Inequality and Happiness. Psychological Science, 22(9), 1095–1100. <https://doi.org/10.1177/0956797611417262>

There is a crisis of reproducibility and replication in the data science field. In an attempt to bolster confidence in scientific results, the following paper is a replication of “Marriage and happiness: Providing evidence against a relationship between inequality and happiness in Oishi, Kesebir, and Diener (2011)” (Grunberg, Kim, and Kim [2014a](#ref-marriagehappiness)). This paper follows the ReScience replication method(Rougier and contributors [2020](#ref-rescience)) which strongly encourages open source practices as underlined by the format of this project: a github compendium. This particular replication is unique in that the paper being replicated is an alternate implementation of the paper:“Income Inequality and Happiness” (Oishi, Kesebir, and Diener [2011](#ref-incomeinequality)). While the first paper draws a connection between happiness and income, the second calls in the question the data used for this conclusion, and pulls in additional data for analysis. This second analysis, focused on the relationship between marriage and happiness, is the subject of replication in this paper.

# Introduction

# Methods

## Overview

The paper of interest(Grunberg, Kim, and Kim [2014b](#ref-data)) used linear models to show that the correlations drawn by the original findings(Oishi, Kesebir, and Diener [2011](#ref-incomeinequality)) do not hold once the interactions of variables known to be correlated with the response variables are added to the model.

In this replication paper, three artifacts from the paper(Grunberg, Kim, and Kim [2014b](#ref-data)) were reproduced. The first artifact, Figure 3., plots the mean happiness by year and race, the second, Table 1, shows the linear regression model’s coefficients and the third, Table 2, shows the correlation coefficients of the different independent variables used in the linear models created.

## Dataset

Marriage and Happiness(Grunberg, Kim, and Kim [2014b](#ref-data)), obtained data from the General Social Survey (GSS) and the US Census Bureau. The subset of GSS data provided omitted the RACE and MARITAL columns crucial to the replication of the figures and models. In order to complete the missing data, the original dataset was queried directly using the gssr(Healy [2019](#ref-gssr)) R package.

The dataset from gssr contained a larger time interval than the original paper, it was filtered back to match the original intervals. This dataset was also merged with the original US Census Bureau data found in Gini\_families.csv(Grunberg, Kim, and Kim [2014b](#ref-data)) by appending the total column to the GSS data using year as the key. The “happy” variable was re-encoded so that the numerical factor would reflect an increasing level of happiness as per the original code.

## Replicating Figure 3

Figure 1: Figure 3 from the original paper

Figure 1: Figure 3 from the original paper

The original R code used the standard R libraries for visualizations. The ggplot2 package was leveraged to replicate the figure with the new dataset. Axis labels, lines and colors were adjusted to match the original output.

## Replicating Table 1

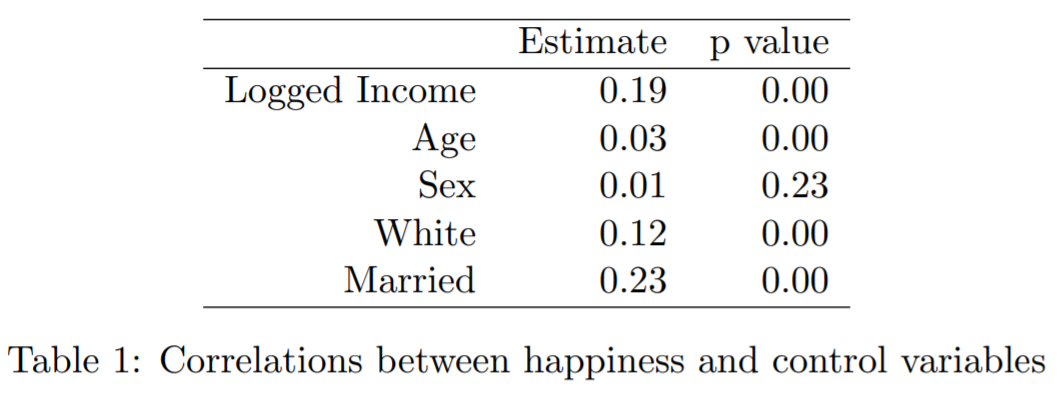


Figure 2: Table 1 from the original paper

The R function cor.test() was used to get the Pearson correlation coefficient between the different independent variables (income, age, sex, white, married) and happiness. Grundberg et al.’s code was slightly modified to keep the same encoding of variables given the new dataset input.

## Replicating Table 2

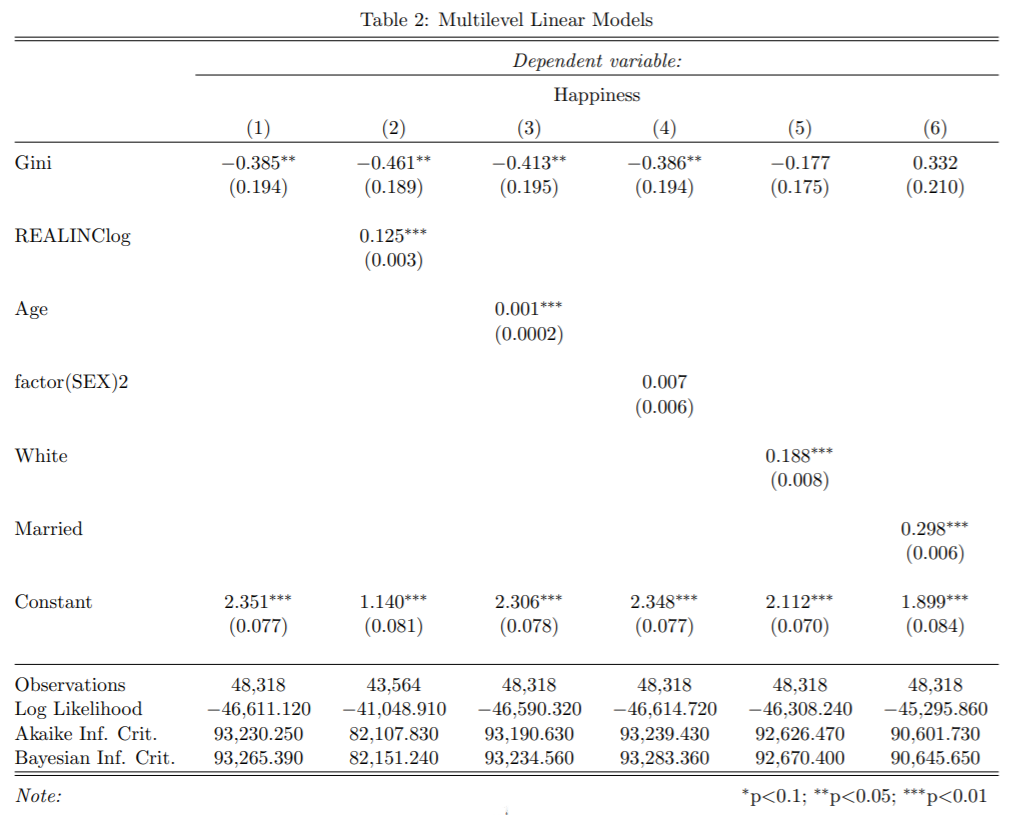
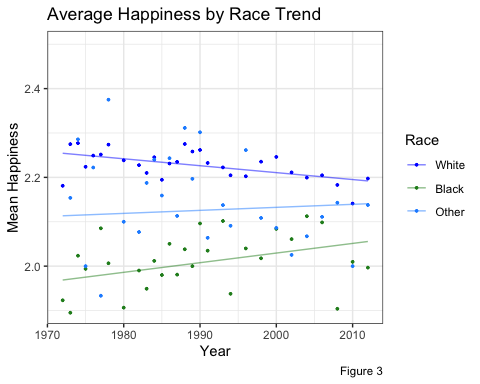


Figure 3: Table 2 from the original paper

Grundberg et al used the lme4 package to generate coefficients of linear mixed-effects models using the intercept to capture year as the random effect (see (Bates et al. [2015](#ref-lmefour)) section 2).

# Results



Estimate

p value

Logged Income

0.19

0.00

Age

0.03

0.00

Sex

0.01

0.23

White

0.12

0.00

Married

0.23

0.00

The to almost exactly reproduce the results from Table 1 in the original paper that shown in Figure 2. The only difference is the estimate for Sex was 0.01 in the publication compared to 0.00 in our replication. This may be due to rounding versions or an underlying exclusion of outliers we saw with the above plot.

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu % Date and time: Sat, Feb 22, 2020 - 13:05:30

Again we were able to almost identically reproduce the results from Table 2 in the original paper, shown in Figure 3. Aside from slight variability in the second and third decimal places, our results from the linear models are very close. There are several differences wit the number of observations, listed in the bottom of the table, which is to be expected as we are using a package with the GSS data and it is likely not an identical snapshot to the one used in the original paper. Overall though these two tables were indeed highly reproducible even with different input data.

# Conclusion

Replication of this paper was ulimately contingent on tracing back the original dataset. Without this resource being cited and publicly available, it would not have been possible to replicate the different tables and figure presented in the original paper.

# References Cited

Reference 1: Grunberg, Kim, and Kim ([2014a](#ref-marriagehappiness))  
Reference 2: Grunberg, Kim, and Kim ([2014b](#ref-data))  
Reference 3: Oishi, Kesebir, and Diener ([2011](#ref-incomeinequality))  
Reference 4: Healy ([2019](#ref-gssr))  
Reference 5: Rougier and contributors ([2020](#ref-rescience))  
Reference 6: Bates et al. ([2015](#ref-lmefour))

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using Lme4.” *Journal of Statistical Software, Articles*. <https://doi.org/10.18637/jss.v067.i01>.

Grunberg, Rebecca L., Hyejun Kim, and Minjae Kim. 2014a. “Marriage and Happiness Grunberg Kim Kim.pdf.” Harvard Dataverse. <https://doi.org/10.7910/DVN/25655/MB980L>.

———. 2014b. “Replication data for: Marriage and happiness: Providing evidence against a relationship between inequality and happiness in Oishi, Kesebir, and Diener (2011).” Harvard Dataverse. <https://doi.org/10.7910/DVN/25655>.

Healy, Kieran. 2019. “gssr: General Social Survey data for use in R.” <http://kjhealy.github.io/gssr>.

Oishi, Shigehiro, Selin Kesebir, and Ed Diener. 2011. “Income Inequality and Happiness.” *Psychological Science* 22 (9): 1095–1100. <https://doi.org/10.1177/0956797611417262>.

Rougier, Nicolas P., and contributors. 2020. “ReScience C: Overview of the submission process.” <http://rescience.github.io/write/>.