

“The effect of various socio-economic factors on major depression scores in unemployed individuals”

Cristen McCann, Yilin Song, and Vitaly Ievlev

12/11/2017

I. Introduction

In this project we are exploring statistical relationships between depression and different socio-economic factors that might contribute to it. Although anyone is susceptible to major depression (especially those with family history), social and economic stresses are important environmental factors that make some people more vulnerable than others, prompting us to investigate this underlying phenomenon. In order to accomplish the task at hand, we examined the data collected as a part of Job Intervention Program evaluation that was conducted in the University of Michigan in 1992. We are focusing only on the socio-economic explanatory variables of sex, age, economic hardship, marital status, and education to explain depression score (the latter of which was measured quantitatively).

It makes intuitive sense that worsening of socio-economic environment negatively impacts depression. More specifically, a 2007 study by Lorant et al. showed that a decrease in material standard of living was associated with an increase in depressive symptoms and incidents of major depression. Similar findings were observed by Madianos et al in 2010 in their article about relationship between economic hardship during the 2008 economic crisis. They found out that economic hardship and depression were strongly correlated. Based on this information we would also expect to see a statistically significant relationship between financial status and depression. Besides income, depression status is known to be influenced by other socio-economic factors. One such factor is family situation. Family separation can seriously impact depression status in a negative way (Lorant et al., 2007). This information gives us a rough idea of what to expect from a statistical comparison of marital status and depression.

Past findings regarding the relationship between education and depression indicate that higher education levels correlate with lower depression scores (Crespo et al., 2014). It would be interesting to see how education may affect depression after accounting for other variables such as economic hardship and marital status, in our own study.

In addition, past studies show that age may be a confounding factor and can directly influence depression scores, and that many variables may interact to produce unexpected results. For example, a Canadian depression study from 1996-2013 suggests that, although females tend to have higher depression scores, the gap between female and male depression scores closes as the subjects reach middle age (Patten et al., 2016). The study provides insight for our own research, because it is evident that the effect of gender on major depression varies with age, so our variables may interact and influence each other in ways that we may not expect. We would like to investigate and take into account how and if the interplay between our variables of gender, age, economic hardship, marital status, and education affects depression scores.

Based on past findings, we hypothesize that depression scores are higher with greater economic hardship, lower levels of education, the female sex, and divorced subjects.

II. Materials and Methods

Study description:

As aforementioned, respondents were recruited from four offices of the Michigan Employment Security Commission (MESC) in southeastern Michigan in 1992 in order to participate in the Job Search Training Program. The information was originally collected from 1801 non-randomly sampled unemployed adult respondents who participated in the program. The set we are using is a condensed version of the original

dataset and contains 899 observations of 17 different variables. It is found in the mediation package of the R server. The non-random recruitment method used for gathering respondents suggests that we cannot generalize our findings from this dataset onto the population.

The original study focused on effectiveness of the job search training program, therefore the original dataset included variables before and after the training (treatment). We did not need the after-treatment data for our purposes. The reason for it is that we decided to take a different approach to analyze these data by focusing only on the socio-economic aspects and depression before treatment. Therefore, we deleted all the Job Intervention Program-related variables from the dataset leaving only socio-economic factors and depression scores. There were no missing data (NULL values) in our data set, which made it very easy to clean.

Defining our variables:

The depression variable is continuous quantitative, with values from 1-5. This information was gathered by asking each participant 11 questions about their own experiences with depression, for which they answered on a categorical scale from 1-5 (meaning “not at all” to “extremely”) and the answers to the 11 questions were averaged, converting the categorical variable into a continuous quantitative variable. This method is based on the Hopkins Symptoms Checklist, which had 25 questions about overall mental health. The answers are in a different scale ranging from 1-4, meaning “Not at all,” “A little,” “Quite a bit,” and “Extremely” respectively (the depression score was calculated by averaging answers for 15 questions about depression specifically).

The economic hardship variable is discrete quantitative, for which the study took the mean of 3 questions with answers from 1-5 (5 indicated highest economic hardship level). In addition to this variable, we also created a new variable called “economic”, that divides the quantitative economic hardship variable into 4 categories: mild (0-2), moderate (2-3), high (3-4), severe (4-5). This allowed us to perform an ANOVA test and make a boxplot that was easier to read and analyze. We used the “economic” variable in addition to the economic hardship variable, not as a substitution.

A Job Intervention Program evaluation survey was used to gather demographic information of age, gender, education, marital status, occupation, family income, and ethnic/racial identification.

The age variable is continuous.

The sex/gender variable is categorical, with the 2 categories (coded 0 as male and 1 as female).

The education variable is categorical with 5 categories (less than high school, high school, community college, bachelors, and graduate).

The marital status variable is categorical, with 5 categories (married, never married, separated, divorced, and widowed).

The occupation variable is categorical with 7 categories (professionals, operatives, managerial, sales workers, craftsmen, clerical, and laborers/service).

The family income variable is categorical, with 5 categories (50k+, 40-49k, 25-39k, 15-24k, less than 15k).

The ethnic/racial variable is categorical, with 2 categories (white and non-white).

Analyses:

For each of our explanatory variables, we found summary statistics, primarily mean, median, and standard deviation. The statistical analysis method of ANOVA were used on depression scores (quantitative variable) paired with economic hardship, education, and marital status (categorical with over 2 levels). For analysis of depression and gender (binary) we used a Welch’s t-test. We also did a multiple linear regression on depression scores with economic hardship, gender and age. This allowed us to consider age as a confounder for both economic hardship and gender. Also, we could test for the individual impacts of economic hardship and gender on depression after adjusting for one another. We used 0.05 as our significant level in our tests.

In order to visually represent the data we relied primarily on boxplots (depression vs marital status, depression vs economic hardship, depression vs education, depression vs gender), scatter plots (depression vs age) and histograms (age, depression).

III. Results

1. Gender and depression

Comparison of gender and depression (Table. 1) showed that the mean depression score is 1.92 for women, and 1.82 for men. The box plot shows a slightly larger depression score median and range for females than males (Fig.1). From the t-test (Table. 2) of gender and depression, the p-value is 0.008852 which is smaller than 0.05. Therefore, we can conclude that there is a statistically significant relationship between gender and depression status. The confidence interval (Table. 2) is [-0.1724, -0.0248] meaning we are 95% certain that the true difference in proportions of depression between males and females excludes zero, which is in favor of the alternative hypothesis. From the means of male and female depression scores, there is an evidence to suggest that females are more depressed on average than males are.

2. Economic hardship and depression

We represented the relationship between economic hardship and depression in a form of a box plot. By looking at the plot it is clear that there is an association between economic hardship and depression scores (Fig.2). People who experience severe economic hardship are clearly more depressed with a mean depression score at 2.13 (Table. 3). While people with little or no economic hardship have a mean depression score at 1.59 (Table. 3) From the summary statistics, we can see a clear trend that the mean depression score increases as we go from mild to severe economic hardship. The p-value from ANOVA test (Table. 4) is smaller than 0.00001, so we have statistically significant evidence to reject the null hypothesis in favor of the alternative, which suggests there is an association between economic hardship and depression. Then, we compared each possible pair of combinations. All of the pairs of categories in our Tukey analysis (except the difference in mean depression of severe vs high economic hardship) had confidence intervals that excluded the value 0, showing that we have statistically significant p values in the first 4 categories, so we are 95% confident that the true mean depression scores for each of those first four categories are in ranges that exclude 0, as described in our Tukey table (Table. 5). Thus, we reject our null hypothesis in favor of the alternative.

3. Education and depression

Comparison of education level and depression by a box plot and summary statistics yielded some interesting statistics (Fig.3, Table. 6). For example, it was clearly evident that people who didn't finish high school were the most depressed on average with a mean depression score at 2. Meanwhile, people who received high school, bachelors and professional degrees all have the mean depressions at approximately 1.84. However, the p-value in ANOVA is 0.313 which means we fail to reject our null hypothesis and there is no statistically significant evidence to suggest there is a relationship between education and depression (Table. 7).

4. Marital status and depression

Assessment of the boxplot of the marital status and depression showed that married people had a lower mean of depression than all the other groups (Fig.4). Mean depression score for married people was 1.81, while divorced and widowed participants had the highest means of depression (1.94 and 1.97 respectively) (Table. 8). We can see from both the box plot and summary statistics that there is a difference between the depression scores in married people and divorced people. However, the p-value from ANOVA analysis is 0.0854, which is slightly above our significance level (Table. 9). It is possible that there might be some relationships between marital status and depression, but we still fail to reject the null hypothesis in our model.

5. Economic hardship, gender, and age vs depression

Economic hardship (treated as a quantitative variable in this case), gender and age were all included in a multiple linear regression model. After adjusting for gender and age, economic hardship still has a statistically significant relationship with depression scores (p value <0.0001) (Table. 10). The slope is 0.1878 which means that controlling for gender and age, for every additional increase in economic hardship, the depression score is expected to increase 0.1878. After adjusting for age and economic hardship, gender is still statistically significantly associated with depression (p value <0.0001). The slope is 0.121 which means that controlling for economic hardship and age, females are expected to have 0.121 higher depression scores than males. The p -value for the coefficient of age on depression in this study was found to be not significant.

IV. Discussion

Overall, our results demonstrate that there is a relationship between several socio-economic factors and depression. Specifically, out of the socio-economic factors that we examined (namely: economic hardship, income, education, marital status, gender, race, occupation and age) only three factors (economic hardship, income and sex) were suggested to be significantly associated with depression.

Even after adjusting for age and gender, economic hardship had a significant positive correlation with depression, which is consistent with other research (Larant et al., 2007; Batterworth et al., 2009; Madianos et al., 2010). Interestingly, people experiencing high and severe economic hardship had approximately the same depression scores, which were higher than those of subjects from the mild and moderate economic hardship groups. That means that beyond a certain limit of economic hardship, depression score reaches saturation. Although income was found to be significantly negatively correlated with depression, we do not discuss it in much detail because it directly reflects economic hardship discussed above.

Subjects of the female sex had a significantly higher depression score than male subjects. We are not the first to observe this kind of trend in the population. Another study of the sources of major depression in Canada also found females to be more susceptible to depression than males (Patten et al., 2006). Even though the relationship between gender and depression might be affected by income and age, which are a confounding variables, after adjusting for both of these potential confounders the p -value remained significant.

Although marital status did not significantly affect depression in our study, some literature on the subject suggests otherwise. Lorant et al. in 2007, reported that married people have a significantly lower depression than those who were recently separated. That makes logical sense and our data supports this trend, yet our study did not yield a small enough p -value to make it statistically significant.

Education level was also not significantly associated with depression in our study. Our initial hypothesis was that lower level of education was associated with higher depression. Our data, however, did not show any significant association between advanced degrees and depression, although the trend according to the box plot supported our hypothesis. The education category with the highest depression score was the high school dropouts. These finding are in line with our hypothesis and with the findings of other researchers in that that higher education decreases overall depression (Crespo et al., 2014) The relationship between education and depression is most likely confounded by income.

Race, age and occupation variables did not show any interesting trends with depression scores and were all found to be statistically insignificant. Overall, it looks like we are only sure that females and those with more economic hardship tend to have a higher major depression score.

Our results are not generalizable to the entire US population. Since respondents were recruited and were only unemployed individuals who participated in job search training program, the data is not randomly sampled. This non-random sampling introduces bias in our conclusions. Therefore, the sample is not representative of the entire US population. There are also too many confounding variables that we do not have the means of testing given the scope of this class, such as occupation or income, since they are not binary categorical variables.

There are several limitations to the study. For example, as stated above, it was only conducted with unemployed people in job training program which is not random sampled, so it can't even be generalized to the entire U.S population, let alone the world. In addition, the depression score scale used in the study was not very clear at defining each number on the scale, so we only know 1-5 as "not at all" to "extremely". This makes a number such as 2 not very meaningful, since we cannot accurately describe what a score of 2 means, and respondents may have different interpretations of the scales when answering. In addition, our depression score variable is not perfectly quantitative, because it is only quantitative after taking the average of values for an ordinal categorical variable (1,2,3,4,5). In addition, while we were able to perform multiple linear regression with several categorical variables, the nearly normal residuals condition is not fully met, so we cannot be sure of its statistical significance. Having it meet all conditions for multiple linear regression would allow us to create a model where each of the variables would be accounted for. This way we would be able to assess contribution of each individual explanatory variable to the overall depression score. The multiple R² value for our multiple linear regression is 0.1154 and the adjusted R² is 0.1124 which means that the variability in depression scores is hardly explained by our model.

A strength in our analysis was that when we found that the variable economic hardship did not meet the assumptions that would allow us to do a regression test, we converted economic hardship into a categorical variable in order to use ANOVA. Therefore, through data modification, we were able to successfully analyze the relationship of economic hardship and depression and found a significant p-value. We still used economic hardship as a quantitative variable when we did multiple linear regression, since it fitted better into our model in this way.

In our dataset, the quantitative variables were presented within a very limited range (1-5 for economic hardship and 1-3 for depression score). Therefore, it is more difficult to see the differences in depression score or economic hardship between different categories of other variables, especially in visuals such as linear regression. If we were to recollect the data we would expand the scale of quantification of these variables. It would also be interesting to perform another survey with a wider sample (not just unemployed subjects). Further research questions may also explore how the aforementioned socio-economic variables may affect the length or recurrences of depressive episodes, rather than just a depression score.

V. Tables and Figures

Figure 1: Boxplot of Depression Status and Gender

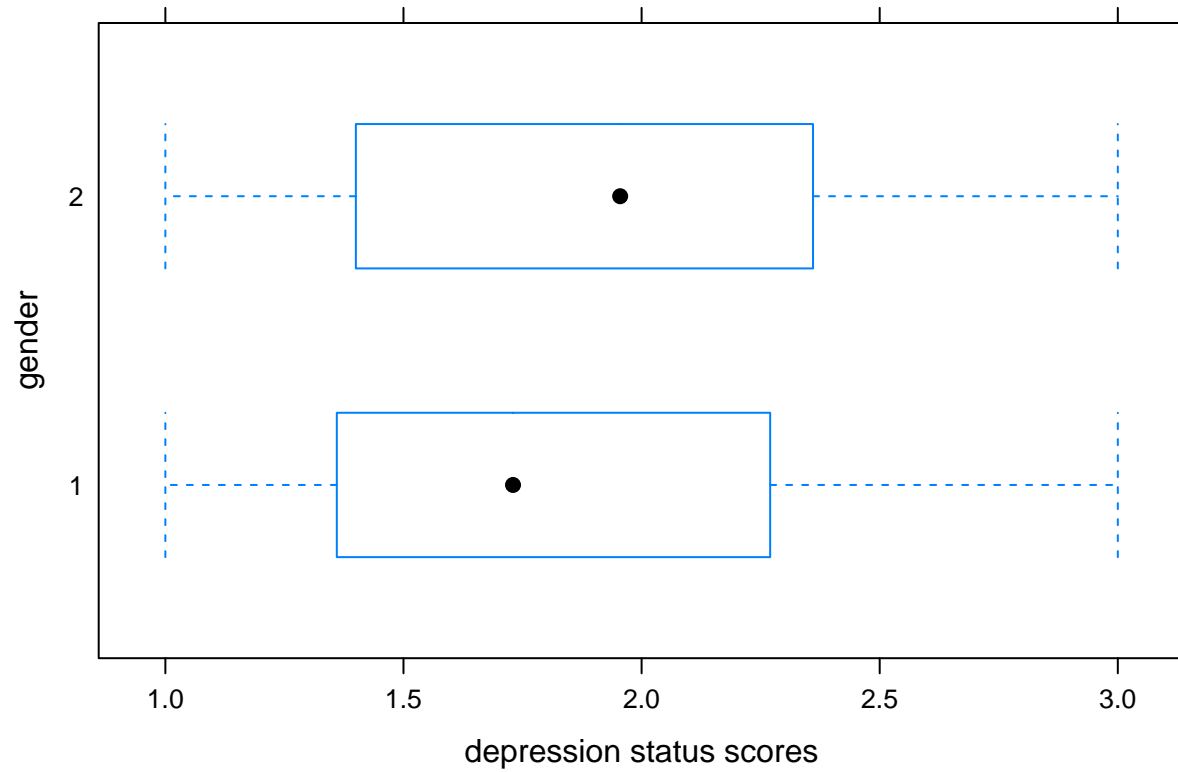


Figure 2: Boxplot of Economic Hardship and Depression Status and ANOVA

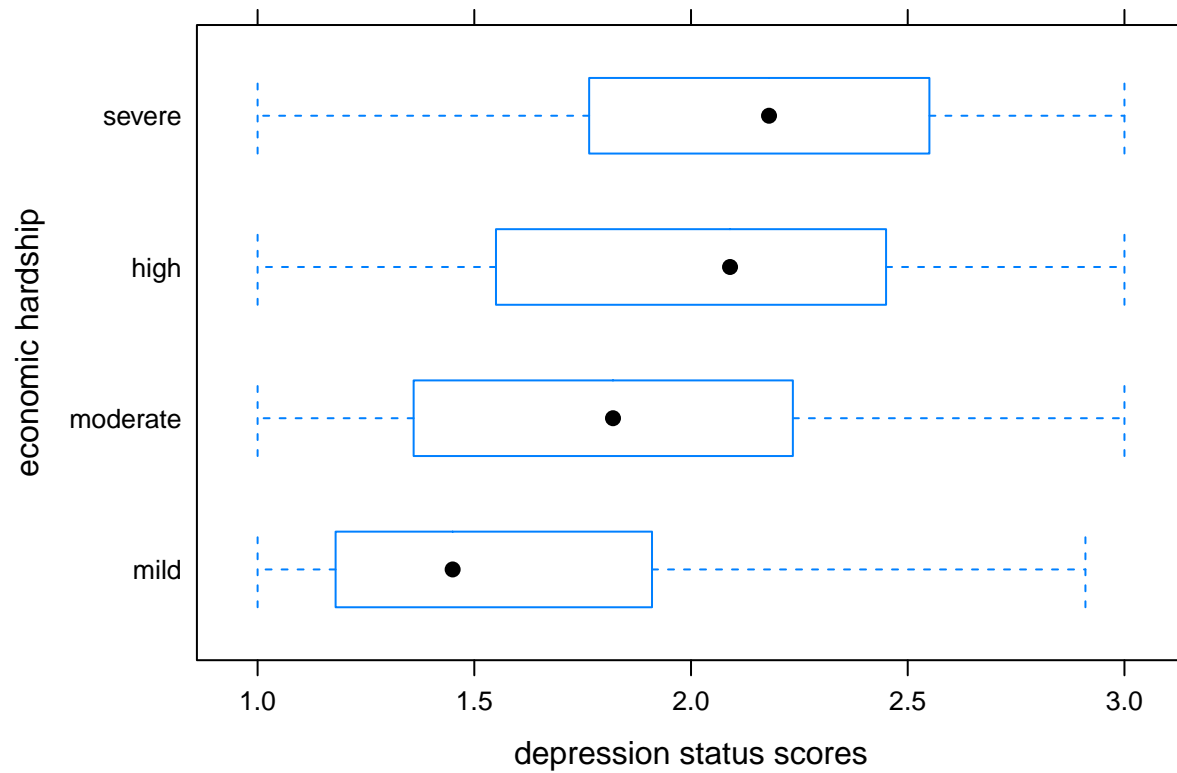


Figure 3: Boxplot of Depression Status and Education and ANOVA

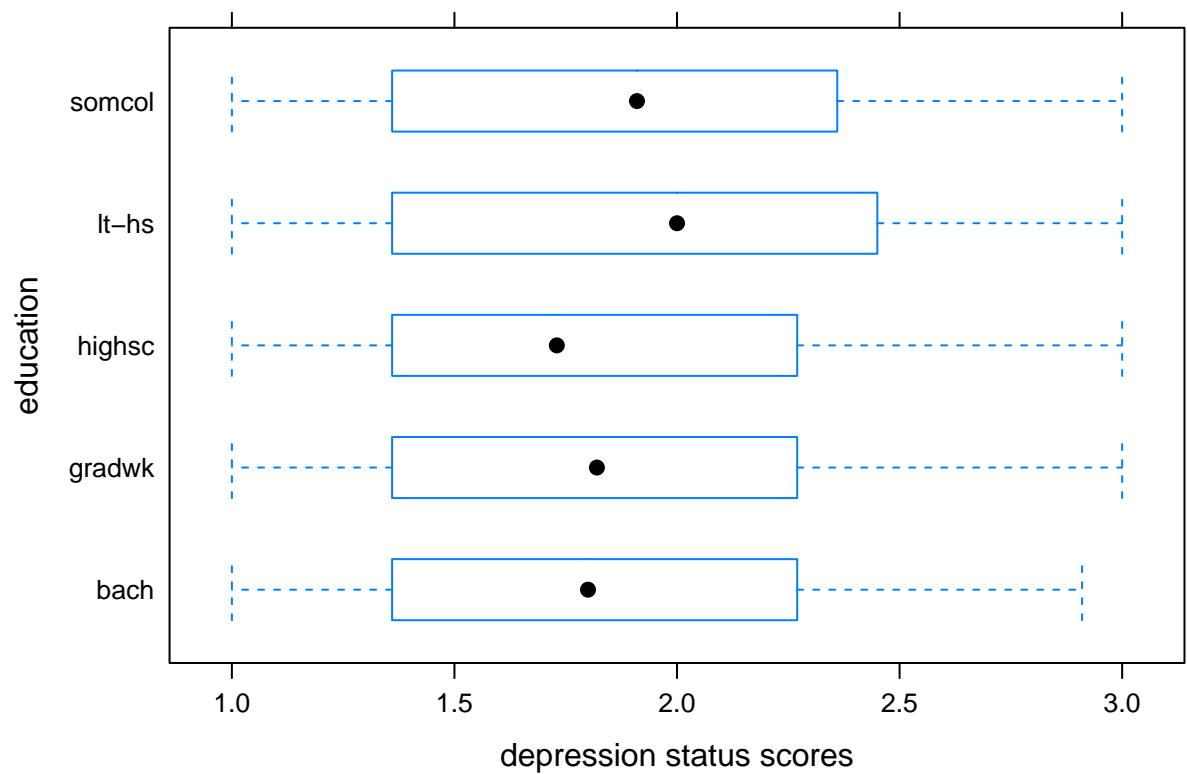


Figure 4: Boxplot of Depression Status and Marital Status and ANOVA

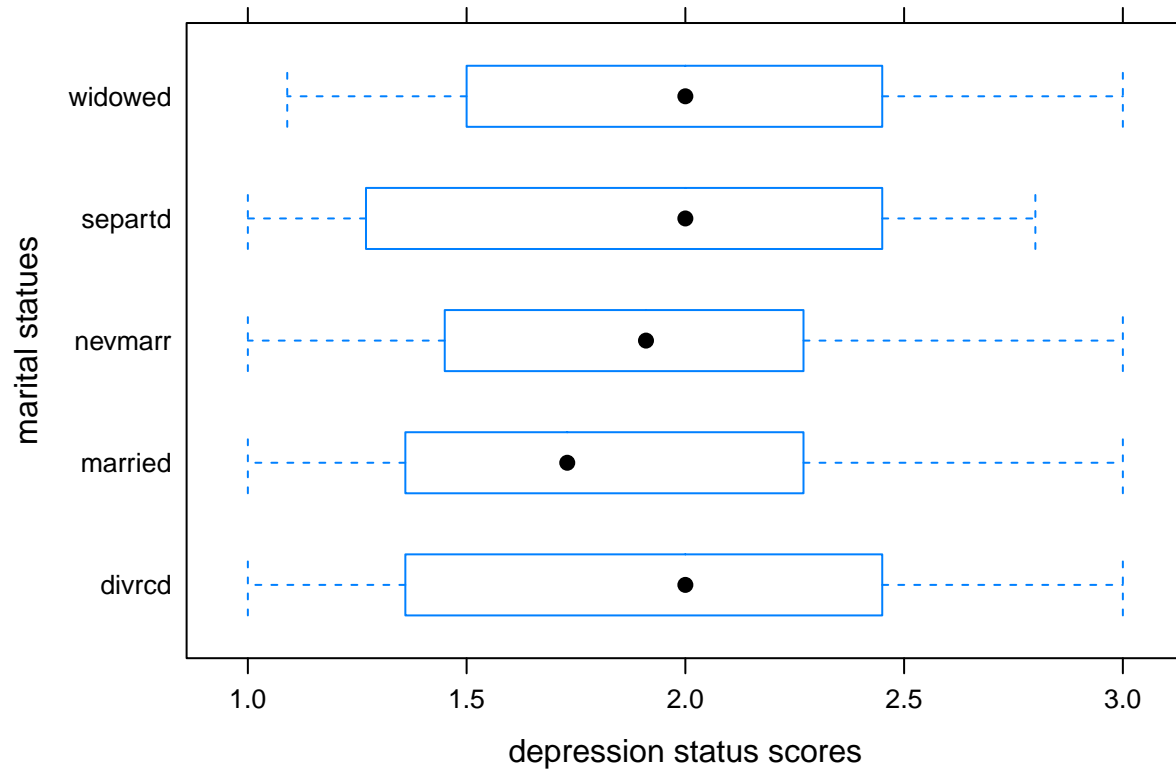


Table 1: Summary statistics of gender and depression scores

##	mean	sd	sample size	missing
## male	1.82	0.55	417.00	0.00
## female	1.92	0.58	482.00	0.00

Table 2: Results of t.test for gender and depression scores

##	t-statistics	obs.diff	in means	p-value	lower-bound of CI
##	-2.6235		-0.1000	0.0089	-0.1724
##	highr-bound of CI				
##	-0.0248				

Table 3: Summary statistics for economic hardship and depression scores

##	mean	sd	sample size	missing
## mild	1.59	0.52	189.00	0.00
## moderate	1.82	0.54	319.00	0.00
## high	2.01	0.56	267.00	0.00
## severe	2.13	0.51	124.00	0.00

Table 4: Results of ANOVA for economic hardship and depression scores

##	F-statistics	p-value
##	33.1700	0.0001

Note: p-value: 0.0001 actually means p-value < 0.0001

Table 5: Results of TukeyHSD for economic hardship and depression scores

##	difference in mean	p-value	lower-bound of CI
## moderate-mild	0.2224	0.0001	0.0953
## high-mild	0.4157	0.0001	0.2841
## severe-mild	0.5340	0.0001	0.3740
## high-moderate	0.1933	0.0001	0.0785
## severe-moderate	0.3116	0.0001	0.1651
## severe-high	0.1183	0.1799	-0.0321
##	highr-bound of CI		
## moderate-mild	0.3493		
## high-mild	0.5473		
## severe-mild	0.6939		
## high-moderate	0.3081		
## severe-moderate	0.4581		
## severe-high	0.2687		

Note: p-value: 0.0001 actually means p-value < 0.0001

Table 6: Summary statistics for education and depression scores

##	mean	sd	sample size	missing
## bachelors	1.8413	0.5276	146.0000	0.0000
## graduate	1.8399	0.5470	112.0000	0.0000
## high school	1.8407	0.5882	272.0000	0.0000
## less than high school	1.9970	0.6043	50.0000	0.0000
## some college	1.8988	0.5631	319.0000	0.0000

Table 7: Results of ANOVA for education and depression scores

##	F-statistics	p-value
##	1.193	0.313

Table 8: Summary statistics for marital status and depression scores

##	mean	sd	sample size	missing
## diverced	1.9437	0.5963	163.0000	0.0000
## married	1.8142	0.5577	408.0000	0.0000
## never married	1.9001	0.5495	279.0000	0.0000
## separated	1.8887	1.6156	30.0000	0.0000
## widowed	1.9689	0.5709	19.0000	0.0000

Table 9: Results of ANOVA for marital status and depression scores

##	F-statistics	p-value
##	2.0510	0.0854

Table 10: Results of linear regression for gender, economic hardship and age with depression scores

##	slope parameters	standard error	t-value	p-value
## intercept	1.3022	0.0883	14.7400	0.0001
## economic hardship	0.1878	0.0180	10.3970	0.0001
## gender	0.1215	0.0357	3.4000	0.0007

age -0.0017 0.0017 -1.0220 0.3071

Note: p-value: 0.0001 actually means p-value < 0.0001

VII. References

Butterworth, P., Rodgers, B., & Windsor, T. D. (2009). Financial hardship, socio-economic position and depression: Results from the PATH Through Life Survey. *Social Science & Medicine*, 69(2), 229-237. doi:10.1016/j.socscimed.2009.05.008

Crespo, L., López-Noval, B., & Mira, P. (2014). Compulsory schooling, education, depression and memory: New evidence from SHARELIFE. *Economics of Education Review*, 43, 36-46. doi:10.1016/j.econedurev.2014.09.003

LORANT V., CROUX C., WEICH S, DELIÈGE D., MACKENBACH J, ANSSEAU M. (2007). Depression and socio-economic risk factors: 7-year longitudinal population study. *The British Journal of Psychiatry*, 190(4), 293-298. doi:DOI: 10.1192/bjp.bp.105.020040

Madianos, M., Economou, M., Alexiou, T., & Stefanis, C. (2010). Depression and economic hardship across Greece in 2008 and 2009: two cross-sectional surveys nationwide. *Social Psychiatry and Psychiatric Epidemiology*, 46(10), 943-952. doi:10.1007/s00127-010-0265-4

Patten, S., Williams, J., Lavorato, D., Wang, J., Bulloch, A., & Sajobi, T. (2016). The association between major depression prevalence and sex becomes weaker with age. *Social Psychiatry and Psychiatric Epidemiology*, 51(2), 203-10.

Patten SB, Wang JL, Williams JV, Currie S, Beck CA, Maxwell CJ, El-Guebaly N (2006) Descriptive epidemiology of major depression in Canada. *Can J Psychiatry* 51:84-90

VIII. Peer Review Sessions Feedback Incorporation

After receiving feedback for our rough draft, we changes several aspects of our paper. First, we changed the title to account for the basis of our study, rather than just focus on a few of our many variables. In addition, we added details such as the fact that the study was non-randomly sampled, and that the depression score variable was originally categorical, to further clarify our writing. We also fixed our visuals so that the caption would be on the same page as the graph or table. We clarified our suggestions for further research, described bias, added ANOVA, Tukey and summary statistics in tabular form, and added more sources to back up the hypotheses for each of our variables.