

# The Effect Parental Divorce or Separation has on a Person Relationships

Isaac Casey

Monday October 19, 2020

## Abstract

In this study I will look at the divorce and marriage rates of people whose parents are also divorced to people whose parents are not. I use data from the General Social Survey on Family 2017 to conduct my analysis. Using a simple logistical regression model, I am able to find the results of the study. I find that people who have parents that are divorced are more probable to get a divorce in their marriages. They are also less likely to get married in general when compared to people whose parents are not divorced.

## Introduction

In this study, I will be examining the effects a parental divorce or separation has on a person's marital status. This topic of study peaked my interest as I am in this situation so I was inevitably curious to find out what the statistical outcomes are for a person whose parents are no longer together. Specifically I will be looking at the likelihood a person whose parents are separated or divorced also goes through a divorce compared to a person whose parents did not split. Furthermore I will also be comparing the probability a person gets married and such person's parents are divorced to the probability a person gets married and their parents are together still.

In the data section I will explain where I got the data used in this study from, and I will explain the variables I used and how I used them. In the model section I will explain the model I used and why I chose such model instead of others. In the results section I will display the outcomes produced by my model of the data and I will elaborate on the insights of said results in the discussion section. Lastly, I will state any weakness in this analysis and next steps for further analysis on this study in the weakness and next steps section respectively.

## Data

The data used in this survey is from the General Social Survey on Family 2017. The objectives of the GSS is "to gather data on social trends in order to monitor changes in the living conditions and well being of Canadians over time and to provide information on specific social policy issues of current or emerging interest." The target population of the GSS are individuals over the age of 15 who are non-institutionalized and live in Canada. Although the general social survey on family contains a wide range of variables and questions, this study only focuses on a few of those variables. The main variables I used are as followed:

Age when parents got divorced and age when parents separated: I used these two variables to find out whether a person's parents were together or not. For this study I did not take into account the age when the parents separated or divorced, I was merely concerned only if the parents were together or not. Have you ever been married: I used this variable to see whether or not the respondent was ever married. Number of divorces or separations: I used this variable to see if the person had ever gone through a divorce or a separation. A drawback of this variable is that one could go through separations from a common law relationship without getting married.

## Model

For this study, I could have used many different models. The two that I considered were a simple linear regression model or a simple logistical regression model. A reason for why simple linear regression would work because for both of objectives I was looking at, I wanted to predict the outcome on the basis of a single predictor variable, whether their parents were separated or not. I could also use the logistic regression model due to the outcome variable of both of my objects being binary. Have you ever been through a divorce or separation? Have you ever been married? Both of which are binary yes or no answers. Thus I decided to use a simple logistical regression model.

To compute the logistical regression model I went through the following steps. Firstly I cleaned the data obtained from the GSS survey modifying the R script template provided by Rohan Alexander and Sam Caetano while also making some minor adjustments to suit my specific objectives. Next I split my data into train and test data sets. The purpose of this is so I can test the accuracy and validity of my model using the whole data frame as my reference. Then I used the `glm()` function which allows me to compute a logistical regression model. To do so you need to specify the parameter family = binomial. Now that the model is made, I can predict the probability of my objectives, that is if someone has been in a divorce or if someone has ever been married based on the predictor of whether their parents are divorced. Next to test the accuracy of the model, I check the proportion of predictions that were correct.

## Results

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2    v purrr  0.3.4
## v tibble 3.0.3     v dplyr  1.0.2
## v tidyr  1.1.2     v stringr 1.4.0
## v readr  1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

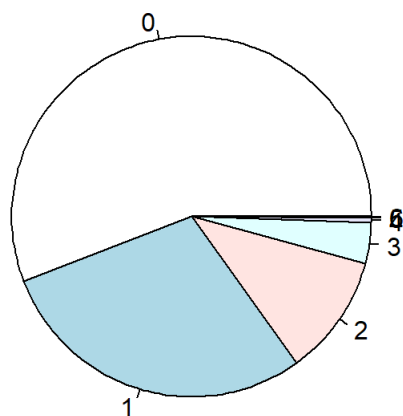
```
## Warning: package 'caret' was built under R version 4.0.3
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'caret'
```

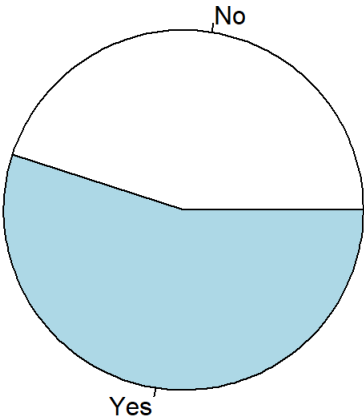
```
## The following object is masked from 'package:purrr':
##
## lift
```

**Figure 1**



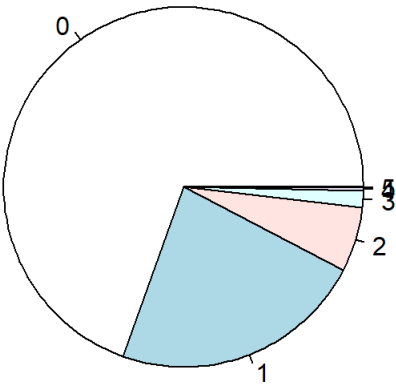
```
## [1] "This figure shows the number of divorces for people whose parents are also divorced"
```

Figure 2



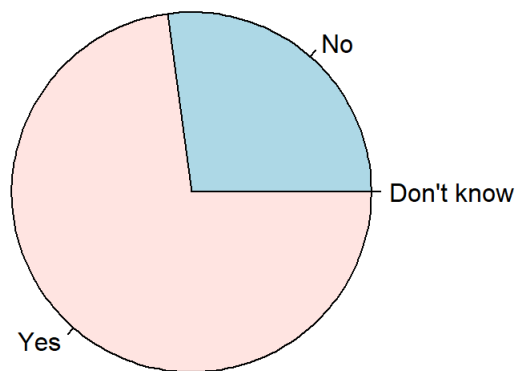
## [1] "This figure shows the ratio of people who got married and whose parents are divorced"

Figure 3



## [1] "This figure shows the number of divorces for people whose parents are also together"

**Figure 4**



```
## [1] "This figure shows the ratio of people who got married and whose parents are together"
```

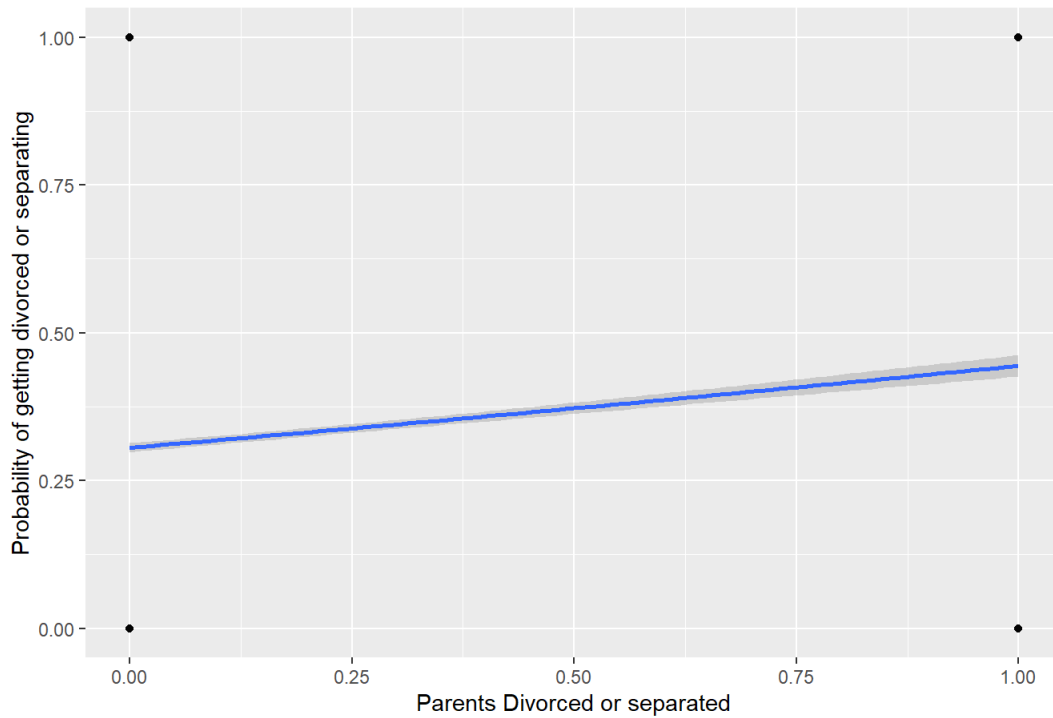
```
## [1] "A summary of the simple logistical regression for the probability if someone is divorced based on if their parents are divorced or separated.
From this we can see  $\log(p/(1-p)) = -0.81986 + 0.59516 \cdot \text{parents\_divorced}$ "
```

```
##
## Call:
## glm(formula = separate_or_divorce ~ parents_divorced, family = "binomial",
## data = train.data)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -1.0836 -0.8544 -0.8544 1.5394 1.5394
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.81986 0.01861 -44.06 <2e-16 ***
## parents_divorced 0.59516 0.04188 14.21 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 20904 on 16481 degrees of freedom
## Residual deviance: 20705 on 16480 degrees of freedom
## AIC: 20709
##
## Number of Fisher Scoring iterations: 4
```

```
## [1] "The model accuracy is 0.675485436893204"
```

```
## `geom_smooth()` using formula 'y ~ x'
```

Figure 5 - Logistic Regression Model, Probability of Getting Divorced or Separating



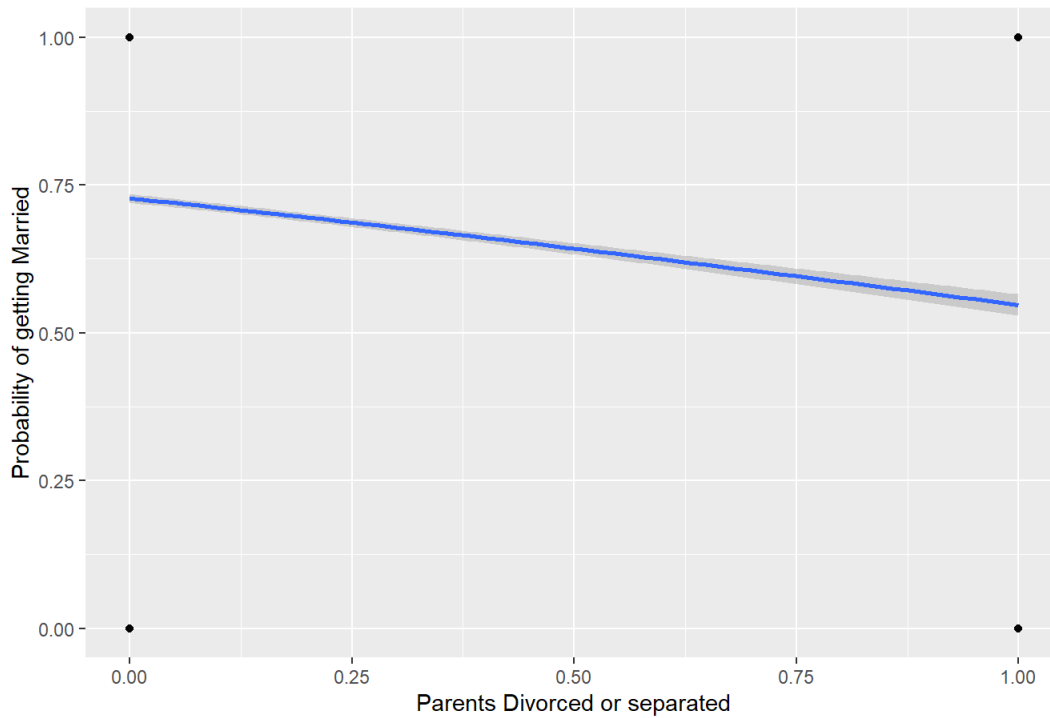
```
## [1] "A summary of the simple logistical regression for the probability if someone is married based on if their parents are divorced. From this we can see  $\log(p/(1-p)) = 0.98292 - 0.79320 \cdot \text{parents\_divorced}$ "
```

```
##
## Call:
## glm(formula = ever_married ~ parents_divorced, family = "binomial",
## data = train.data)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -1.6130 -1.2590 0.7974 0.7974 1.0980
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.98292   0.01926  51.03  <2e-16 ***
## parents_divorced -0.79320   0.04212 -18.83  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 20237 on 16477 degrees of freedom
## Residual deviance: 19890 on 16476 degrees of freedom
## AIC: 19894
##
## Number of Fisher Scoring iterations: 4
```

```
## [1] "The model accuracy is 0.702354940519544"
```

```
## `geom_smooth()` using formula 'y ~ x'
```

Figure 6 - Logistic Regression Model, Probability of Getting Married



## Discussion

The first objective of this study was to compare divorce rates between people whose parents are divorced or separated to people whose parents are still together. Just from comparing figure 1 to figure 3, we can see a big discrepancy between the two cases. Also from figure 6, it is clearly seen that people whose parents are divorced or separated are shown to have a higher probability of one or multiple divorces or separations in their future relationships. More specifically using the simple logistical model, the probability of someone getting a divorce or separation is found to be  $\log(p/(1-p)) = -0.81986 + 0.59516 \cdot \text{parents\_divorced}$ , where  $\text{parents\_divorced}$  equal 1 if the persons parents are divorced or separated or equal 0 if they are not. After calculating, the probability of getting a divorce or separating when your parents also have is 44.4% whereas the probability of divorce when your parents are still together is 30.6%. As represented in Figure 5. The accuracy of this model was found to be 67.5%.

The second objective of this study was to compare the marriage rates between people whose are divorced or separated to those whose parents are still together. By looking at figure 2 and figure 4, we can see that in both situations the marriage rate is greater than 50%, although it appears to be higher when the persons parents are still together. When looking at the simple logistical model produced from the data we can see the probability to get married is  $\log(p/(1-p)) = 0.98292 - 0.79320 \cdot \text{parents\_divorced}$ . Solving for probability  $p$ , we get the marriage rate for people whose parents are not together is 54.7% whereas for people whose parents are together, the probability of marriage is 72.8%. Figure 6 accurately depicts this. The accuracy of this model was found to be 70.2%.

## Weaknesses

A weakness of this study is that I didn't take age into consideration. Since the GSS survey is open to any individual 15 years and older, there can be caveat where multiple 15 year olds who are children of divorce are recorded and obviously they are not married or have gone through divorces. Thus skewing the results of the study. Another weakness is that the number of divorces and separations are in a single variable. Thus there can be the case of a person never being married but have a number greater than zero in the divorce or separation variable. This is due to a common-law relationship being counted toward the separation statistic. For this study to be more accurate, it'd need a separate variable for number of divorces and number of separations.

## Next Steps

A follow-up study to this study could be to take age into account. See if divorce or separation of parents at different age groups such as infant, adolescent, etc yields different outcomes in marital status. A study like this could show that divorces that happen when the person is still a child could impact their future relationships more than if they are a young adult. I would also be interested in if the terms on which the parents relationship ended in, whether it was mutual or not plays a factor in the person relationships. You could also take into account whether or not the individual has a sibling or not at the time of divorce/ separation. The possibility of going through this time with someone else could positively affect them thereby changing the results. This can play into a multiple logistical regression where more categorical or continuous variables are taken into account rather than just whether or not the parents are divorced/separated or not.

## References

Kassambara, et al. "Logistic Regression Essentials in R." STHDA, 11 Mar. 2018, [www.sthda.com/english/articles/36-classification-methods-](http://www.sthda.com/english/articles/36-classification-methods-)

essentials/151-logistic-regression-essentials-in-r/.

Smith, Tom W., Davern, Michael, Freese, Jeremy, and Morgan, Stephen L., General Social Surveys, 1972-2018 [machine-readable data file] /Principal Investigator, Smith, Tom W.; Co-Principal Investigators, Michael Davern, Jeremy Freese and Stephen L. Morgan; Sponsored by National Science Foundation. –NORC ed.– Chicago: NORC, 2019. 1 data file (64,814 logical records) + 1 codebook (3,758 pp.). – (National Data Program for the Social Sciences, no. 25).