

# How Age Factors Influence Duration of Marriage

Xuetong Tang, Yilin Lan, Lida Liu

3/20/2022

## Abstract

This paper generally investigates the relationship between people's current marriage time and age as well as whether they have children and whether it is their first marriage. By using dataset created by the Canadian 2017 General Social Survey (GSS), we construct some main variables and create histograms analyzing the distribution of the marriage time and age difference. Also, two scatter plots have been drawn to show the relationship between marriage time and difference age related components. At last, two box plots have been drawn to compare the medium of marriage time between whether they have children and whether they are in their first marriage. By further creation of the scatter plots and multiple linear regression model, we find that variables age of respondent, age of respondent when they got married and whether they have children have will affect the marriage time.<sup>1</sup>

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Data</b>	<b>3</b>
2.1	Methodology . . . . .	3
2.2	Data cleaning . . . . .	3
2.3	Data visualization . . . . .	4
<b>3</b>	<b>Result</b>	<b>7</b>
3.1	Linear relationship between marriage time and explanatory variables . . . . .	7
3.2	Model . . . . .	9
<b>4</b>	<b>Discussion</b>	<b>11</b>
4.1	What this paper done . . . . .	11
4.2	Findings and discussion . . . . .	11
4.3	Weakness of this paper . . . . .	13
4.4	Next Step . . . . .	13
<b>5</b>	<b>Appendix</b>	<b>14</b>
	<b>Reference</b>	<b>17</b>

---

<sup>1</sup><https://github.com/YilinLan/STA304Paper3>

# 1 Introduction

Nowadays, with the development of society, people care more about the marital status. There are many factors can influence the marital status, including age difference, education quality, life style, ect. Research shows that Preliminary results suggest that there does not appear to be a strong association between marital age difference and likelihood of divorce (Wilson and Smallwood 2008). On the other hands, analysis of age at marriage suggests that there may be an association between age difference and probability of divorce for individuals who marry above age 30 and are older than their spouse by more than ten years (Wilson and Smallwood 2008). The purpose of the paper is to investigate whether there is a relationship between Canadian citizens' current marriage time and age as well as whether they have children and whether it is their first marriage.

By using dataset created by the Canadian 2017 General Social Survey (GSS), we construct some main variables, including the response variable current marriage time, the age difference between respondent and respondent's spouse, age of respondent at the start of current marriage, as well as some categorical variable indicate whether the couple is the first marriage and whether they have children. We remove the answer "valid skip," "Don't know," "Refusal" and "Not stated" in the survey. For the variable age\_difference, we remove the leverage point and only select observations that has age\_difference less than 20. Then we create histogram analyzing the distribution of the marriage time and age difference. Also, two scatter plots have been drawn to show the relationship between marriage time and difference age related components. At last, two box plots have been drawn to compare the medium of marriage time between whether they have children and whether they are in their first marriage.

Furthermore, we construct scatter plot showing the relationship between marriage time and age at start of marriage and divided them by whether they have children. Also, a scatter plot showing the relationship between marriage time and age at start of current marriage and group by whether it is their first marriage. We find that those who have children generally have longer duration of marriage than those who don't have children. For respondents who start current marriage above 37 years old, those who have children generally have shorter duration of marriage than those who don't have children. At last, we construct a multiple linear regression model and find that variables age of respondent, age of respondent when they got married and whether they have children have will affect the marriage time.

All the data analysis in this paper uses R studio (R Core Team 2021) with tidyverse (Zhu 2021), patchwork (Pedersen 2020), ggplot2 (Wickham 2016) , kableExtra (Wickham et al. 2019) and dplyr (Wickham et al. 2021) packages.<sup>2</sup>

---

<sup>2</sup><https://github.com/YilinLan/STA304Paper3>

## 2 Data

### 2.1 Methodology

The target population includes all non-institutionalized persons 15 years of age and older, living in the 10 provinces of Canada. The survey uses a new frame, created in 2013, that combines telephone numbers (landline and cellular) with Statistics Canada’s Address Register, and collects data via telephone. Data are subject to both sampling and non-sampling errors.

The population of the data includes all persons 15 years of age and older, living in the 10 provinces of Canada, excluding residents of the Yukon, Northwest Territories, and Nunavut; and full-time residents of institutions. To carry out sampling, each of the ten provinces were divided into strata (i.e., geographic areas). There are 14 Census Metropolitan Areas (CMAs) were considered to be separate strata. The remaining CMAs were divided into 3 strata. At last, the non-CMA areas each of the ten provinces were grouped together and considered to be the rest 10 strata. In total, Canada had been divided in to 27 strata to carry out sampling. The citizens’ telephone numbers in use available to Statistics Canada from various sources were combined with Statistics Canada’s Address Register (AR) to form a new frame in year 2013. There are about 86% of available telephone numbers that could link to AR. The result 14% of telephone numbers were also included in the frame. When more than one telephone numbers were linked to a household, the first one was considered to be the best one. Then, each telephone number was assigned to a stratum within its province and a simple random sample without replacement of telephone numbers was performed in each stratum. In this survey, there are 91.8% of the selected telephone numbers reached eligible households. That is, there are at least one person aged 15 year or order. The target sample size for the survey was 20,000 while the actual number of respondents was 20,602. Data in the survey was collected through computer assisted telephone interviews (CATI). Respondents were interviewed in the official language of their choice. The calls were made from approximately 9:00 a.m. to 9:30 p.m. Mondays to Fridays using centralized telephone facilities in five of Statistics Canada’s regional offices. For participants that refuse to take the survey, interviewer explained the importance of the survey to the society and encouraged them to take part in it. For participants that were inconvenient to make the call at that time, the interviewer made an appointment with them and recalled at the appointment time. If there is no one at home, several call backs were made. The overall response rate for the survey was 52.4%.

The questionnaire included all the age group people and asked very specific questions. In the future investigation, people can create lots of variables and there are many aspects for people to focus on. Through the questionnaire is overall very specific, there might by some bias behind it when it comes to detailed personal questions. Some people may not able to completely remember their marriage details because it has been a long time since they get married and live in the household. For example, the questionnaire asked for the age of their second marriage. If it is a lone time ago, people may not able to give a correct response. The relationship between the marriage time and age is the aspect that we choose so that there may be some error behind the results. The overall result and discussion will not be largely affected.

### 2.2 Data cleaning

We select the following variables from the dataset for our investigation:

1. `current_marriage_time` (response variable): the total marriage time of respondent current marriage. We want to investigate the relationship between current marriage time and age. So that we choose these two variables, `ASEPMA0C`: Age of respondent at time of separation at current marriage, and `AGEMA0C`: Age of respondent at the start of current marriage. Then we take the difference between them and get the new variable called `current_marriage_time`.

Other important numerical explanatory variables include:<sup>3</sup>

---

<sup>3</sup><https://github.com/YilinLan/STA304Paper3>

2. age\_respondent: age of respondent at time of the survey interview. The original variable name is AGE\_C.
3. ageofmarriage\_respondent: age of respondent at the start of current marriage. The original variable name is AGE\_MA0C.
4. age\_difference: the age difference between respondent and respondent's spouse. We take the difference between variable aprma0c: age of respondent's spouse at start of current marriage and variable age\_ma0c: age of respondent at start of current marriage.

We also choose 2 dummy explanatory variables to discuss different conditions in our research problem:

5. children\_born: if the respondent have children born or adopted in his/her current marriage.
6. is\_first\_marriage: if the respondent's current marriage is his/her first marriage.

In the process of data cleaning, all "valid skip," "Don't know," "Refusal" and "Not stated" value in the survey answer are removed. For the variable age\_difference, we only select observations that has age\_difference less than 20 since those small number of observations with age\_difference larger than 20 are leverage points which could influence our linear model outcome too much.

There is no similar dataset because this survey was conducted from February 2nd to November 30th 2017. It takes a long time to collect the data and it include a large number of populations. The survey is not conducted every year so that there is almost no similar dataset.<sup>4</sup>

## 2.3 Data visualization

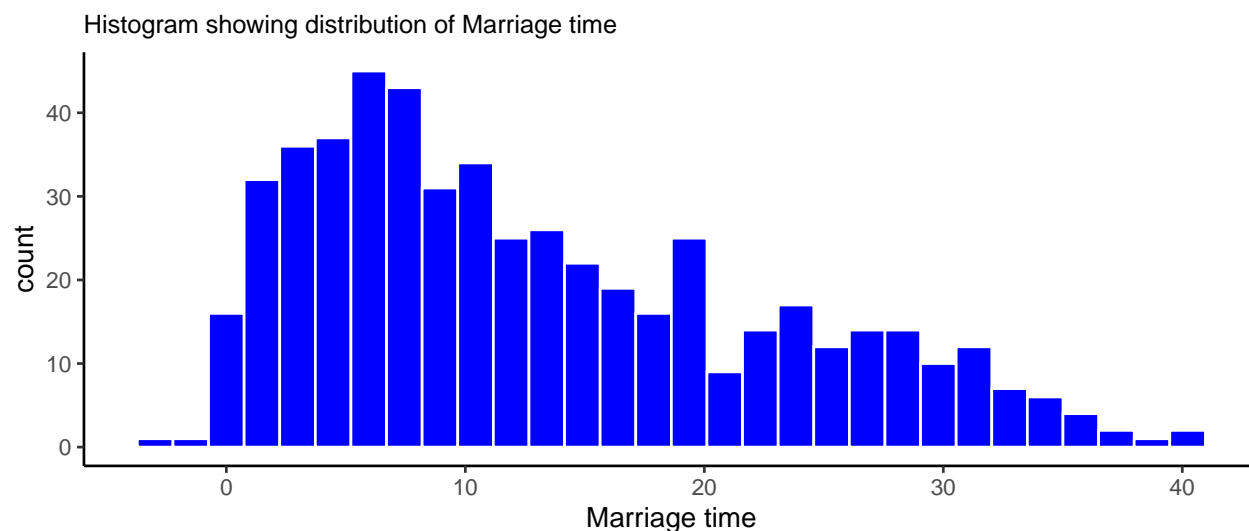


Figure 1: A histogram

Table 1: Summary table of Marriage time

median	mean	sd	IQR	max	min
10.8	13.21801	9.575435	14.1	40.7	-2.5

<sup>4</sup><https://github.com/YilinLan/STA304Paper3>

To give a brief glance of our sample, a histogram 1 of marriage time could be created to reveal the general distribution of marriage time. In this histogram, we could see that the distribution is right-skewed, which illustrates that the majority of marriage time are approximately from 2 years to 12 years, and the most marriage time of the chosen respondents is around 6 years. Also the table 1 shows that the average duration of marriage is 10.8 years.

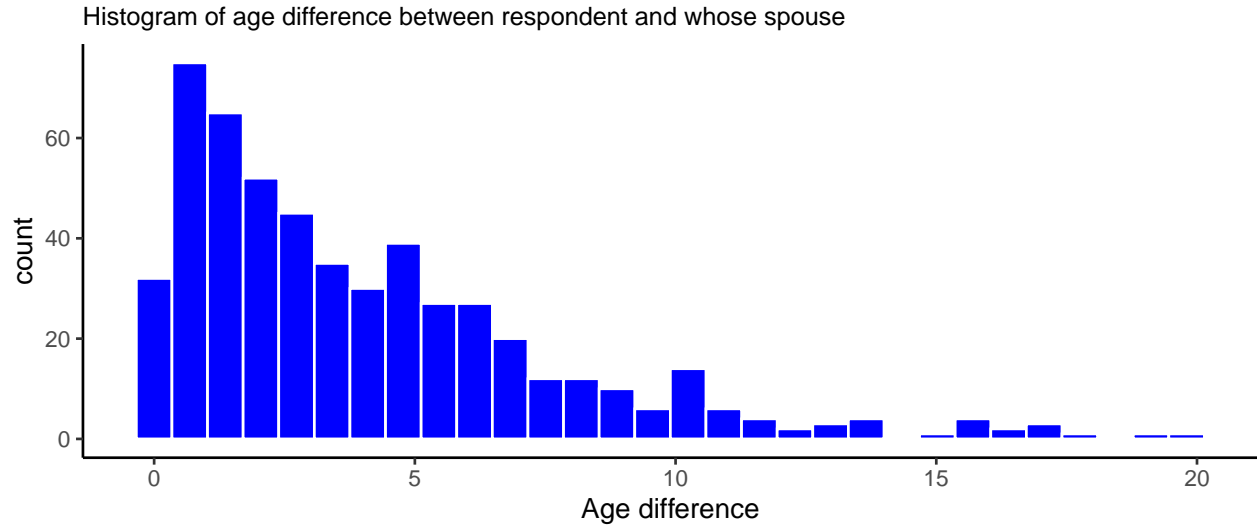


Figure 2: A histogram showing the age difference between respondent and whose spouse

Similarly, a histogram 2 of age difference could illustrate the distribution of age difference between the chosen couples as well. In this histogram, we could see that the distribution is right skewed as well. This phenomenon shows that the majority of age difference between chosen couples are around 1 to 5 years, with a peak of 1 year.



Figure 3: Scatterplots showing the relationship between marriage time and different age components

By drawing the scatter plot 3 between the marriage time and the age of respondent, the distribution of the<sup>5</sup>

<sup>5</sup><https://github.com/YilinLan/STA304Paper3>

relationship between these two factors could be demonstrated. According to the scatter plot between the marriage time and the age of respondent, we could see that there is a positive correlation between these two variables, which make sense in real life, so that our sample is solid for further analysis.

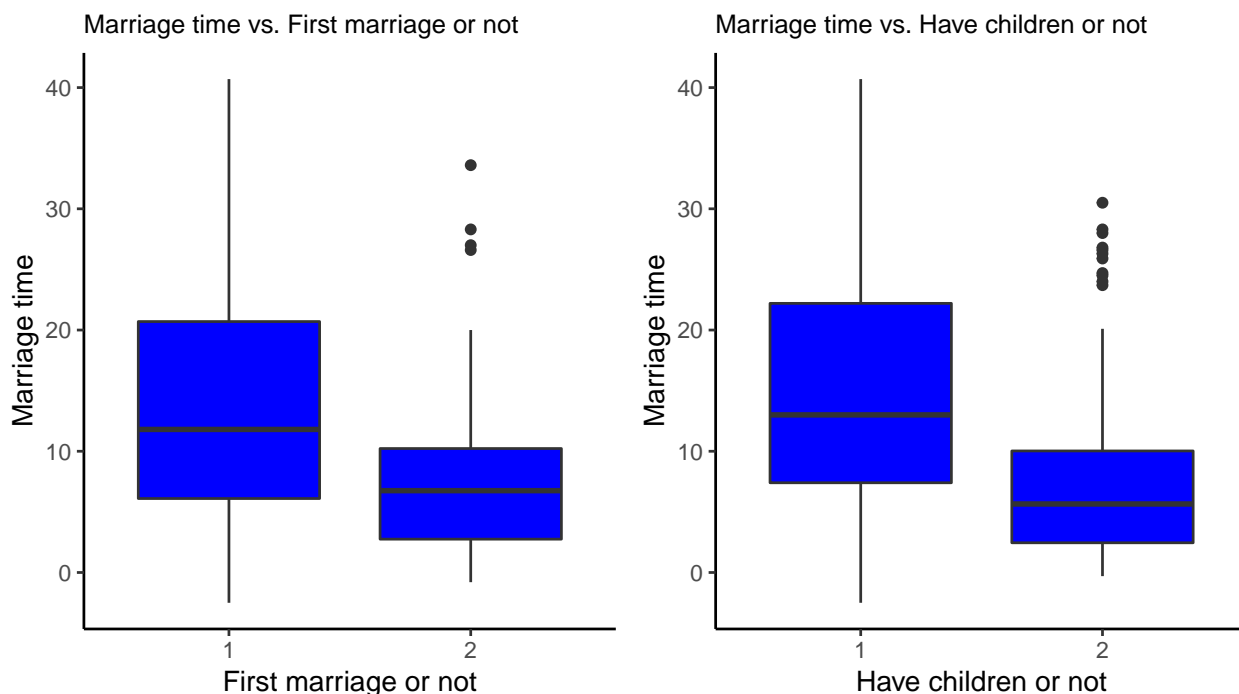


Figure 4: Boxplots showing the medium difference between marriage time and different categorical variables

According to the Figure 4 between marriage time and first marriage or not, we could see the difference of our sample between first marriage and not first marriage and the distribution of marriage time of these two types of marriage. For respondents who are in a first marriage, a median marriage time of 12 years is shown in the boxplot, and the more precise concentration of marriage time is from 6 years to 12 years. For respondents who are not in a first marriage, a median marriage time of 6 years is revealed, and the concentration of marriage time is from 4 years to 9 years, with several outliers indicating marriage time more than 25 years.

Similarly, the Figure 4 between marriage time and having children or not could illustrate the difference of our sample between having children or not having children and the distribution of marriage time of these two conditions. For respondents who are having children, a median marriage time of 13 years is revealed, and the marriage time is concentrated from 8 years to 12 years. For respondents who are not having children, a median marriage time of 6 years is demonstrated, and the concentration of marriage time in this case is from 4 years to 6 years, with several outliers of more than 23 years.<sup>6</sup>

<sup>6</sup><https://github.com/YilinLan/STA304Paper3>

### 3 Result

#### 3.1 Linear relationship between marriage time and explanatory variables

Marriage time by Age difference between respondent and whose spouse

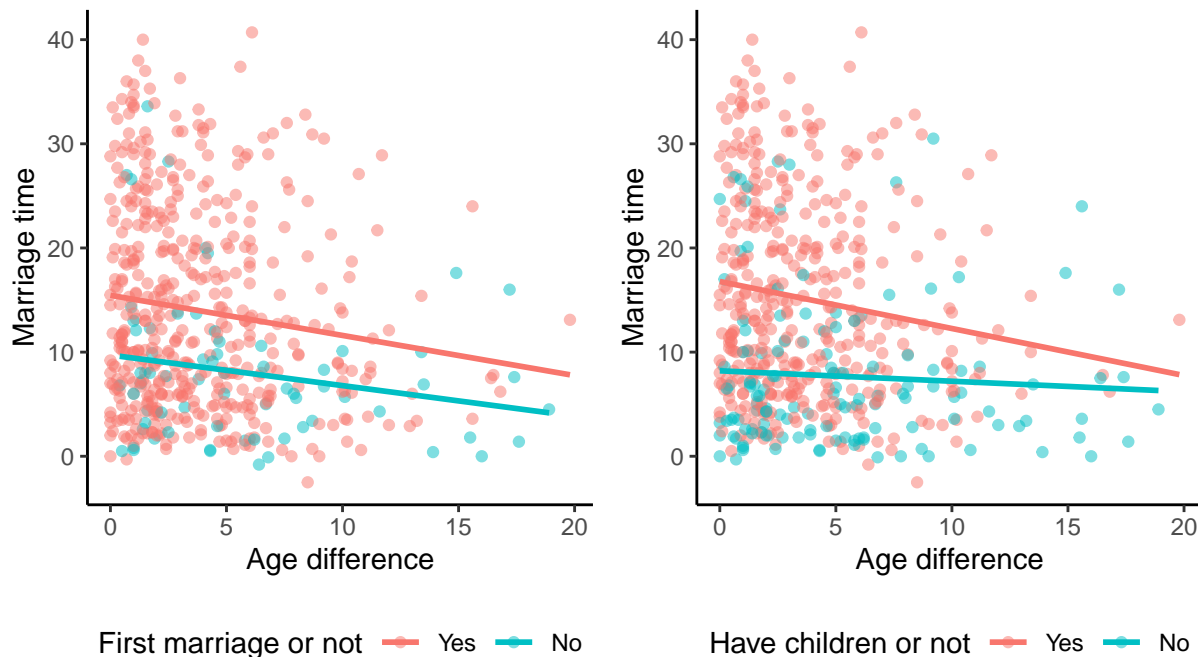


Figure 5: Scatterplots showing the relationship between marriage time and age difference with different categorical variables

By creating these two scatter plots 5 between marriage time and age difference, we can see a negative relationship. The graph on the left hand side shows the linear relationship with condition first marriage or not. The red line in the graph on the left hand side represent the group whose marriage is first marriage while the blue line represent the group whose marriage is not first marriage. From the fitted line on the scatterplot we can see that if the respondent is in the first marriage, their average marriage time is longer than the respondents who are not in the first marriage. The graph on the right hand side shows the linear relationship with condition having children or not. The red line in the graph on the right hand side represent the group who have children while the blue line represent the group who don't have children. By linear regression, the fitted line tells that couple have children generally have longer marriage duration than those who don't have children.

Over all, in general, we can conclude from both graph that when the age difference is smaller, couples' marriage time would be generally longer than the couples who have larger age difference under all the circumstances. Moreover, when couple don't have children, the line is very flat. Thus, it is quite interesting to see that when the couple don't have children born or adopted, the age difference is not a factor which affects marriage time much.<sup>7</sup>

<sup>7</sup><https://github.com/YilinLan/STA304Paper3>

### Marriage time by Age of respondent

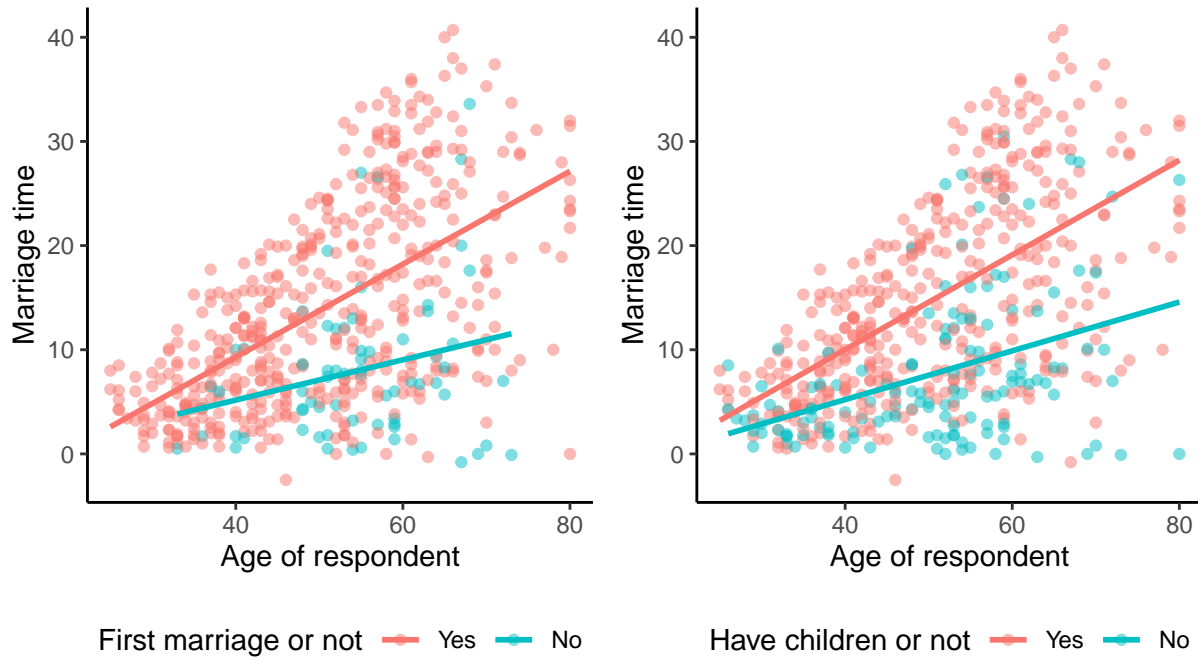


Figure 6: Scatterplots showing the relationship between marriage time and different age components with different categorical variables

In this case, by constructing these two scatter plots 6 between marriage time and age of respondents with conditions of first marriage or not and having children or not, the relationship between these two variables is clearly illustrated. The graph on left hand side discusses the linear relationship under circumstances that current marriage is first marriage or not while graph on left hand side discuss circumstances that couple have children or not. In this case, on both graphs, with the increase of the age of respondent, the marriage time of respondent are increasing as well, which make sense in real life. For same age respondents, those respondents who are in their first marriage generally have longer duration of marriage than those respondents who are not in their first marriage. Moreover, For same age respondents, those who have children generally have longer marriage time as well than those who don't have children.<sup>8</sup>

<sup>8</sup><https://github.com/YilinLan/STA304Paper3>



### Marriage time by Age at start of current marriage



Figure 7: Scatterplots showing the relationship between marriage time and age at start of marriage with different categorical variables

These two scatter plots 7 between marriage time and respondent's age at start of marriage with conditions of first marriage or not and having children or not could also be analyzed for more information. The graph on left hand side discusses the linear relationship under condition that current marriage is first marriage or not while graph on left hand side discuss conditions that couple have children or not. In this case, there are intersections in both two plots. For respondents who start current marriage under 37 years old, those who have children generally have longer duration of marriage than those who don't have children. On the other hand, For respondents who start current marriage above 37 years old, those who have children generally have shorter duration of marriage than those who don't have children. Thus, it is very interesting to see that having children or not influence older people's marriage in a different way compare to younger people's marriage.

Meanwhile, from the graph on the right hand side we can see that for respondents who start current marriage under 35 years old, those whose current marriage is their first marriage generally have longer duration of marriage than those who are not in their first time marriage. However, For respondents who start current marriage above 35 years old, those whose current marriage is their first marriage generally have shorter duration of marriage than those who are not in their first time marriage. Here, the gap between 2 group first marriage and not first marriage is not that large as the gap between group have children or not in graph on the left. This imply that the affect of factor first marriage on marriage time or not is not as large as the factor have children or not.

## 3.2 Model

Here, to have a better view on the whole picture of our research problem, we build a multiple linear regression model which conveys the relationship between marriage time and all explanatory variables we considered in our research including age of respondent, have children or not, first marriage or not, age difference between<sup>9</sup>

<sup>9</sup><https://github.com/YilinLan/STA304Paper3>

respondent and spouse and age of respondent at start of marriage. The equation of the model is shown as below:

$$(current\_marriage\_time)_i = \beta_0 + \beta_1(age\_respondent)_i + \beta_2(children\_born)_i + \beta_3(is\_first\_marriage)_i + \beta_4(age\_difference)_i + \beta_5(ageofmarriage\_respondent)_i + \epsilon_i$$

where:

$\beta_0$  is the term of intercept

$\beta_1 \dots \beta_5$  are the coefficients for each explanatory variable

$i = 1, \dots, n$  where  $n$  is the number of observations in the sample data

$\epsilon_i$  is the  $i^{th}$  error term for this regression model

Table 2: Summary of regression model

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	8.2165602	1.6288639	5.0443502	0.0000006
age_respondent	0.4422128	0.0227571	19.4318909	0.0000000
children_born2	-2.2193054	0.7336869	-3.0248673	0.0026088
is_first_marriage2	-0.3762604	0.9674343	-0.3889261	0.6974879
age_difference	-0.0660721	0.0803193	-0.8226185	0.4110970
ageofmarriage_respondent	-0.5808285	0.0474245	-12.2474359	0.0000000

Above is the summary table 2 of the model. We could interpret the estimated value of coefficients in the summary table as follows:

As age of respondent increases by 1 unit, the mean duration of marriage increase by 0.4422128 unit with all else been constant, this matches the positive relationship shown by scatter plot which we discussed before. As age difference between respondent and whose spouse increase by 1 unit, the mean duration of marriage decrease by 0.0660721 unit with all else been constant, which matches the negative linear relationship shown in the scatter plot. As age of respondent at the start of marriage increase by 1 unit, the mean duration of marriage decrease by 0.5808285 unit with all else been constant. This also matches the negative relationship in scatter plot.

The estimated coefficient value for two dummy variables in the model show the difference between groups. The estimated value of  $\beta_2$  is -2.2193054. This means that on average the difference between group “have children” and group “not have children” is 2.2193054. This value means that the predicted marriage duration for couples who have children is 2.2193054 years longer than couples who don’t have children with all else been constant. On the other hand, The estimated value of  $\beta_3$  is -0.3762604. This means that on average the difference between group “first marriage” and group “not first marriage” is 0.3762604. It is predicted that the marriage duration for respondent whose current marriage is whose first marriage is 0.3762604 years longer than respondent whose current marriage is not whose first marriage with all else been constant. This reflects the same result as scatter plots we have above. Couples who have children and whose current marriage is their first marriage tend to have longer marriage time.

What’s more, the p-value conducting by hypothesis test on each variable shows the significance of each explanatory variable in this model. Here, variable age\_respondent, ageofmarriage\_respondent and children\_born all have p-value less than 0.05, which shows that these predictors are statically significant in this model. In contrast, variable age\_difference and is\_first\_marriage has p-value larger than 0.05, which shows that these two variable are not statistically significant in this model. Thus, we could deduct that the age difference of couple and if the marriage is the first marriage does not affect marriage duration much.<sup>10</sup>

<sup>10</sup><https://github.com/YilinLan/STA304Paper3>

## 4 Discussion

### 4.1 What this paper done

The paper aims to investigate the relationship between marriage time and age with combination of whether people have children and whether it is the first marriage. To focus on this aspect, we set our response variable to be the current marriage time, the predictors we choose are age of the respondents, age of the respondents when they got married, age difference between respondents and their spouse, categorical variables indicate whether the respondents have children and whether it is their first marriage.

By using these variables, we first create two histograms showing the distribution of marriage time and age difference. Then, we create two scatter plots in order to show the relationship between marriage time and different age components. We find that the longer the marriage time, the younger that the respondent got married and the older the respondent's current age. After that, we construct two boxplots indicating the medium differences in marriage time between whether they have children and whether it is their first marriage. It can be seen that the for responses that don't have children and are not in their first marriage, the medium of marriage time is absolutely lower than response who have children and in their first marriage.

Above is the brief investigation of the relationship between response variable and predictors. For the next part, we start a further investigation by creating scatter plots with fitted line with the combination of response variable, numerical variable and categorical variable. First, we create two scatter plots indicate the relationship between marriage time and age difference and group by whether they have children and whether they are in their first marriage. The graphs show that there is a slightly negative relationship between marriage time and age difference. Also, as we mentioned above, for respondents who have children and are in their first marriage, the marriage time will be generally higher than those who do not have children and are not in their first marriage.

Second, we create two scatter plots to show the relationship between the marriage time and current age of respondents and group by whether they have children and whether they are in their first marriage. There is a positive relationship between marriage time and age of respondents. Third, intersections exists when we construct scatterplots to show the relationship between marriage time and age at start of the marriage. That means for respondents who start current marriage under 37 years old, those who have children generally have longer duration of marriage than those who don't have children. At last, to continue verify the result we find according to the scatter plots, we build a multiple linear regression model including all the variables we choose at the beginning. The p-value shows that variable current age of respondents, respondents age when they first got married and categorical variable indicating whether they have children are considered to be statistical significant.

### 4.2 Findings and discussion

According to the linear regression model, the scatterplots we made previously, our research question could be discussed. Due to the fact that the sample we analyzed are currently separated couples, we have avoided the influence of respondent's age when conducting our data analysis. Since all couples investigated are in a separation status, the duration of their marriage is not directly related to their age. Thus, these results truly reflect some marital issues in real life.

#### 4.2.1 Having Children, Age of respondent and age at start of marriage

From our analysis result, we find that in general, couples that have children born or adopted have longer duration of marriage. This make sense in our real life. According to the Article Psychosocial Factors That Influence Marital Couple Duration, children are negatively affected by divorce, both psychologically and socially (Wilson and Smallwood 2008). Nowadays, divorce become more and more common, and the divorce<sup>11</sup>

---

<sup>11</sup><https://github.com/YilinLan/STA304Paper3>

itself could be traumatic, not only for the couples, but also for their friends and children. For marital couples who are having children, children may become their representation and link of the marriage. If they divorce, their children will suffer from lots of pressure, like psychological pressure or financial pressure. Consequently, couples with children would be harder to decide to separate because of the existence of this link. On the other hand, according to the article of How to Manage the Higher Risk of Divorce When Baby Come Along, 67% percent of couples show a decline in satisfaction of marriage life after the arrival of the first baby (Brittle 2015). This phenomenon is easy to understand because the arrival of the first baby will make the marital life be more complicated, and things that would disturb the original marital life will occur as well.

Moreover, from our result, we find that age of respondent have positive relationship with duration of marriage. Elder people in our sample tend to have longer duration of marriage. This reflect the reality as well. With the longer time of the marriage, the marital issues within couples could be digested more. Additionally, the age of respondent starting their marriage is also influencing the marriage time. Based on our data analysis, we find that there is negative linear relationship between marriage time and age of respondent starting their marriage. We deduct that the reason behind may be when people are very old, they tend to be in the saperation status in their marriage. Thus, for those people who start marriage late, their marriage duration is shorter compare to those who start marriage when they are young. What’s more, we find that for respondents who start current marriage under 37 years old, those who have children generally have longer duration of marriage than those who don’t have children. This is very interesting and we haven’t find a reasonable explanation for this phenomenon. According to the article of What Is The Right Age To Get Married, the most appropriate age to get married and avoid divorcing in 5 years is around 28 years old to 32 years old. It makes sense as well, because 28 years old to 32 years old is an age that people could truly understand the difference between mature love and puppy love.

#### 4.2.2 Age difference and first marriage

According to the summary table of linear mixed model we built, the p-values of each variable has shown. It is shown that the variables: age of the respondent, age of respondent starting their marriage, and whether the respondent having children or not are all statistically significant to our model. On the other hand, the variable age difference between respondent and whose spouse and the variable whether the respondent is in a first marriage are not are not statically significant in our model, which indicates that those two are not main factors influencing the duration of marriage.

Although the result of our model indicate that whether the respondent is in a first marriage or not is not a main factor that influence the duration of the marriage, according to the article of How Long Does the Average Marriage Last, the duration of the first marriage in USA is only 7 years, and most people who divorce in first marriage has 60 percentage of probability to divorce in the second one, and 73 percentage for a third marriage (Writer and Marshall 2020). It is astonishing that, nowadays, the duration of marriage wound not last for a long time for some of the couples in USA, and the second marriage will not probably last for a long time as well. In real life, The factor of whether respondent is in a first marriage or not is going to influence duration of marriage as well.

As for the factor age difference, according to the scientific article Age difference at marriage and divorce, indeed, the age difference in a marriage isn’t a crucial factor of divorce (Wilson and Smallwood 2008). It makes sense in real life, because age difference is a factor of deciding to get married or not, but not a factor of deciding to divorce. According to the article of What Is A Best Age Difference For A Successful Marriage, it states that age gaps work differently on different couples. Some people will feel right with an age gap of 2 years, while some are looking for an age gap of 10 years (“What Is the Best Age Difference for a Successful Marriage?” 2020). Indeed, the age gap seems to be an indivual stuff for different couples.<sup>12</sup>

---

<sup>12</sup><https://github.com/YilinLan/STA304Paper3>

### 4.3 Weakness of this paper

There are some limitations in the data. First, the sample data we have is obtained from survey. The survey is answered voluntarily. So we cannot ensure the accuracy of the data collected. Second, there are non-response problem in the survey.

There are also lots of limitations in our regression model. First of all, the variable age difference doesn't include the gender information which shows whether male is elder, or female is elder. In this case, the sexual difference could influence the result. In real life, in some culture, female tend to choose elder spouse and some tradition believes that elder male spouse take care of their wife better. Thus, in those culture, gender information on age difference between respondent and spouse may have some effect on the duration of marriage. Secondly, the children in this case are either born or adopted, so that it doesn't include the existence of stepchild in different marital conditions. So we are not able to tell exactly if there is any children in the family. Moreover, here we are conducting an observational study, all these relationship we discussed are associations. We cannot tell whether having children would lead to longer duration of marriage, or the longer marriage lead to birth or adoption of children. Couples tend to have children after very careful consideration. Many couples tend to have children two or more years after getting married. For this reason, marriage time may also have an effect on having children. Thus, this cause-and-effect conclusion is very hard to be defined.

### 4.4 Next Step

For the further study, we plan to expand our model to include more considerable variables that may influence the duration of marriage. Moreover, instead of only focus on the duration of marriage, we want to set some criteria in order to measure the quality of a marriage. For example, how happy are the couples in the marriage. This require more information on respondent's personal life and highly depend on respondent's mood and personal feelings. Thus, to find a good way to collect these private information could be a big barrier for future analysis on this research problem.

What's more we could analysis on children of one family or children in one marriage. By analyzing the children number, children's education level and some more information collected, our further discussion could focus more on the relation ship between children and marriage. Children age, children number, children have pet or not and the reason why have children are all considerable variables of interest for our further study.<sup>13</sup>

---

<sup>13</sup><https://github.com/YilinLan/STA304Paper3>

## 5 Appendix

The supplementary survey is intended to get a more accurate result by considering other factors in the marriage. For the further study, investigators are able to collect more data and consider different aspect that may influence the marital status. Contact information: eileen.lan@mail.utoronto.ca The questions include:

1. Age when respondent has the first child in the first marriage. A. less than or equal to 15 years old B. 16~25 years old  
C. 26~35 years old  
D. 36~45 years old  
E. 46~55 years old  
F. older than or equal to 56 years old  
G. Valid skip  
H. Don't know  
I. Refusal
2. Age when the respondent has the first child in the second marriage. A. less than or equal to 15 years old  
B. 16~25 years old  
C. 26~35 years old  
D. 36~45 years old  
E. 46~55 years old  
F. older than or equal to 56 years old  
G. Valid skip  
H. Don't know  
I. Refusal
3. Age when respondent has the second child in the first marriage. A. less than or equal to 15 years old  
B. 16~25 years old  
C. 26~35 years old  
D. 36~45 years old  
E. 46~55 years old  
F. older than or equal to 56 years old  
G. Valid skip  
H. Don't know  
I. Refusal
4. Age when the respondent has the second child in the second marriage. A. less than or equal to 15 years old  
B. 16~25 years old  
C. 26~35 years old  
D. 36~45 years old  
E. 46~55 years old  
F. older than or equal to 56 years old  
G. Valid skip  
H. Don't know  
I. Refusal
5. The current highest education level of respondent's children. A. Lower than elementary school  
B. Elementary school  
C. Middle school  
D. High school<sup>14</sup>

---

<sup>14</sup><https://github.com/YilinLan/STA304Paper3>

- E. Undergraduate
  - F. Graduate
  - G. Higher than graduate
  - H. Valid skip
  - I. Don't know
  - J. Refusal
6. The purpose for the respondent of having the first kid in the first marriage. A. To have someone take care of you when you are old B. You like kids  
C. You parents asked you to do so  
D. You accidentally had a kid  
E. More than one of the options is satisfied  
F. Valid skip  
G. Don't know  
H. Refusal
7. The purpose for the respondent of having the second kid in the first marriage. A. To have someone take care of you when you are old  
B. You like kids  
C. You parents asked you to do so  
D. You accidentally had a kid  
E. More than one of the options is satisfied  
F. Valid skip  
G. Don't know  
H. Refusal
8. Does the respondent have pets (any kind) at the first marriage? A. Yes  
B. No  
C. Don't know  
D. Refusal
9. Does the respondent both have children and pets at the first marriage? A. Yes  
B. No  
C. Don't know  
D. Refusal
10. For respondent who does not have pets in the first marriage, does the marriage goes on until now? A. Yes  
B. No  
C. Don't know  
D. Refusal  
E. Valid skip
11. If respondent's parents both alive at the first marriage, do respondent's parents agree the first marriage?  
A. Yes  
B. No  
C. Don't know  
D. Refusal  
E. Valid skip

An URL for the supplementary survey: <https://forms.gle/sXo2h5nDp51ekVos6> Thanks for answering the above questions and your answers will definitely help us on the research.<sup>15</sup>

---

<sup>15</sup><https://github.com/YilinLan/STA304Paper3>



## Reference<sup>16</sup>

- Brittle, Zach. 2015. “When Three’s Not the Charm: How to Manage the Higher Risk of Divorce When Baby Comes Along.” <https://www.washingtonpost.com/news/inspired-life/wp/2015/06/30/when-threes-not-the-charm-how-to-manage-the-higher-risk-of-divorce-when-baby-comes-along/>.
- Pedersen, Thomas Lin. 2020. *Patchwork: The Composer of Plots*. <https://CRAN.R-project.org/package=patchwork>.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- “What Is the Best Age Difference for a Successful Marriage?” 2020. <https://timesofindia.indiatimes.com/life-style/relationships/love-sex/what-is-the-best-age-difference-for-a-successful-marriage/photostory/87910591.cms>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Wilson, Ben, and Steve Smallwood. 2008. “Age Differences at Marriage and Divorce.” [https://www.researchgate.net/profile/Steve-Smallwood/publication/23168277\\_Age\\_differences\\_at\\_marriage\\_and\\_divorce/links/09e4150b3295d2ba55000000/Age-differences-at-marriage-and-divorce.pdf#page=17](https://www.researchgate.net/profile/Steve-Smallwood/publication/23168277_Age_differences_at_marriage_and_divorce/links/09e4150b3295d2ba55000000/Age-differences-at-marriage-and-divorce.pdf#page=17).
- Writer, LawInfo, and Ally Marshall. 2020. “How Long Does the Average Marriage Last?” <https://www.lawinfo.com/resources/divorce/how-long-does-the-average-marriage-last.html>.
- Zhu, Hao. 2021. *kableExtra: Construct Complex Table with ‘Kable’ and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.

---

<sup>16</sup><https://github.com/YilinLan/STA304Paper3>