To study if the factors influencing the dating preferences are different for men and women

Sneha Banda

# 1. Abstract

The research is a part of an experimental study, conducted to identify differences in mate selection for both the genders. The data has been generated from a speed dating experiment conducted with independent measures. The primary question of interest is to determine if the factors influencing the dating preferences are different for men and women.

From a group of volunteers, subjects have been selected at random and were made to experience 4-minute dates. At the end of the speed date, subjects were made to fill up a closed ended questionnaire, having to rate the participant on six pre-defined parameters and whether they would like to go on a second date or not.

Our research suggests that the factors influencing dating behavior of men and women are different. Men prefer women who are attractive, intelligent and fun to be with. They do not prefer women who are more intelligent than themselves and also ambitious.

On the other hand, women prefer men who are fun to be with, intelligent and have mutually common interests.

# 2. Introduction

Choosing a life partner is one of the most critical decisions a person can take in his/her life. In the modern day, this decision is made after careful evaluation of the potential partner, by investing lot of time into the process of understanding and evaluating the relationship, which is referred to as “Dating”. With the increase in financial independence of women, it seems like there has been a gender role overlap happening in the society. In the ancient days, men were responsible for sourcing the resources and women were responsible for nurturing the family. However, these gender roles have changed drastically, at least over the last 30 years.

In this study, instead of studying the dating pattern, there is enough scope for us to understand the factors influencing the decision to date.

# 3. Literature Review

There are large number of research papers on the subject of marriage and mate selection. Many of these papers have considered the cultural diversities and economic conditions and their effect on the mate selection.

Belot and Francesconi (October 2006) have assessed the importance of isolating individual preferences and opportunities in dating behavior. Preferences are when the partners choose each other, irrespective of different social status and opportunities are the cases when the people, belonging to similar social status, meet each other in informal gatherings and interact more. The paper concludes that the role of opportunities is larger than that of individual preferences.

Townsend (1989) and Townsend & Roberts (1993) examined in their research that the increasing economic independence of women, did not change their traditional approach to mate selection and marital goals. The higher the financial status of the woman, the higher the standards they set for the potential partners, thereby reducing the pool of acceptable mates. Therefore, it gets difficult to find a suitable mate.

Weisfeld, Russell, Weisfeldand and Wells (1992), have concluded that the couples in the United States tend to be similar and the more similar they are, the more stable their relationships are. They also observed that the females find dominant men more attractive and men give more importance to physical attributes.

The above-mentioned research papers shed light on the previously established theories that claim that the men value physical attributes more and women look for factors like affluent backgrounds, intelligence and ambition are more valued in men.

The earlier research papers found the influencing factors affecting the mate selection behavior but they had limitations with respect to determining the extent of the contribution of each of these factors for both the genders.

# 4. Research Objective

The objective of the research is, “To study if the factors influencing dating preferences are different for both men and women”

# 5. Hypothesis

The null hypothesis that is being tested is that,

H0 : There are no differences in the factors influencing the dating preferences for both men and women.

This would mean that the factors which are important for choosing a suitable mate, are common for both the genders.

H1 : Different factors influence the dating preference of men and women.

# 6. Research Methodology

Primary research has been done in the form of an experimental study with Independent measures, conducted by Fisman, Iyengar, Kamenica and Simonson (2006) for their research on the subject. In this study, around 5000 men and women have been randomly selected from a larger set of volunteers. Each subject was made to participate in a 4-minute first date with potential partners. At the end of the date, the subject rates the date on the scale of 1 to 10, over 6 identified important factors. The decision variable is a binary variable, which tells if the subject would like to have a second date with the partner or not.

Before the experiment commenced, the subjects were asked to rate themselves on the chosen set of explanatory variables for reference.

The ratings on the explanatory variables along with the decision variable, will lead the research to identify if the contributing factors are different for both the genders.

# 7. Solution

The solution to the problem statement can be broken down into five subparts.

1. Exploratory Data Analysis – Data preprocessing:

This step is essential to check if the data is ready for analysis. Identifying the requirement for variable transformation, the outlier and missing values treatment, evaluating the skewness of data, were performed in this stage.

1. Sampling of the data:

Each subject had multiple dates and the participant was different each date. Therefore, stratified sampling method was implemented in choosing the random sample from the population, where two dates of each subject have been randomly sampled for the further analysis. Having two observations for each person will also control for variance caused by an individual’s personal bias while decision making.

1. Data partitioning:

The problem statement required to find the influencing factors for both men and women, therefore data is divided based on the gender variable. Further analysis would be performed separately on these subsets.

1. Building Individual Models:

For each subset, the full model is built, with all the explanatory variables. To use model selection process through forward selection, backward elimination or step-wise selection, which compare models using statistics like AIC, BIC or Cp.

1. Testing Individual Models:

The significance of the explanatory variables in both the best models, is tested using t-statistics. This will give us the significant factors for both the genders.

# 8. Analysis

## 8.1 Data Preprocessing

Data preprocessing stage includes the following tasks

## 8.1.1 Identifying the relevant variables and their datatypes

Of all the variables in the dataset, 10 explanatory variables were selected for the research. The “decision” variable is the dependent variable. The complete list of these variables along with their definitions is available in the Appendix exhibit.

To derive more insights from the data, variables mentioned in *Table1* were created. These three variables were also included in the modeling.

Table 1

|  |  |  |  |
| --- | --- | --- | --- |
| S.No | Variable | Values | Description |
| 1 | higher\_ambition | 0 or 1 | 1 if the participant has higher ambition than the subject |
| 2 | higher\_attraction | 0 or 1 | 1 if the participant has higher rating on attraction than the subject |
| 3 | higher\_intelligence | 0 or 1 | 1 if the participant has higher rating on intelligence than the subject |

## 8.1.2 Descriptive analysis of the variables

The descriptive analysis of the basic numerical variables is mentioned in *Table2.*

Table 2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | Min | Q1 | Median | Q3 | Max | Mean |
| Attrp | 0 | 5 | 6 | 7 | 10 | 5.884 |
| Sincp | 0 | 6 | 7 | 8 | 10 | 7.191 |
| Intelp | 0 | 6 | 7 | 8 | 10 | 7.358 |
| funp | 0 | 5 | 6 | 8 | 10 | 6.281 |
| ambp | 0 | 6 | 7 | 8 | 10 | 6.744 |
| sharep | 0 | 4 | 5 | 7 | 10 | 5.428 |

All the numerical explanatory variables are the bound between 0 and 10, as they are the ratings given by the subjects. The median scores of these variables are between 5 and 7. The means of these variables are mentioned in *Table2.*

The descriptive data analysis for the categorical variables is mentioned in *Table3.*

Table 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Count | | Proportion | |
| 1 | 0 | 1 | 0 |
| Gender | 454 | 436 | 51% | 49% |
| Samerace | 338 | 552 | 38% | 62% |
| Decision | 362 | 528 | 41% | 59% |
| Higher\_ambition | 234 | 656 | 26% | 74% |
| Higher\_attraction | 207 | 683 | 23% | 77% |
| Higher\_intelligence | 156 | 734 | 18% | 82% |

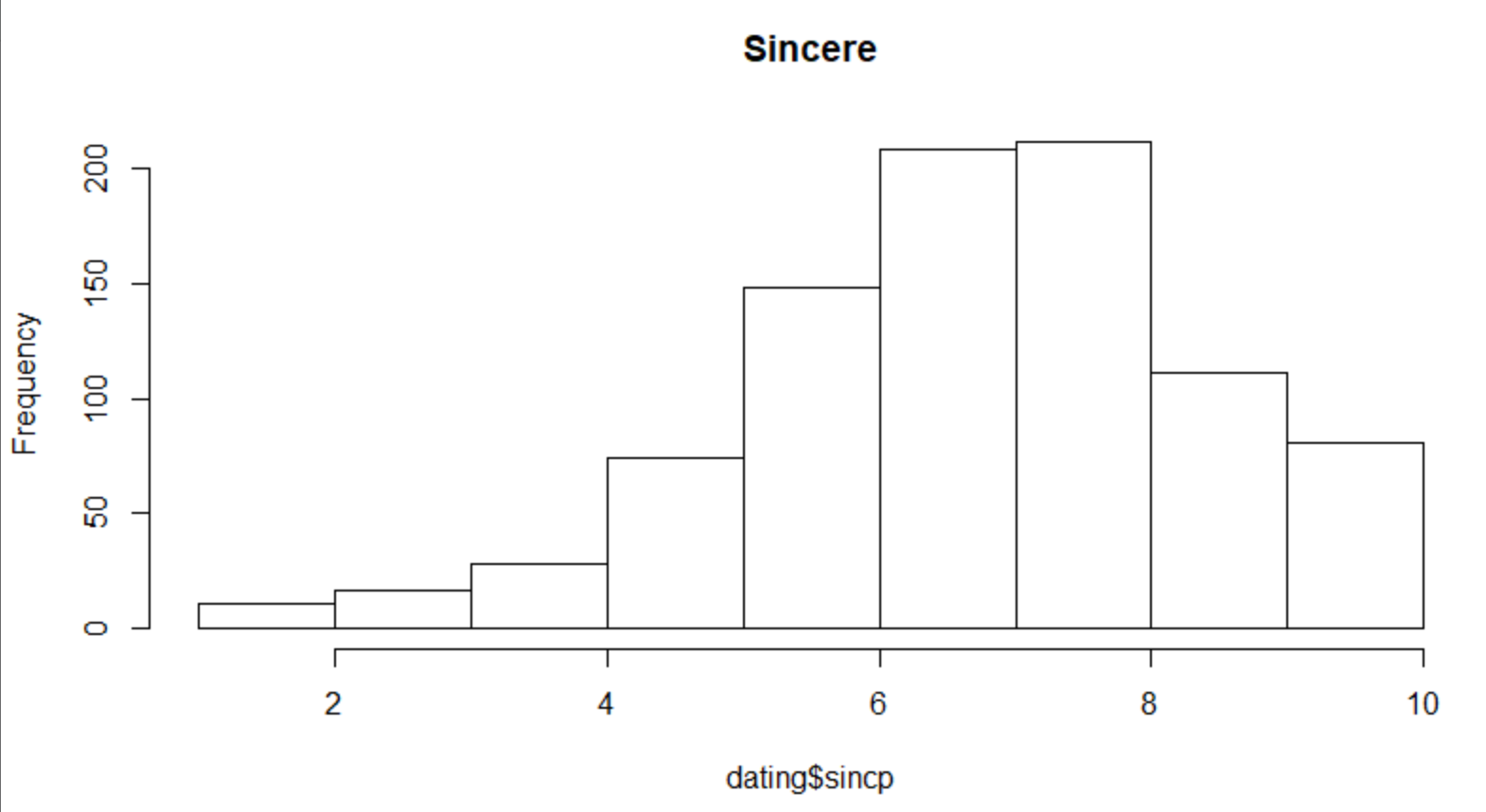
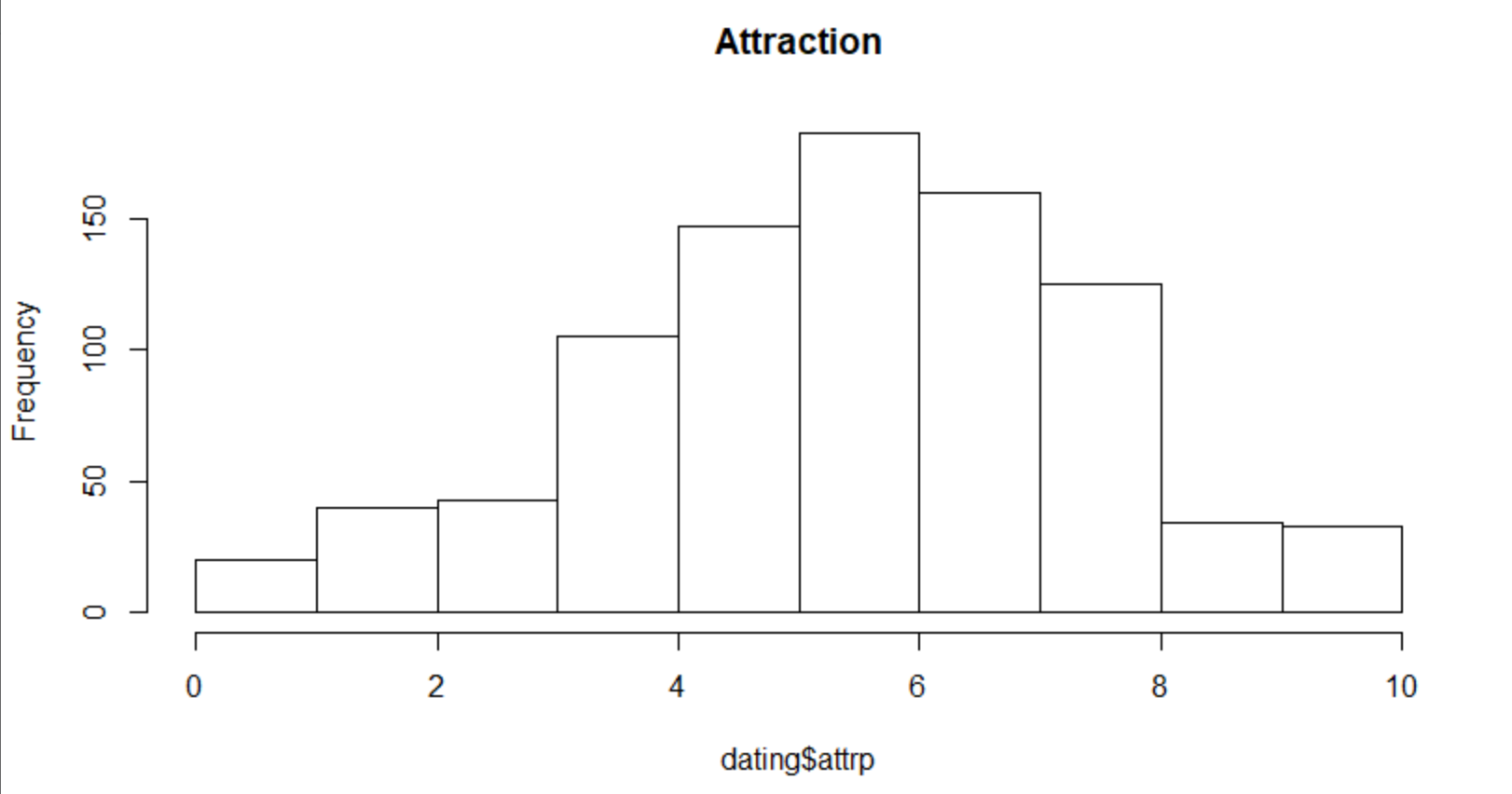
As mentioned in the above table, the data of female and male subjects is balanced, with 51% of the observations belonging to the male respondents and 49% corresponding to the female respondents. Out of the dates conducted within the sample, 38% were between people who belonged to the same race, the rest 62% belonged to couple belonging to different races.

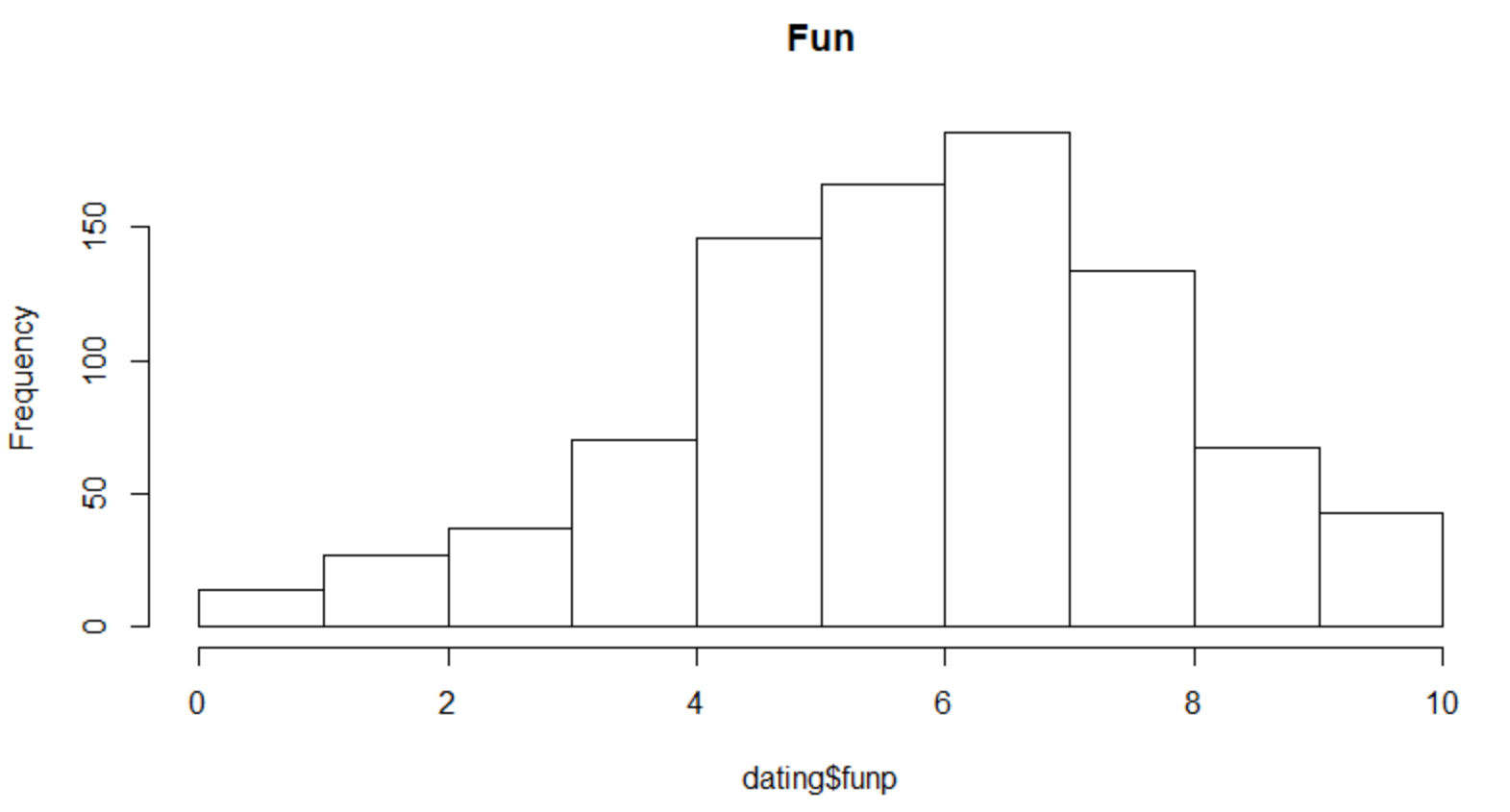
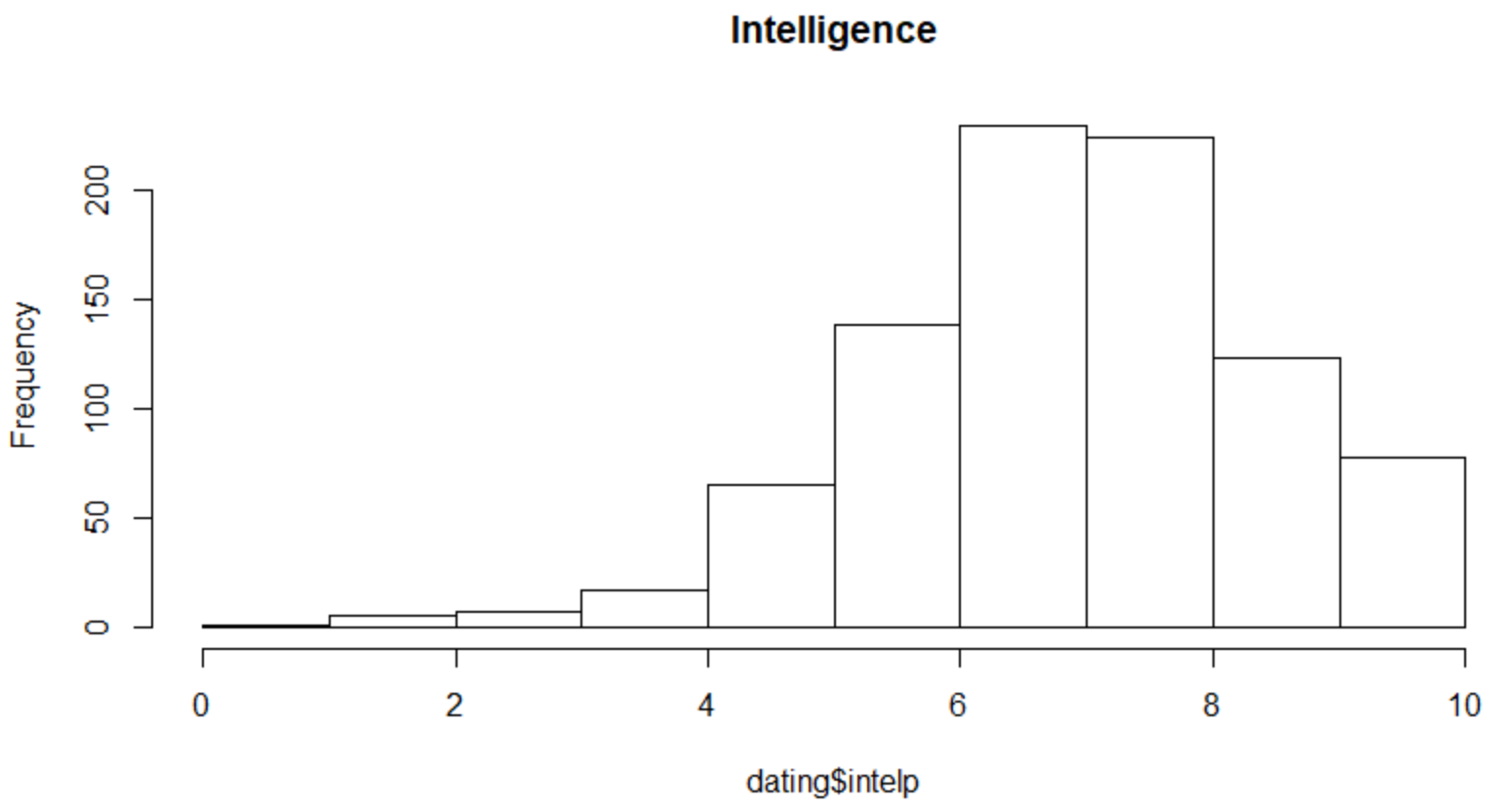
For the dependent variable “Decision”, 41% of the respondents have agreed for a second date and the remaining 59% did not want to go for a second date with the respective participants.

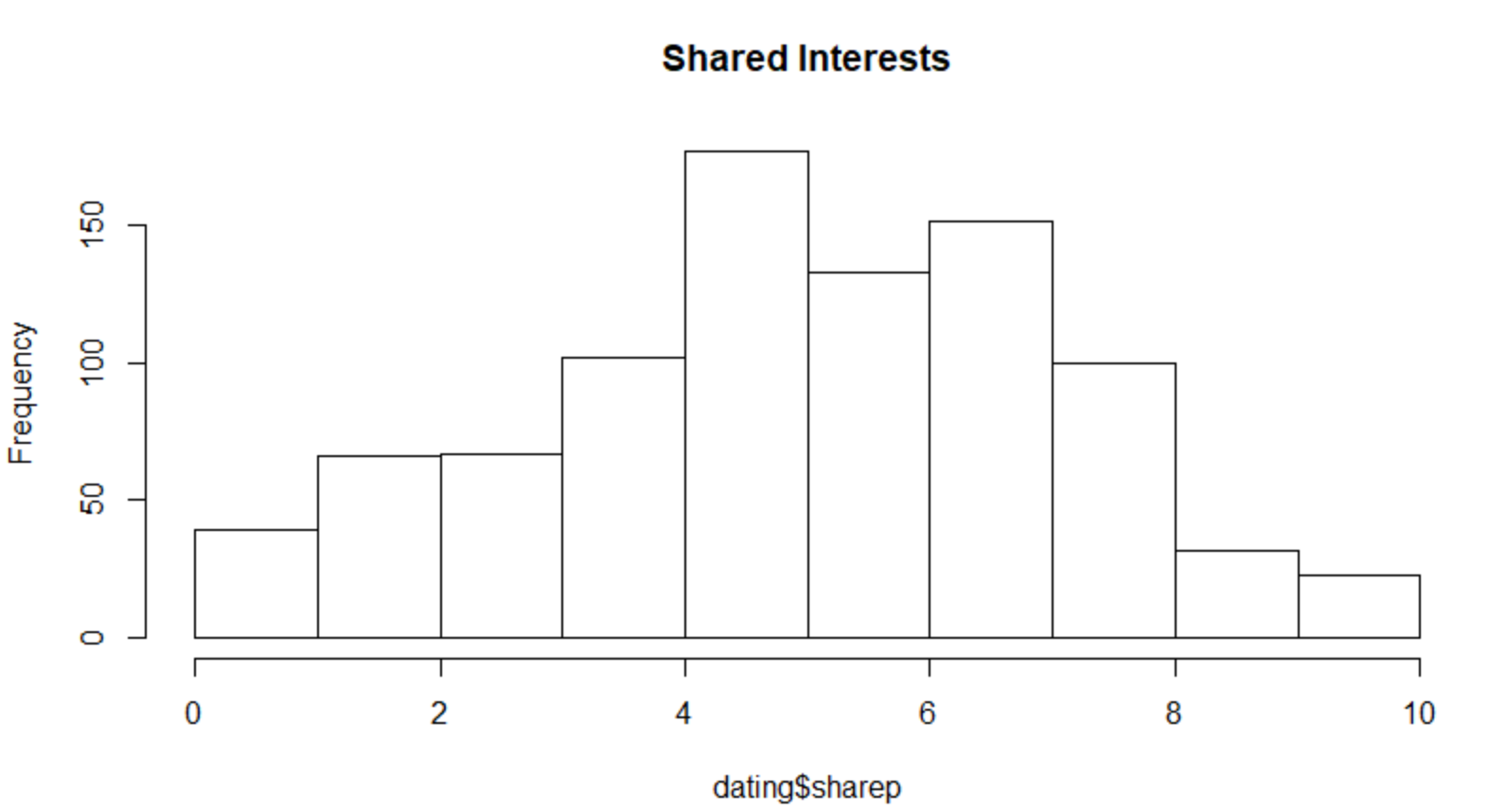
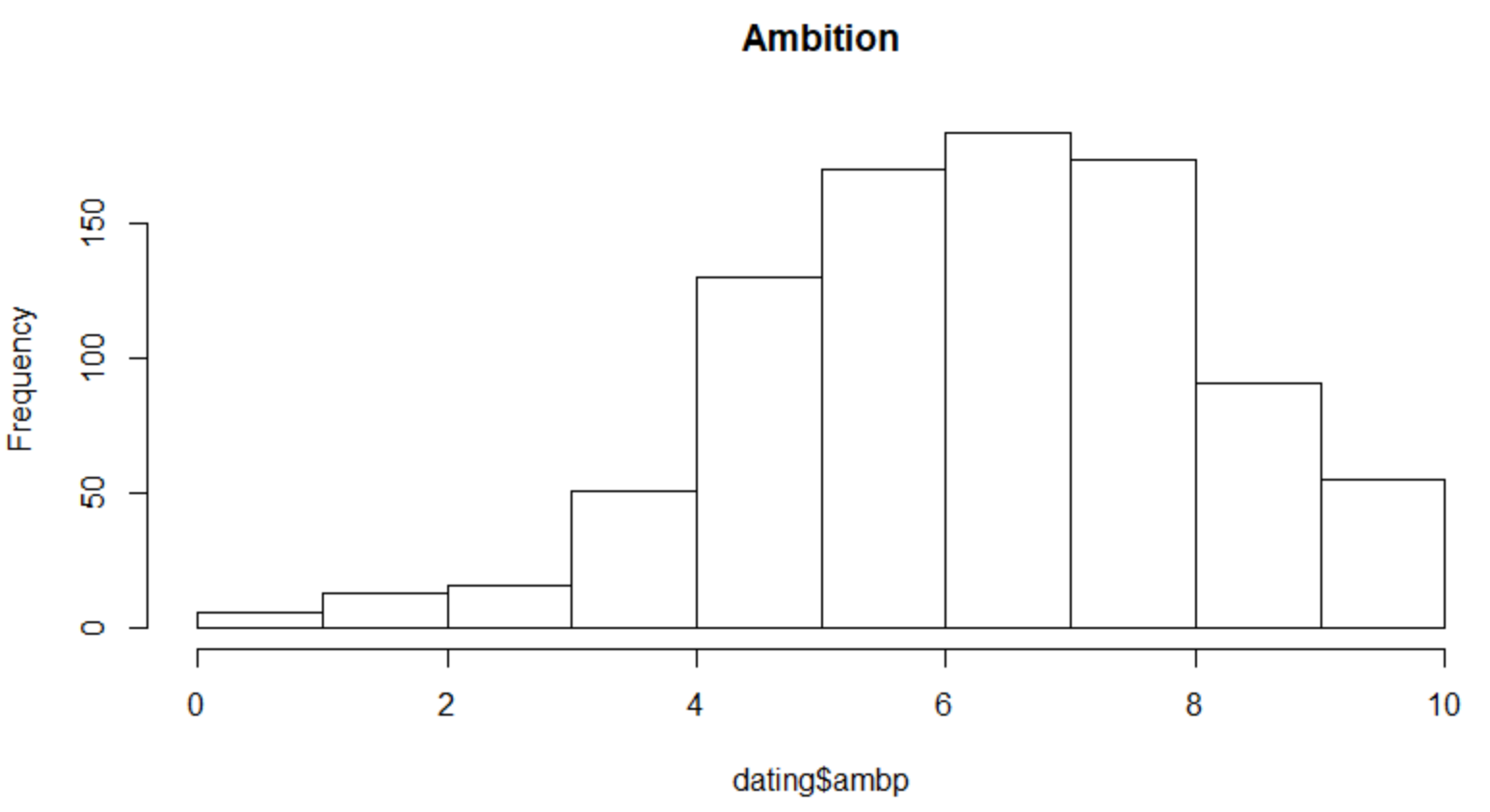
## 8.1.3 graphical univariate analysis of explanatory variables

Most of the data is clean, there are no missing values in the selected sample and there are very few outliers.

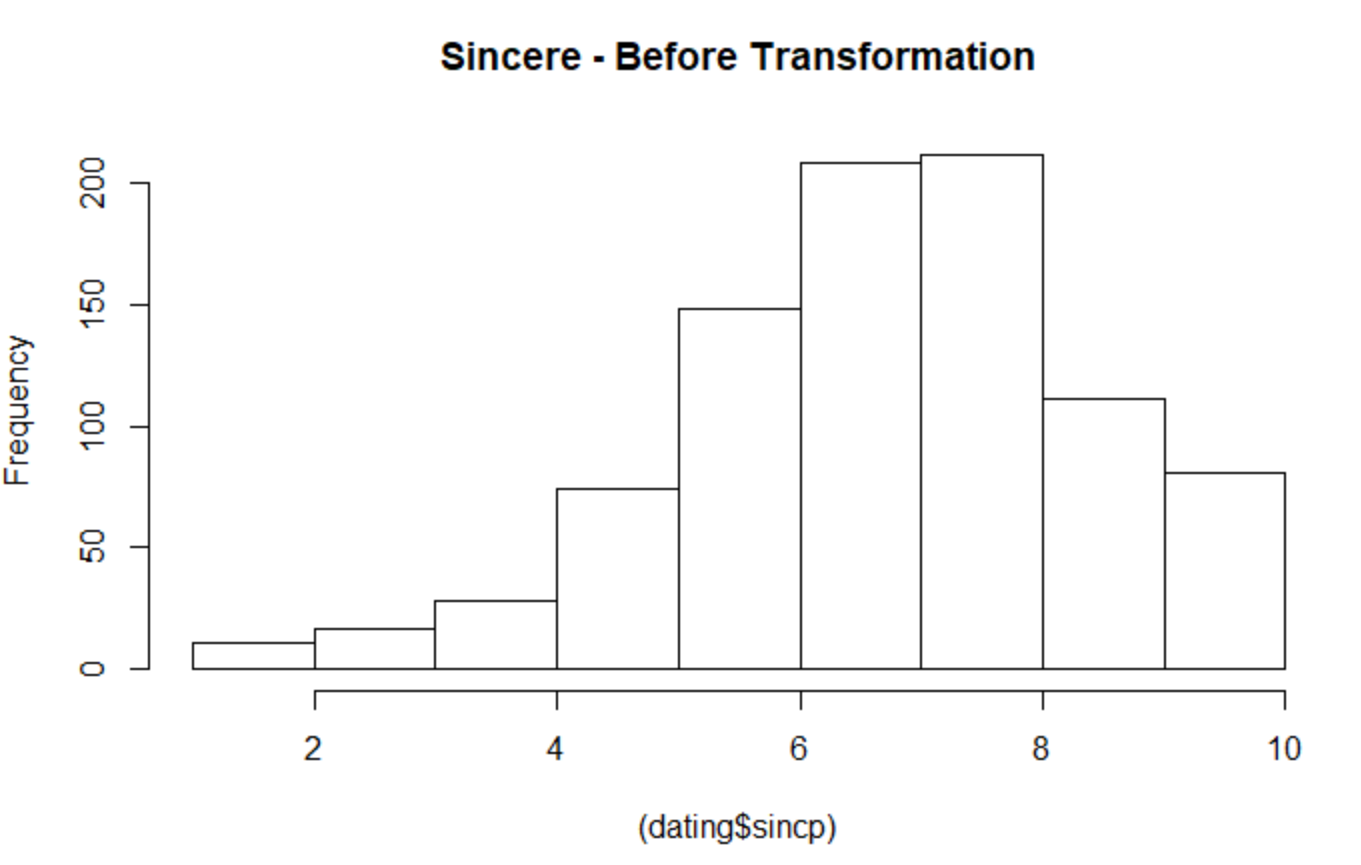
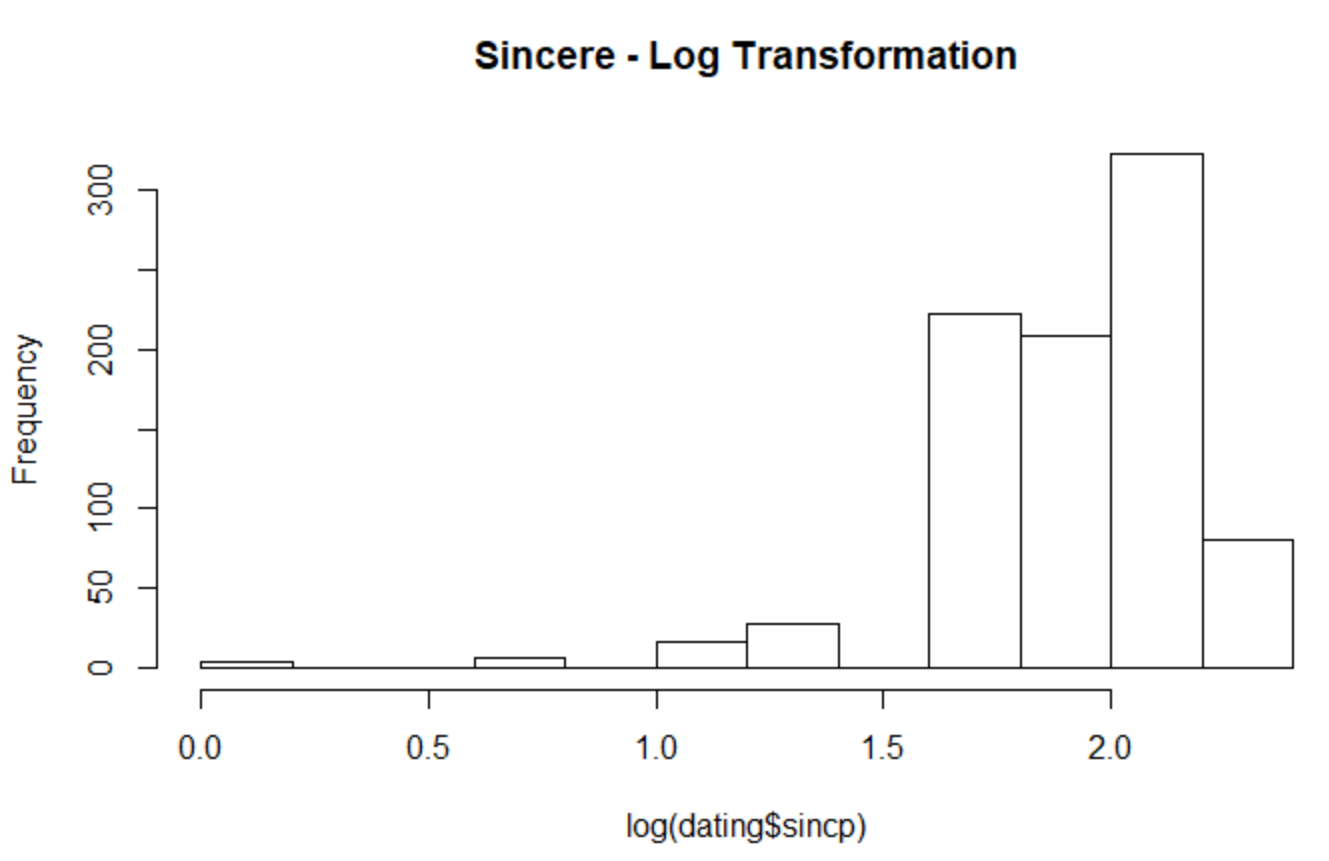
The distribution of the key explanatory variables is almost normal. Variables like “Sincere”, “Ambition” and “Intelligence” are a slightly skewed to the left and log transformation was performed to check if the skewness gets corrected.



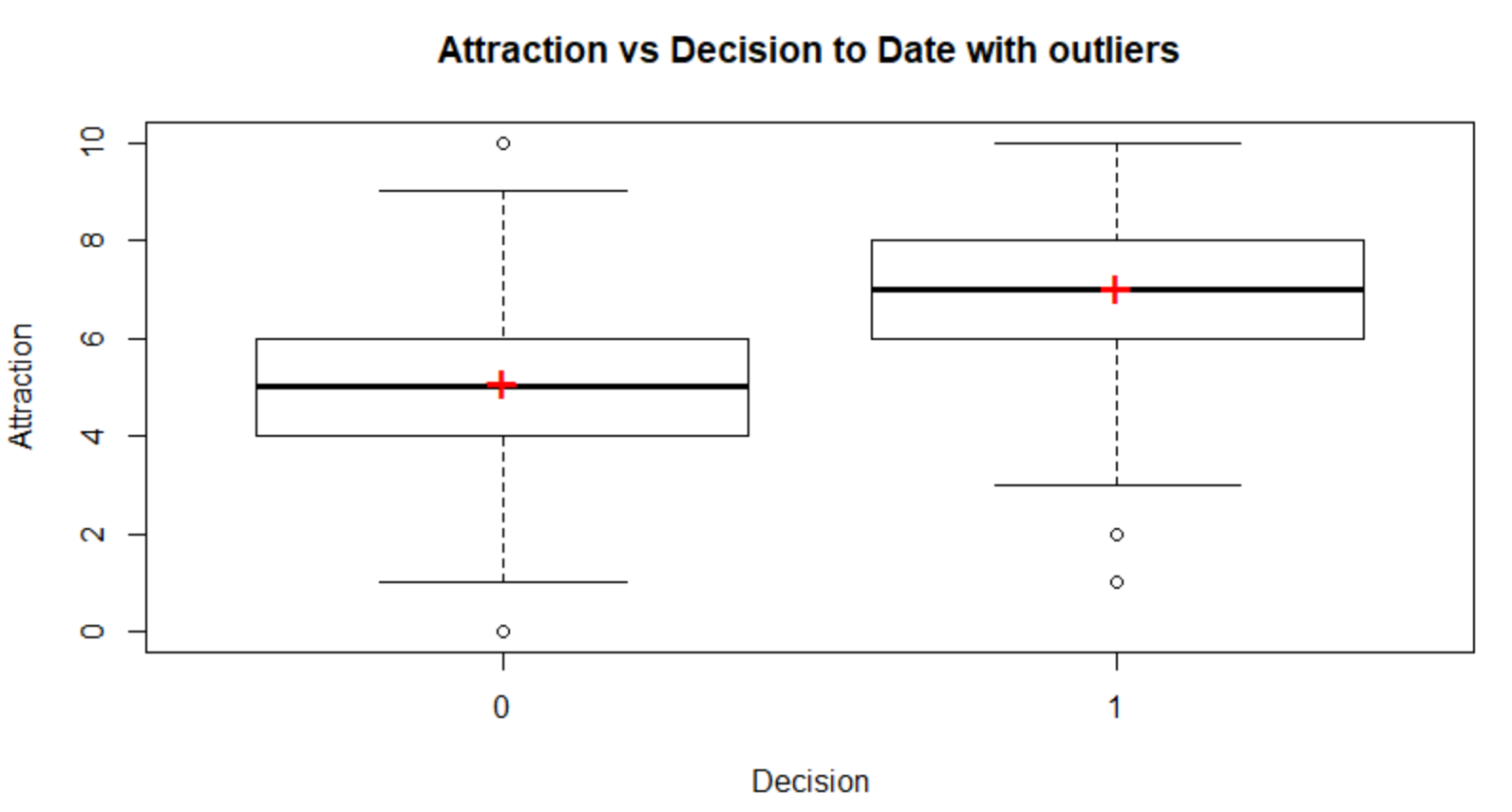




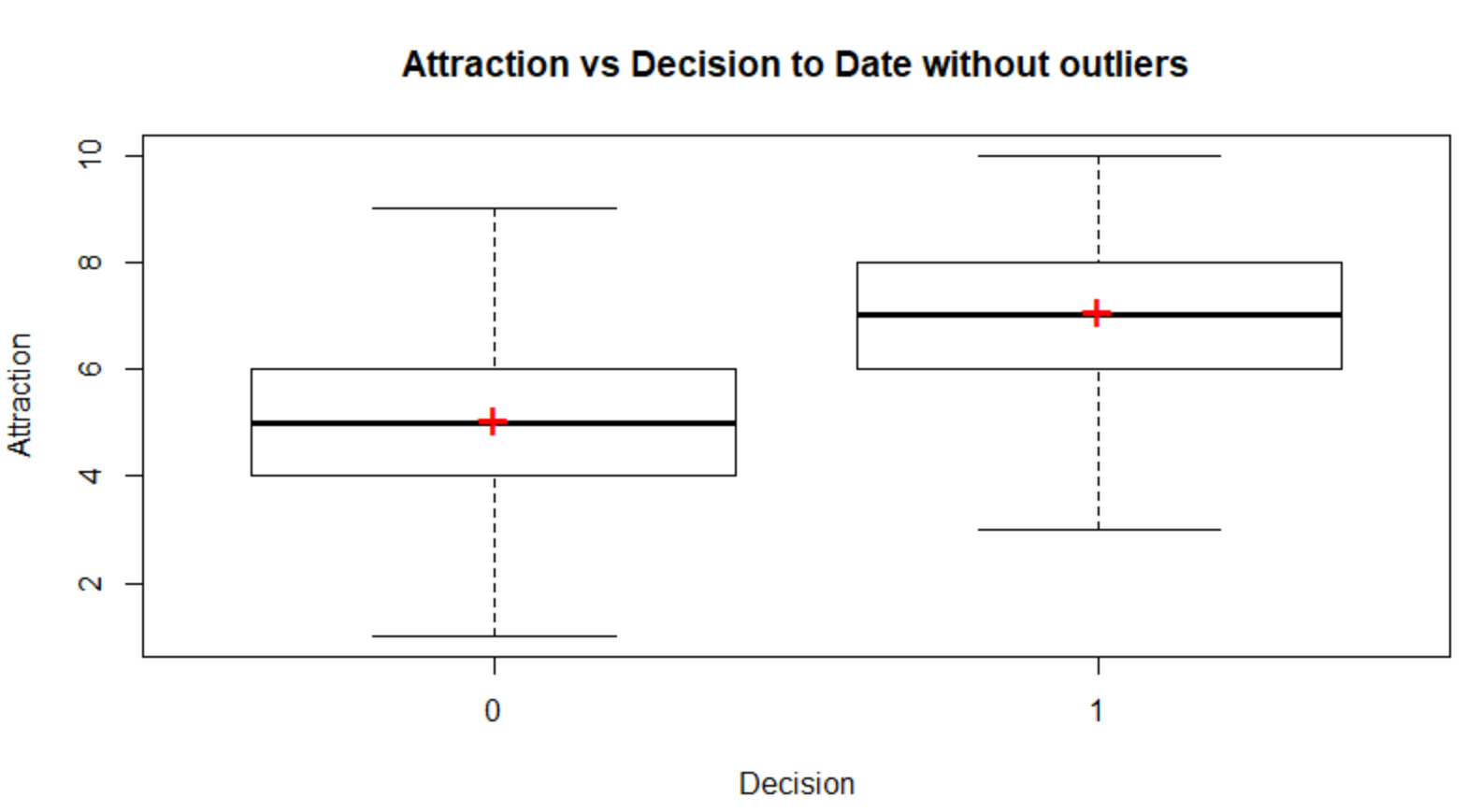
The distribution of the “Sincere” variable without any transformation and with log transformation is presented below. It is evident that the log transformation did not correct the skewness and the same is the case with “Ambition” and “Intelligence” variables. Therefore, the variables are used without any transformation for the further analysis.

## 8.1.4 outlier treatment

The below graphical display shows that there are 4 outliers, out of the total sample size of 890, in the “Attraction” variable. Analyzing mean & median after removing outliers from the dataset. 

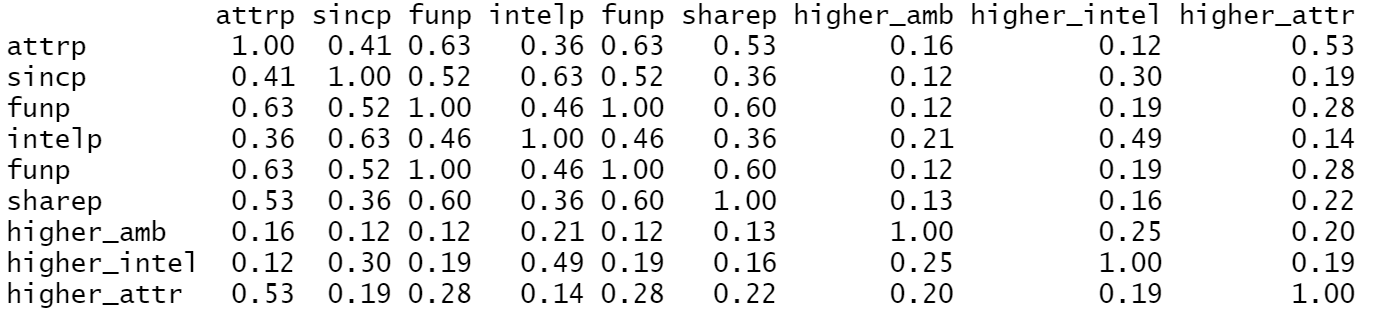
From the below graph, it is to be noted that the mean and median values are not affected largely after removing the outliers. The red plus indicates the mean of the variable and the line dividing the box indicates the median. It is the case because, the values of the “Attraction” variable are bound between 0 and 10 and the proportion of outliers is very less when compared to the sample size. Proceeding for further analysis after outlier removal.



## 8.1.5 Correlation analysis

The correlation between the explanatory variables has been checked, to ensure that there is no multi collinearity issue in the model. The model with multicollinear explanatory variables will give biased estimates and inflated standard errors. Therefore, this check is necessary before model building.

Table 4



It is evident from *Table4* that collinearity between the explanatory variables is very less. Therefore, all the variables can be included in the full model.

## 8.2 Data Partitioning

Before starting the model building, in order to test for our research hypothesis, the cleaned data had to be partitioned into two subsets, on the “male” variable i.e., we divide the data for men and women. The subsets are almost balanced. We have 436 observations for the female subset and 454 observations for the male subset.

In each of the subsets, the data is further divided into training and testing sets. Training set is used for model building and testing set is used for model validation.

The number of observations in each of the subset is mentioned in *Table5*.

Table 5

|  |  |  |  |
| --- | --- | --- | --- |
|  | Women | Men | Total |
| Train | 436 | 454 | 890 |
| Test | 131 | 135 | 266 |
| Total | 567 | 589 |  |

Since the outliers, missing values and skewness has been corrected in the data already, these subsets are ready for model building.

## 8.3 Building Individual Models

## 8.3.1 Model building for female dataset

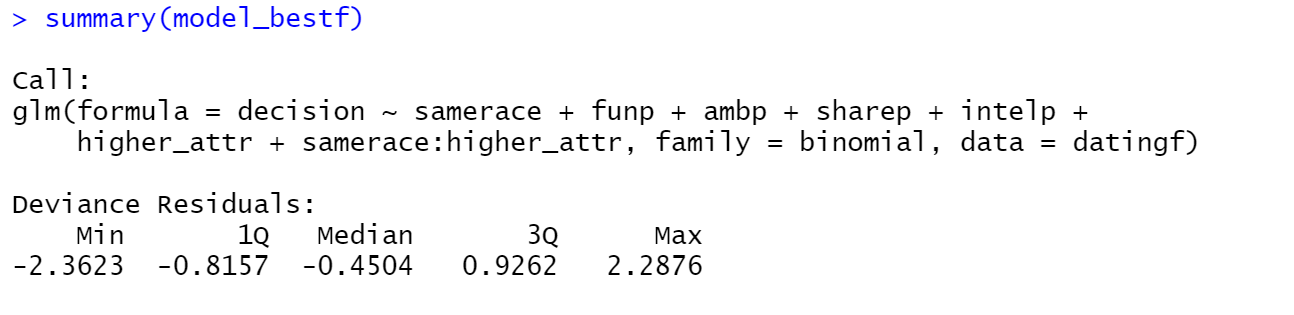
The full model has been built with all the explanatory variables, along with their two-way interaction effects.

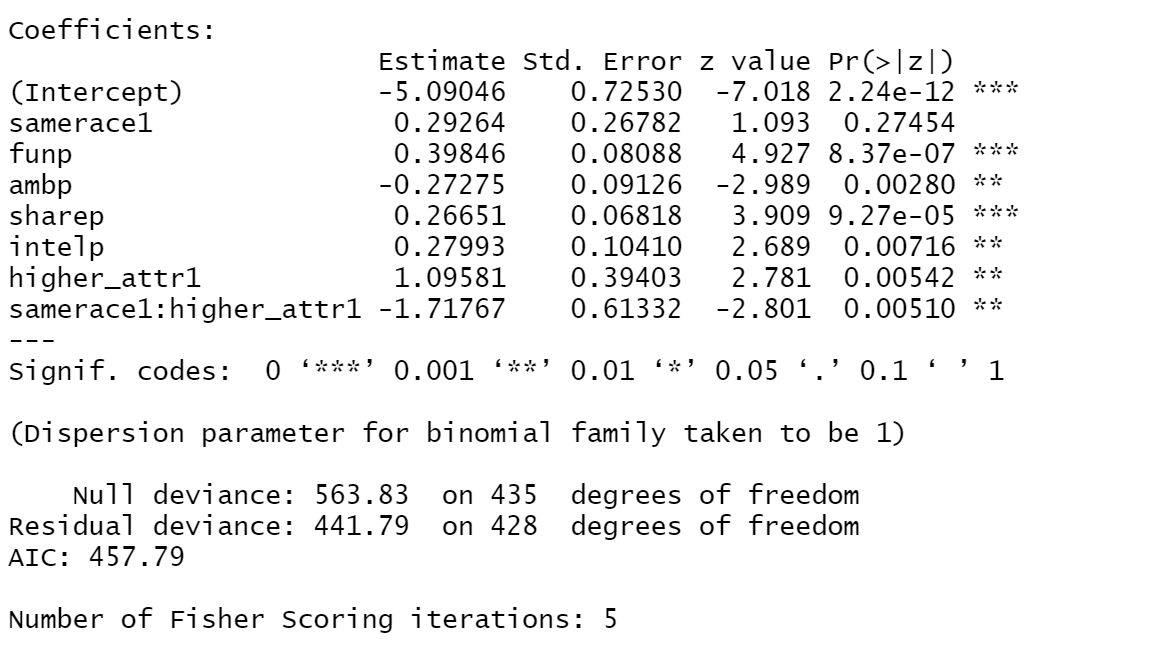
The BIC of the full model is 736.3133.

Sequential variable selection technique, of backward elimination is used to select the best-model, based on BIC penalty. The best fit model resulting from this method had 7 explanatory variables, out of which 6 are significant even at 1% significance level.

The BIC of the reduced model is 490.4132.

The R- summary of the reduced model is listed below.





The significant variables according to this model are listed in *Table6*,

Table 6

|  |  |
| --- | --- |
| Variable | Alpha Level |
| samerace | Insignificant |
| funp | 0.01% |
| sharep | 0.01% |
| ambp | 1% |
| intelp | 1% |
| higher\_attr | 1% |
| samerace:higher\_attr | 1% |

The beta coefficients of the significant variables are interpreted below in *Table7*.

Table 7

|  |  |
| --- | --- |
| 1 Point increase in the variable | Odds of Agreeing (Vs rejecting) for a second date |
| funp | Increases by 0.4895 |
| sharep | Increases by 0.3054 |
| ambp | Decreases by 0.2386 |
| intelp | Increases by 0.3229 |
| higher\_attr (for 1 value) | Increases by 1.9916 |
| samerace:higher\_attr (For 1 value) | Decreases by 0.8205 |

## 8.3.2 Model building for male dataset

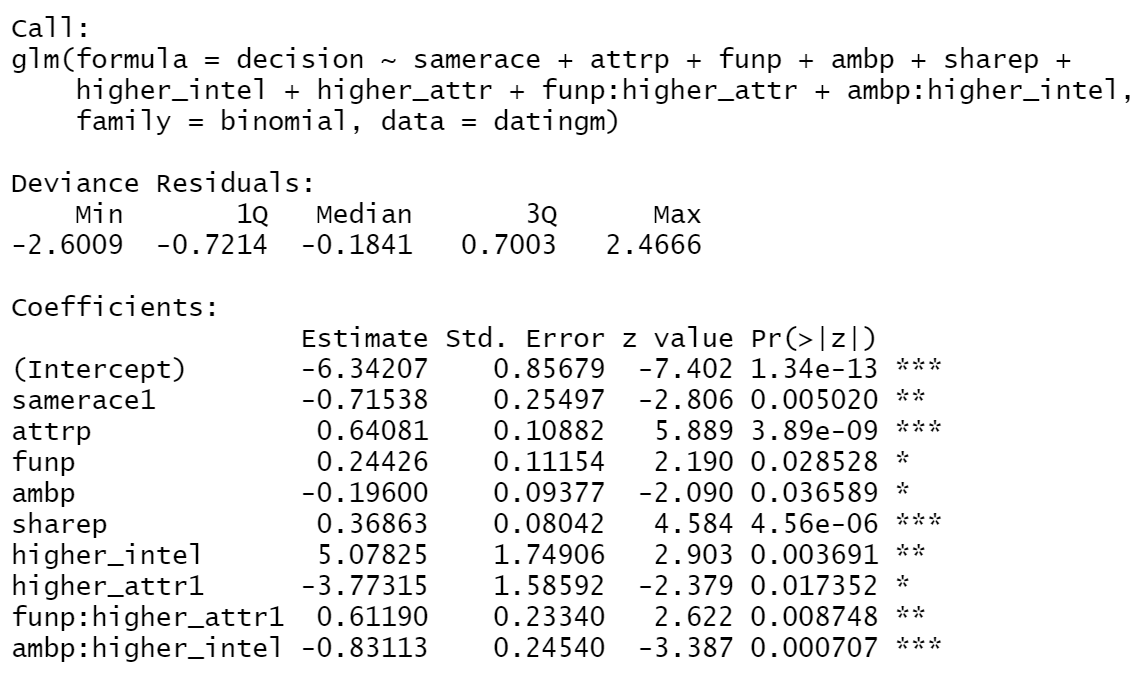
The full model has been built with all the explanatory variables, along with their two-way interaction terms.

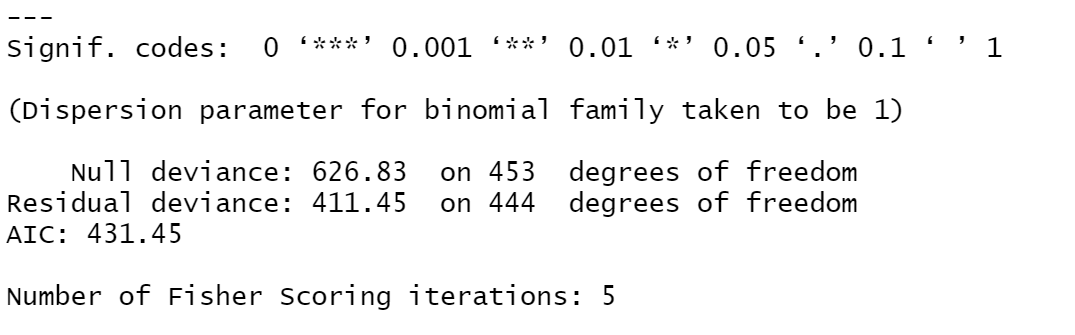
The BIC of the full model is 715.4253.

Backward elimination technique for sequential variable selection has been used with Backward elimination technique. The measuring parameter used is the BIC penalty.

The BIC of the best model suggested by the above technique is 472.6276

The R-summary of the best model is mentioned below.





The significant variables as per the best-fit model are listed in the *Table8*

Table 8

|  |  |
| --- | --- |
| Variable | Confidence level |
| samerace | 1% |
| attrp | 0.10% |
| funp | 5% |
| ambp | 5% |
| sharep | 0.01% |
| higher\_intel | 0.01% |
| higher\_attr | 5% |
| funp:higher\_attr | 1% |
| ambp:higher\_intel | 0.01% |

Coefficients of the significant variables according to the reduced model, along with their interpretation, are listed below in *Table9*

Table 9

|  |  |
| --- | --- |
| 1 Point increase in the variable | Odds of agreeing (Vs rejecting) for a second date |
| samerace | Decreases by 0.5109 |
| attrp | Increases by 0.8980 |
| funp | Increases by 0.2766 |
| ambp | Decreases by 0.1779 |
| sharep | Increases by 0.4457 |
| higher\_intel | Increases by 159.484 |
| higher\_attr | Decreases by 0.9770 |
| funp:higher\_attr | Increases by 0.8439 |
| ambp:higher\_intel | Decreases by 0.5644 |

# 9. Findings

For women, key interpretations as per the best fit model are,

* Women prefer dating men who are fun to be with and have shared interests. These two variables are highly significant at 0.01% alpha level.
* Women do not prefer dating ambitious men, though they prefer dating intelligent men. Both these variables are significant at 1% alpha level.
* If the man and woman belong to the same race and the man is more attractive, the odds of the woman agreeing for a second date decrease.

For men, our key findings from the best fit model are,

* Men prefer dating women who are attractive and share common interests, both the variables being highly significant at 0.01% alpha level.
* Chances of men dating women who are not of same race are higher, than when compared to the women belonging to the same race.
* There is an interesting finding of “Attraction” variable - Men prefer woman who are more attractive compared to themselves and also fun to be with. They don’t prefer women who are only more attractive than them and not fun to be with.
* Similar is the case with “Intelligence” variable - Men prefer dating women who are more intelligent than themselves. However, if the woman is ambitious, along with being more intelligent, then the chances of them dating are significantly reduced.

The social structure theory advocates that the men give more importance to the physical attributes and the women weigh men on the aspect of being resourceful i.e., factors like intelligence take precedence.

Though, the findings of the research are aligned with the above-mentioned theory, some interesting deflections can be observed. Men also prefer women who are more intelligent than themselves and women prefer men who are attractive and share common interests. This can be an indication of the gender role overlap which is being witnessed in the recent times.

# 10. Limitations

* All the participants have volunteered for the experiment, we cannot draw inferences to the population. The sample does not represent the entire population and hence the scope of the findings is limited to the group of participants.
* It is a huge assumption that there would be no other factors influencing the decision to date, other than the ones considered in the research. However, in practicality, it is not the case. There can be many confounding variables like financial status, education qualification etc which may play a role in an individual’s decision regarding dating. Therefore, causality cannot be established. Research can be extended in this domain, by controlling for these confounding variables either in the design stage or the analysis stage.

# 11. Bibliography

Belot, Michèle V. K. and Francesconi, Marco, Can Anyone Be 'the' One? Evidence on Mate Selection from Speed Dating (October 2006). IZA Discussion Paper No. 2377. Available at SSRN: <https://ssrn.com/abstract=941111>

Weisfeld, G. E., Russell, R. J., Weisfeld, C. C., & Wells, P. A. (1992). Correlates of satisfaction in British marriages. Ethology & Sociobiology, 13(2), 125–145. [https://doi.org/10.1016/0162-3095(92)90022-V](https://psycnet.apa.org/doi/10.1016/0162-3095(92)90022-V)

Townsend, J. M. (1989). Mate selection criteria: A pilot study. Ethology & Sociobiology, 10(4), 241–253. [https://doi.org/10.1016/0162-3095(89)90002-2](https://psycnet.apa.org/doi/10.1016/0162-3095(89)90002-2)

# 12. Appendix

## 12. 1 Explanatory variables – definitions

List of explanatory variables considered for the research are listed in *Table10*.

Table 10

|  |  |  |  |
| --- | --- | --- | --- |
| S.No | Variable | Values | Description |
| 1 | Gender | 0 or 1 | 0 for Female and 1 for Male |
| 2 | Samerace | 0 or 1 | 0 if both the participants don’t belong to the same race, else 1 |
| 3 | age\_p | Numeric | Age of the participant |
| 4 | age\_s | Numeric | Age of the subject |
| 5 | decision | 0 or 1 | 0 if the subject doesn’t want to go on a second date, otherwise 1 |
| 6 | attrp | 1 to 10 | Subject's rating for the participant on "attraction", 10 being the highest |
| 7 | sincp | 1 to 10 | Subject's rating for the participant on "sincere", 10 being the highest |
| 8 | funp | 1 to 10 | Subject's rating for the participant on "fun", 10 being the highest |
| 9 | intelp | 1 to 10 | Subject's rating for the participant on "Intelligence", 10 being the highest |
| 10 | ambp | 1 to 10 | Subject's rating for the participant on "Ambition", 10 being the highest |
| 11 | sharep | 1 to 10 | Subject's rating for the participant on "shared interests", 10 being the highest |

## 12. 2 Model evaluation – Reduced Model for Female Dataset

The best fit model considered for the female dataset can be evaluated using the statistical measures listed in this section.

*Table11* presents the confusion matrix of the model, which specifies accuracy, misclassification rate, sensitivity and (1-specificity) measures of the model.

Table 11

|  |  |  |  |
| --- | --- | --- | --- |
|  | Predicted | | |
| Actual |  | 1 | 0 |
| 1 | 79 | 73 |
| 0 | 39 | 245 |

The model’s performance can be evaluated with the below mentioned characteristics.

The model evaluation parameters for train and test datasets can be referred in *Table12*

Table 12

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Train | Test | Description |
| Accuracy | 0.75 | 0.8473 | Proportion of observations correctly classified |
| Misclassification Rate | 0.25 | 0.1527 | Proportion of observations incorrectly classified |
| Senstivity | 0.5238 | 0.5555 | Proportion of true positives classified by the model |
| (1-Specificity) | 0.1338 | 0.0421 | Proportion of false positives classified by the model |
| Area Under the Curve | 0.8 | 0.853 | Range of AUC is 0 to 1, 1 being the Area under the curve for a perfect model |

It is evident that the reduced model fits the test data better than the train data. This is an indication of the model underfitting the train data.

## 12. 3 Model evaluation – Reduced Model for Male Dataset

The best-fit model constructed for the male dataset, can be evaluated using the statistical measures listed in this section.

The *Table13* presents the confusion matrix for the model, from which the other evaluation metrics like Accuracy, Misclassification rate etc are calculated.

Table 13

|  |  |  |  |
| --- | --- | --- | --- |
|  | Predicted | | |
| Actual |  | 1 | 0 |
| 1 | 163 | 47 |
| 0 | 49 | 195 |

The performance of the reduced model with respect to training and test data are listed below.

Table 14

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Training | Test | Description |
| Accuracy | 0.7885 | 0.7852 | Proportion of observations correctly classified |
| Misclassification Rate | 0.2115 | 0.2148 | Proportion of observations incorrectly classified |
| Senstivity | 0.77619 | 0.4411 | Proportion of true positives classified by the model |
| (1-Specificity) | 0.2008 | 0.099 | Proportion of false positives classified by the model |
| AUC | 0.865 | 0.8456 | Range of AUC is 0 to 1, 1 being the Area under the curve for a perfect model |

It can be observed that the reduced model fits the training and test datasets equally well. Therefore, there is no indication of underfitting or overfitting.