Speed Dating, Contrast or Teamwork

[redacted]

Abstract

Can we statistically understand love? This article answers this question by exploring data collected from a series of Speed Dating experiments. We fit a logistic regression model to infer what pair-wise features facilitate a successful match. Pair-wise features were constructed by taking the absolute difference or the sum, which leads to two model frameworks: Contrast vs. Teamwork. We find that in general the contrast between a pair of participants decreases the probability of making a match, with a few exceptions. Teamwork framework helps us to gain clarity on how the two participants are similar to each other. Similarity turns out not always to be helpful for match making. In the Teamwork model framework, we find that a pair of participants with high ambitions or a sense of high importance of religion will less likely to be a match.

KEY WORDS: Logistic Regression, Speed Dating.

1. Introduction

1.1 The Speed Dating Experiment

The Speed Dating Experiment consists of 21 waves of controlled speed dating experiments from 2002-2004. The data was collected for the study done by Ray Fisman and Sheena Iyengar for their paper *Gender Differences in Mate Selection: Evidence From a Speed Dating Experiment*[1].

-*Participants:* the participants are undergraduate and graduate volunteers from Columbia University.

-*Wave:* Each wave is considered as an independent experiment. By independence of wave we mean that there is no overlap in participants in each wave. No participants will show up in two waves of the experiments. The general setting of each wave of experiment are controlled to be identical, with a few exceptions where slight variations in experiment design are introduced. Each wave include 20-40 subjects from both genders (male and female).

-*Procedure:* During each wave of experiment, all possible heterogender pairs of participants go on a 4 minute dates and provide feedback and evaluation of their partners at the end of date. Additionally, demographic data and dating habits and preference of all subjects are collected in the form a survey.

1.2 Evaluations, Match and Survey

Participants were asked to rate their partner from 1-10 on the following six aspects: attractiveness, sincerity, intelligence, fun, ambition, shared interests. Additionally participants rate their partner's overall as a potential date and the probability of their partners agree to date them from 1-10. Both participants were asked to rate whether they met with their partner before (only 45 pairs of participants have met before). Both participants in a pair were asked to make a decision of whether to go out with each other. A successful match is defined as the event when both agree to go out.

Before the Speed Dating event, participants filled out a survey to provide personal information. Demographic information includes age, race, zip code, career and field. Dating preference was surveyed in two ways: 1) allocation of 100 points to the six aforementioned aspects according to importance in an ideal partner 2) score 1-10 rating of the aforementioned aspects of an ideal partner. Additionally, importance of race and importance of religion was surveyed at a scale of 1-10. Dating habit was surveyed by question on dating frequency, go out frequency, both at 1-7 scale (1 frequent, 7 never) and dating goal going into the experiment (serious level varies). The participants were also required to rate their interests in a list of activities.

A very interesting, yet unused part of the survey questions involves participants perception of self, same sex population and opposite sex population. The participants were asked to rate themselves from 1-10 in 5 aspects: attractiveness, sincerity, intelligence, fun, ambition. Then they were asked to rate what they think most of their fellow sex look for in the opposite sex in these five aspects. Finally they were asked to rate what the most opposite sex look for in their sex in these five aspects. Although these questions provide us with interesting insights of how people perceive the opposite sex and their own sex, it is too involved to be related to our project and therefore utilized as part of the feature construction in our final model.

1.3 Prior analysis

The major goal of in Fisman and Iyengar's study [1] was to detect if men and women search for different characteristics in a potential partner. The researchers performed simple linear regression models and univariate logistic regression models, often using the subjects decision (of whether or not they would be interested in going on a real date with their speed date partner) as the response variable. Taking their analyses into account, we instead focus on the successful match of a pair of participants and how their contrast or similarity influence the outcome. We also developed more sophisticated models that account for a larger number of variables.

2. Data Manipulation

The raw data comes in a csv file we downloaded from Kaggle. The raw data consists of 8378 rows indexed by iid, the identifiers for both participants in a pair for a certain speed date. Each row includes the evaluation from the partner (pid) on the participant (iid) and participant's (iid) survey answers. Individual features (feature related iid) are repeated in rows for the same iid, while evaluation from the partner changes as pid changes in each row.

-*Sanity check:* We wrote a function to detect the missing values in the dataset, and found that most missing values were caused by a few participants failed to fill in survey questions such as age, career field and zip codes. We removed all missing values rows for the features we selected to be included in the model.

-*Wave selection:* We subset the data from two aspects: wave selection and feature selection. Since there are known variations to the experiment design in certain waves, we chose the waves without such variations to ensure homogeneity among selected waves. One major variation that would potentially impact out analysis is the way dating preferences were surveyed. We picked all the waves where participants were asked to allocate 100 points to the 5 aspects to ensure unity of metrics. Without better knowledge of how these variations effect participants' decision-making, we excluded wave 5 (in which the participants were all undergraduates), wave 12 (in which the participants were only allowed to say yes to no more than 50% of those they encountered), waves 13 and 14 (in which there was a different M.C. than usual), and waves 18 through 21 (in which the participants were asked to bring either a book or a magazine).The final waves we selected are wave 1, 2, 3, 4, 10, 11, 15, 16, 17, dated from 10/16/2002 to 2/25/2004.

-*Feature selection:* we put the features into two major categories: evaluation from the partner and survey data. Among survey data, we have a few subcategories: demographics, dating preference, dating habit, personal hobby and perception related feedbacks. Starting from the basic, we selected evaluation from the partner and dating preference. Under demographics, we selected age, race, career, field. Although an estimate of income based on home address zip code was also provided, we chose to exclude this feature due to many missing values. Under dating habits, we include dating frequency, go out frequency and dating goal. We excluded personal hobby and perception related feedback since they are less relevant to our major goal.

2.1 Pair-wise Data Structure

We transformed the raw data into a true pair-wise structure, with each row indicating a certain pair of participants in the experiments. First we selected the subset of data with the selected waves and features. For each selected wave, we then split the wave data by gender, select individual features that is related to iid and merged these two gender subset by iid=pid and pid = iid. Individual-features includes identifiers such as iid, gender, wave, pid and individual survey answers.

More specifically, for each male iid_m, to find their partner pid_f's individual feature, we have to look for data where $iid = pid_f$. Thus, we pair up male and female participants by appending individual features of pid_f to the row of $iid = iid_m$ and $pid = pid_f$ and get rid of the data where $iid = pid_f$.

After transformation we got 1880 rows of pair-wise records of individual features of unique pair (iid, pid). Among these rows, the ones without NAs sum up to be 1188, which gives us a decent sized sample.

2.2 Pair-wise Feature Construction

We define pair-wise-features as the outcomes of combining the same feature from the two participates in a pair. Numerical features were combined by sum of absolute difference. And factor features were combined using logic function, resulting in whether the two participants have the same level for a certain factor feature. Pair-wise-features were the appended to the pair-wise individual feature data as additional columns.

3. EDA

We first examined the correlation between continuous features. In figure 1 we plotted the correlation between pairwise absolute difference for each feature. We can see that the difference in preferences are all highly positively correlated. Using Preference of Intelligence and Preference of attractiveness as an example, a positive correlation means that if the difference between the two participants' preference for intelligence is above average, it is very likely for the difference between their preference for attractiveness to be above average as well.

We then plotted the correlation between pairwise sum for each feature and found that the correlation matrix is the same as that of pairwise absolute difference. Thus, Figure 1 also illustrates the correlation between pairwise sum features. In terms of sum, we can interpret positive correlation between preference for intelligence and preference for attractiveness as if both participants have a high preference for intelligence (therefore their sum is above average), it is likely for them to have a high preference for attractiveness as well.

Furthermore, the rating of partner's traits are also positively correlated. Figure 1 reveals what are the most important aspects that factor into the overall score, which is a general rating of the partner as a potential date. The overall score is positively correlated with rating of attractiveness, rating of fun and rating of shared interests. Additionally, rating of attractiveness was moderately positively correlated with rating of ambition, rating of sincerity and rating of shared interests

Next we examined our key assumption, that the waves we selected are homogeneous. We first used logistic regression to regress match on wave as a factor. The result is shown in Table 1. Here we see that all the estimates are negative. With wave 1 as the intercept, negative estimates of the rest of the waves means that pairs in waves other than wave 1 had a lower probability of making matches. The effects of waves except for wave 16 compared to wave 1 are significant. Compared to being in wave 1, being in wave 2, 3, 4, 10, 11, 15, 17 makes it 3-4 times less likely to make a match.

To confirm our findings we performed an ANOVA use match rate, which is the number of matches divided by the total number of pairs. Here we want to test the significance of pair-wise difference between waves. Tukeys test gives us the 95% pairwise confidence interval shown in figure 2. We see that wave 1 differs from most of our other waves, with the exceptions of waves 10 and 16.

With these two sets of results, we conclude that the difference between wave 1 and the rest of the waves is significant with a few exceptions. This motivated us to add wave as a fixed effect to our model.

Table 1: Logistic regression of Match on Wave

	Dependent variable:
	Match
wave2	-1.538^{***} (0.354)
wave3	-1.480^{***} (0.463)
wave4	-0.847^{***} (0.315)
wave10	-0.887^{**} (0.444)
wave11	-1.306^{***} (0.309)
wave15	-1.084^{***} (0.341)
wave16	-0.582 (0.455)
wave17	-1.161^{***} (0.371)
Constant	-0.448^{*} (0.267)
Observations	1,188
Log Likelihood	-544.061
Akaike Inf. Crit.	1,106.123
Note:	*p<0.1; **p<0.05; ***p<0.01

4. Method

We propose to regress match on pair-wise-features using logistic regression. Match is coded 1 if it is a success (both participants agree to go on a date) or failure (if at least one of the participants refused). Therefore we define p as the vector of probability of making a successful match. X is the design matrix that includes all pair-wise features

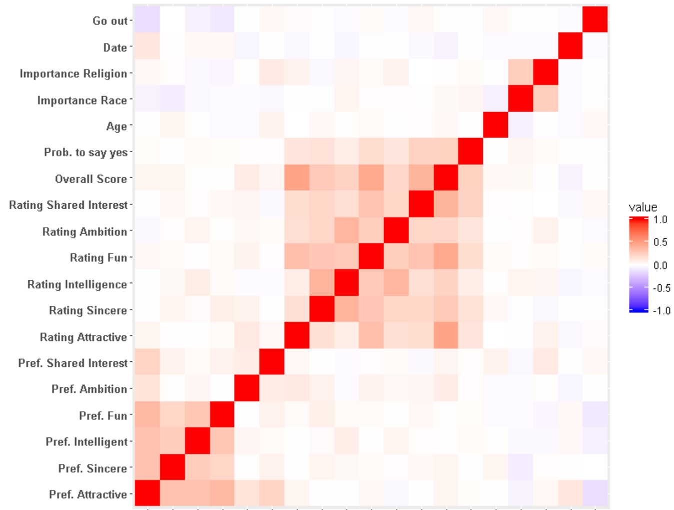

Figure 1: Correlation plot of continuous pair-wise absolute difference and sum. (The correlation matrix are the same for these two measures)

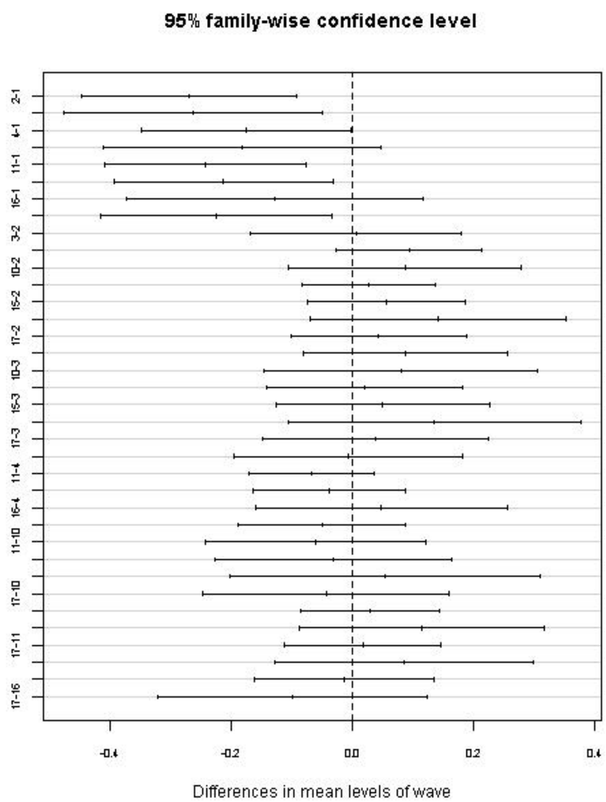

Figure 2: Tukey's 95% confidence interval

and the intercept 1. The continuous pair-wise features are constructed by either taking the absolute difference for Contrast model or sum for Teamwork model. The factor features are same in both model. p is of dimension $N \times 1$, β is of dimension $(m+1) \times 1$ and X is of dimension $N \times (m+1)$.

$$logit(p) = X\beta \tag{1}$$

4.1 Wave Fixed Effect

With the EDA, we can see that despite of our effort to select homogeneous waves, there are still unobserved variations among waves. To capture such variations, we propose to put fixed effect on wave. We refit the model as the following. Here $\overline{wave}$ is a vector of wave number as factor of dimension $N \times 1$. $\beta_w \in R$

$$logit(p) = X\beta + \beta_w \overline{Wave} \tag{2}$$

4.2 Individual Fixed Effect

Since the experiment is done with all possible hetero-gender pairs, the sample of pairs within a wave are not independent. Since a male participants dates all female participants in a wave, the individual effect of this participants may influence all of the dates he goes on. E.g. a participant could be very tall, and therefore generating dependent responses from all his partners. Such effect is not captured by any of the features collected by the experiment. Thus, we propose the following model

$$logit(p) = X\beta + \beta_{iid}\overline{IID} + \beta_{pid}\overline{PID} \tag{3}$$

Here the $\overline{IID}$ is a vector of factor indicating the effect of a specific male participant and $\overline{PID}$ is a vector of factor indicating the effect of a specific female participant. $\overline{IID}$ and $\overline{PID}$ are both of dimension $N \times 1$.

However, the problem with this model formulation is that with so many parameters, we lose a lot of degrees of freedom which leads to a failure of convergence. To fix this issue, we propose an alternative formulation:

$$logit(p) = X\beta + \beta_{iid}\overline{IID} \tag{4}$$

$$logit(p) = X\beta + \beta_{pid}\overline{PID} \tag{5}$$

We want to fit model with iid and pid separately to see if a certain male or female participants have a significant influence on the match. The estimates for factor iid or pid represents the average effect of a particular male participant or female participant on match compared to the one individual chosen to be the reference. We hope to see if individual fixed effects improves our model by lowering the AIC.

4.3 Model Selection

With inference as our main goal, we want to select the subset of significant features. We used two approaches: 1) significant subset 2) both-way stepwise regression.

We automated the feature selection process that is usually done manually: first fit the model with all reasonable features to get M_0, and select the variables with a p-value smaller than 0.05 under the Wald test. Then we refit the model with only the subset of significant features from step 1 to get M_1. Finally we select from M_0 and M_1 by comparing the AICs of the two models. The final model is the one with a smaller AIC.

The second approach we use is both-way stepwise regression. Here we propose a null model M_{null} to regress match on only intercept and a full model M_{full} to regress match on all reasonable features. Both-way stepwise regression works like a combination of stepwise forward and stepwise backward regression. At a certain step, the algorithm will either add or drop a feature based on the whether such action decreases the AIC. Such algorithm is more flexible than the forward or the backward regression, since it examines more combination of features.

There is an important difference between how these two approaches deal with factor features. For significant subset, the algorithm will count the factor as significant if at least one level of the factor is significant. The stepwise regression on the other hand, decides whether to add or drop factor feature as a whole depending on whether it will decrease the AIC by doing so. Therefore, with significant subset, we may end up with factors with only certain levels being significant.

The drawbacks of the stepwise regression is that it takes a long time to converge, since it might take many steps to find a optimal model. Therefore, the first approach can gives us some quick results to locate the set of potential significant features to choose from.

Eventually, we choose between the best models proposed by these two approaches by comparing the AICs and select the model with the smaller one.

5. The Contrast Model

In the contrast model, we fit match on pair-wise absolute difference between the participants. The hypothesis is that contrast between the two participants makes it less likely for the two participants to make a match. In this section we will fit the models on three levels: pair-wise features, pair-wise features with wave fixed effect, and pair-wise features with individual fixed effect.

5.1 Pair-wise features

We used both model selection methods discussed before and got two separate models as result. The estimates are shown in Table 2.

Based on AIC of these three models, we select the best model given by both-wave stepwise regression, M2, since its AIC is lowest.

Not surprisingly, most of the estimates for continuous features are negative. This confirms our hypothesis since negative estimates mean that difference between the two participants in a pair leads to a decrease in probability of making a match.

Among continuous features, the estimate for go out frequency is -0.281, meaning a unit increase in go out frequency difference increases the odds of fail to match by 32% holding every other feature constant. The estimates for Overall Score is -0.241, meaning that a unit increase in overall score difference increase the odds of fail to match by 27%. Ordered in terms of descending magnitude of estimates, Rating of fun, rating of attractiveness, Importance of religion, Age, Preference of Intelligence respectively increases the odds of fail to match by 13%, 13%, 9%, 7% and 3% per unit difference increase. Here we assume that other features are held constant when interpreting the estimate for a certain feature.

Notice that Probability of yes is not significant at $\alpha = 0.1$ level under the Wald test. The null hypothesis here is that β is 0. With p-value = 0.052 for Probability of yes, we fail to reject the null hypothesis and conclude that β for this feature is 0. Additionally, we did an ANOVA using Chi-square test to compare the model given by step wise regression and the same model excluding Probability of yes feature and got p-value > 0.1. Here we failed to reject the null hypothesis that $beta = 0$ for Probability of yes and conclude that it is not a significant predictor.

It seems surprising that the stepwise regression selects a non-significant feature. We found that the model fitted with features given by the stepwise regression other than Probability of yes, M3, has a higher AIC. In light of this, it makes sense that stepwise regression would select Probability of yes since it decreases the AIC. Here we see that with different standard for the "best" model, model selection methods lead to conflicting result. We decided to choose the model with smallest AIC to be consistent in our model selection approach.

The estimate for Preference of Attractiveness is positive while the estimate for Rate of Attractiveness is negative. This means that a larger difference in the preference for partner's attractiveness makes it more likely for a pair to make a match, while a larger difference in rating for partner's attractiveness is counter-productive. Since the preference is indicated by 100 points allocated 6 aspects, the similarity in preference for attractiveness potentially means the two participants are more selective and therefore less likely to make a match. By looking at the magnitude of estimates, difference in rate of attractiveness has a larger influence to make a match than that of preference of attractiveness does to break a match.

The estimate for Met no is -1.905, meaning that for people who has not met with each other before almost 6.7 times more likely to fail to match. Unexpectedly, having the same dating goal has a negative effect on making successful match with an estimate -0.448, which means it is 1.56 times more likely for the pair to fail to match.

The fact that having the same goal turns out to be counterproductive is counterintuitive. This effect is driven by the fact that people with same goals have are mostly "low commitment" goals. Among the 380 pairs of participants having the same goal, 224 pairs have the goal "Seemed like a fun night out" and 141 pairs have the goal "To meet new people". Only 2 pairs have the goal "looking for a serious relationship" and 2 other pairs have the goal "looking for a date". Thus the factor same goal effectively profiles the group of the population with lower commitment level going into the experiment.

5.2 Pair-wise features with wave fixed effects

In this model, we include wave fixed effects. The result is presented in Table 3. From the results of two model M4 and M5, we can see that we still prefer result given by both-way step-wise regression, i.e. M5, since it has a smaller AIC.

It should be noticed that the in the model given by significant subset M4, the estimates of all waves other than wave 1 are negative. Since wave 1 is missing from the estimates, the model uses wave 1 as reference and the estimates for the rest of waves are the effect of these waves compared to wave 1. This means that the participants in waves other than wave 1 are less likely to make a successful match compare to those in wave 1. This confirms what we found in the EDA with logistic regression on wave alone and ANOVA. Wave 1 seems to have a higher success rate at making matches. This may be due to the fact that it is the first experiment and people who signed up and participated first were more open to new experience than the rest of the participants. Without better knowledge of the participants, we are unable to justify this finding.

However, the model given by stepwise regression M5 didn't include wave as a significant predictor of match. And the fact that we favor this model over the previous one suggests that with enough features other than wave that are predicative enough, wave heterogeneity stops being a huge factor in predicting match. This is what we hope to achieve by creating a new feature "Met". "Met" is defined as the 1 if both participants in a pair claim that they have met before and 2 if at least one participants in a pair claims that he/she had not met his/her partner. With this additional feature, we are able to get rid of wave as a predictor and achieve a lower AIC compared to the model presented in the presentation where stepwise regression still picks up wave as significant.

The use of fixed effect is more out of necessity as a control. The interpretation of fixed effects doesn't shed meaningful insight other than "these subgroups are different". By loading the model with more explanatory variables, we are able to pin-point the actual difference between subgroups, if there is any, other than simply attribute the explained variance to subgroup difference.

5.3 Pair-wise features with individual fixed effects

We fit this model first by setting both participants in a pair as factors to see how both as a pair influence match.

Table 2: Pair-wise feature model under Contrast Model

	Dependent variable:		
	Match		
	M1: Significant Subset	M2: Stepwise Regression	M3: Stepwise Regression Refined
Preference of Attractiveness	0.018*** (0.005)	0.016*** (0.006)	0.016*** (0.006)
Preference of Intelligence	−0.030** (0.012)	−0.031*** (0.012)	−0.031** (0.012)
Rate of Attractiveness	−0.143** (0.059)	−0.119** (0.060)	−0.124** (0.060)
Probability of yes		−0.084 (0.052)	
Overall score	−0.306*** (0.070)	−0.241*** (0.074)	−0.258*** (0.073)
Age	−0.069** (0.032)	−0.067** (0.032)	−0.067** (0.032)
Importance of religion	−0.095*** (0.031)	−0.089*** (0.031)	−0.091*** (0.031)
Go out freq.	−0.282*** (0.082)	−0.281*** (0.083)	−0.283*** (0.083)
Met no	−1.072*** (0.336)	−1.095*** (0.339)	−1.054*** (0.336)
Same goal yes	−0.458** (0.180)	−0.448** (0.181)	−0.439** (0.181)
Date frequency		0.128** (0.063)	0.129** (0.062)
Rating of fun		−0.122** (0.062)	−0.129** (0.062)
Constant	1.172*** (0.383)	1.210*** (0.414)	1.049*** (0.399)
Observations	1,188	1,188	1,188
Log Likelihood	−503.994	−498.416	−499.737
Akaike Inf. Crit.	1,027.988	1,022.832	1,023.475
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 3: Pair-wise feature with wave fixed effects under Contrast Model

	Dependent variable:	
	Match	
	M4: Significant Subset	M5:Stepwise Regression
Rating of Attractiveness	−0.137** (0.059)	−0.119** (0.060)
Date Frequency		0.128** (0.063)
Preference of Intelligence		−0.031*** (0.012)
Preference of Attractiveness		0.016*** (0.006)
Probability of yes		−0.084 (0.052)
Overall score	−0.306*** (0.070)	−0.241*** (0.074)
Age	−0.076** (0.033)	−0.067** (0.032)
Importance of religion	−0.080** (0.032)	−0.089*** (0.031)
Go out Frequency	−0.287*** (0.082)	−0.281*** (0.083)
Wave2	−1.329*** (0.383)	
Wave3	−1.469*** (0.499)	
Wave11	−1.065*** (0.342)	
Wave15	−0.814** (0.376)	
Wave17	−0.934** (0.396)	
Met no	−1.001*** (0.345)	−1.095*** (0.339)
Same goal yes	−0.556*** (0.187)	−0.448** (0.181)
Rating of Fun		−0.122** (0.062)
Constant	1.956*** (0.455)	1.210*** (0.414)
Observations	1,188	1,188
Log Likelihood	−498.658	−498.416
Akaike Inf. Crit.	1,029.317	1,022.832
Note:	*p<0.1; **p<0.05; ***p<0.01, wave 4, 10, 16 omitted and are not significant	

However, with 231 participants each with individual fixed effects, we lose too many degrees of freedom when we fit the model in this way. The glm function failed to converge, despite that we tried to reduce the amount of features we put in the model. It is unclear why convergence failed, but we suspect it's due to the sheer amount of parameters put in the model.

Without better solution, we then propose to reduce the amount of individual fixed effects by proposing two models, first with male participants fixed effect and the second with female participants fixed effects. The estimate we get is effectively the average "influence" of this participants on all of his/her partners, compared to the individual that is omitted. It is not as an elegant solution, since we aim to examine the pair-wise effect of the participants which will be ideally represented by interaction between iid and pid as factors. However, realistically this separate model still can shed some light on whether a specific individual stands out on the dating spectrum.

Again we fit the model with two model selection approaches explained earlier. With significant subset approach, model fitted with male identifier (iid) as a factor and all pairwise features leads us to a group of significant individuals under the contrast model. Model fitted with female identifier (pid)) as a factor and all pairwise features fails to render any female individual as significant. With stepwise regression approach, neither iid nor pid were selected in th two separate models and both lead to same model presented in table 2 Stepwise Regression column. Among these two approaches, we prefer models given by stepwise regression due to their lower AICs. Thus, we conclude that individual effects are not significant to predict match.

Table 7 in the appendix presents the result of the model fitted with male identifier as factor and all pairwise features under the contrast framework. Although 9 male participants stand out (turns out to be significant), the interpretation of these estimates are not really insightful. These estimates represents the effect of other male participants compared to the individual with iid = 4, who was randomly assigned as the intercept by glm algorithm. Further analysis can be done by assigning a meaningful intercept such as someone with median age or median preference. Yet such analysis is out of the scope of this study since we want to know more about the general trend rather than individual traits in dating behaviors.

6. The Teamwork Model

In the Teamwork Model, we regress match on the sum of continuous features of participants in a pair to understand how similarity of participants influence whether they make a match. The sum gives us some idea of the direction in which the two participants are similar to each other. We hope to see how similarity represented by either large sums or small sums will influence match.

6.1 Pairwise features

From the result given by Table 4, we would still favor the model given by stepwise regression, M7, with its lower AIC. Note that stepwise regression picks up only one additional feature Go out frequency, which is only significant at $\alpha = 0.1$ level. In this case, we are indifferent to these two models and conclude that if we want to simpler model, model given by significant subset would suffice.

The estimates are positive for overall score, rating of attractiveness, rating of shared interest and probability of partner saying yes in descending order of estimates' magnitude. Positive estimates mean that for these features, a larger sum of participants will make it more likely for them to make a match. A unit increase in the sum of score in overall score will makes it almost 1.5 times more likely to make a match, while a unit increase in sum of rating of attractiveness makes it 1.25 times more likely to make a match.

The estimate is negative for having the same goal (same goal yes), rating of ambition and importance of race in descending order of estimates' magnitude. As a factor, the estimates for having the same goal being -0.462 means that having the same goal makes two participants 1.58 times more likely to fail to make a match compare to those who have different goals.

6.2 Pair-wise features with wave fixed effects

The result with save as fixed effect given by Table 5. We still favor the model given by stepwise regression, M9, due to its lower AIC. Wave does not show up to be significant as a major predictor in this model. And the model given by stepwise regression is the same as M7 in Table 4, i.e. model given by stepwise regression without wave fixed effects. In this case we conclude that wave is not an significant predictor of match.

6.3 Pair-wise features with individual fixed effects

Under the Teamwork model framework, setting both male identifier and female identifier as features in a model leads to a failure of convergence. Here the reasoning is the same as in the previous model: we suspect that this is due to the amount of parameters put into the model. Again we fit male identifier and female identifier separately in two models and examine the "average" effect of either a male or a female under the Teamwork framework.

The result is similar to when fitting individual fixed effects under the Contrast Model. With significant subset approach, model fitted with male identifier (iid) as a factor and all pairwise features leads us to a group of significant individuals: this time more male participants stands out (see Table 8 for details). Model fitted with female identifier (pid)) as a factor and all pairwise features fails to render any female individual as significant. With stepwise regression approach, neither iid nor pid were selected in th two separate models and both lead to same

Table 4: Pairwise wise feature under Teamwork Model

	Dependent variable:	
	Match	
	M6: Significant Subset	M7: Stepwise Regression
Rating of Attractiveness	0.231*** (0.051)	0.221*** (0.051)
Rating of Ambition	−0.184*** (0.047)	−0.187*** (0.048)
Rating of Shared Interest	0.092** (0.040)	0.088** (0.040)
Overall score	0.422*** (0.066)	0.421*** (0.066)
Prob of yes	0.076** (0.035)	0.075** (0.036)
Go out freq.		−0.117* (0.062)
Importance of race	−0.083*** (0.024)	−0.080*** (0.024)
Same goal yes	−0.462** (0.205)	−0.486** (0.206)
Constant	−8.892*** (0.742)	−8.194*** (0.820)
Observations	1,188	1,188
Log Likelihood	−394.201	−392.388
Akaike Inf. Crit.	804.401	802.777
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 5: Pairwise feature with wave as fixed effects under Teamwork Model

	Dependent variable:	
	Match	
	M8: Significant Subset	M9: Stepwise Regression
Rating of Attractiveness	0.232*** (0.050)	0.221*** (0.051)
Rating of Ambition	−0.163*** (0.046)	−0.187*** (0.048)
Rating of Shared Interest		0.088** (0.040)
Overall score	0.468*** (0.062)	0.421*** (0.066)
Prob of yes	0.098*** (0.034)	0.075** (0.036)
Go out freq.		−0.117* (0.062)
Importance of race	−0.084*** (0.024)	−0.080*** (0.024)
Same goal yes	−0.480** (0.204)	−0.486** (0.206)
Constant	−8.954*** (0.739)	−8.194*** (0.820)
Observations	1,188	1,188
Log Likelihood	−396.977	−392.388
Akaike Inf. Crit.	807.954	802.777
Note:		*p<0.1; **p<0.05; ***p<0.01

model presented in table 4 Stepwise Regression column. Among these two approaches, we prefer models given by stepwise regression due to their lower AICs. Thus, we conclude that individual effects is not significant to predict match.

7. Goodness of Fit

We use Hosmer-Lemeshow Test to examine the goodness-of-fit of our model. The test separates the sample by fitted value given by logistic regression, i.e. the expected probability. The samples were then ordered by fitted value in ascending order and divided into deciles. Within each decile, the test counts the number of observed 1 responses and 0 responses, and calculates the expected 1 responses and 0 responses by summing up the success probability for observations with response 1 and summing up 1-success probability for observations with response 0. The statistics then is calculated as

$$H = \sum_{q=1}^{10} \frac{(Obs.1 - Exp.1)^2}{Exp.1} + \frac{(Obs.0 - Exp.0)^2}{Exp.0} \qquad (6)$$

The distribution of H is a χ^2 with $Q - 2$ degrees of freedom.

Technically, the null hypothesis is that the observed proportion of success and failure is the same as the expected proportion of success and failure. More generally, the null hypothesis is that the model fits the data well. If we fail to reject the null hypothesis, we can conclude that the model fits the data well.

Here we summarize the model selection result with goodness-of-fit test results in the following table. M5 and M2 are the same; M9 and M7 are the same since stepwise regression did not select wave in the final model. Similarly when we fit all pairwise features with stepwise regression, we end up with M2 and M7 for Contrast and Teamwork model respectively. In the last row of table 6, we present the chi-square statistics for Hosmer-Lemeshow test and its p-value for M2 and M7. Since both p-values are greater than 0.05, we failed to reject the null hypothesis in both cases and conclude that in both cases, the model fits the data well.

Table 6: Pairwise feature with wave as fixed effects under Teamwork Model

	Contrast Model	Teamwork Model
Pairwise Feature	M2	M7
+ Wave	M5 = M2	M9=M7
+ IID	M2	M7
+ PID	M2	M7
Hosmer-Lemeshow	4.2914(0.83)	10.412(0.23)

8. Conclusion

We find that in general contrast between the two participants in a pair is counterproductive to make match. This partly statistically confirms the common sense that similarity draws people closer. Under the context of speed-dating, such finding is quite reasonable. When meeting strangers, people subconsciously try to find something in common to avoid awkwardness. Difference between the pair may be more constructive in a longer conversation. But with the 4 minute time limit, similarity simply makes communication more effective. And ore effective communication is more likely to reduce the feeling of strangeness and distance, which results in a match.

Under the Teamwork framework, we further explore the different directions of similarities and found that both participants having large ambition is counterproductive to make a match. This is probably due to the perception of ego that brought by ambition. Additionally if both participants weights religion highly, it's likely for them to make a match. Religion seems to still be the barrier of finding true love, at least suggested by the experiment.

References

[1] Fisman, Raymond, et al. *Gender Differences in Mate Selection: Evidence from a Speed Dating Experiment.* The Quarterly Journal of Economics, vol. 121, no. 2, 2006, pp. 673697. JSTOR, JSTOR, www.jstor.org/stable/25098803.

A. Individual Fixed Effect Results

Table 7: Male individual effect (iid as factor) under Contrast Model

	Dependent variable:
	Match
part_att	-0.192^{***} (0.072)
part_lik	-0.385^{***} (0.087)
age_part	-0.085^{**} (0.041)
date	0.164^{**} (0.081)
go_out	-0.296^{***} (0.105)
iid8	3.737^{***} (1.306)
iid9	2.934^{**} (1.183)
iid77	1.928^{*} (1.125)
iid86	2.508^{**} (1.112)
iid242	3.108^{**} (1.302)
iid268	2.451^{**} (1.156)
iid384	2.313^{**} (1.177)
iid416	3.648^{***} (1.304)
iid432	2.088^{*} (1.128)
part_met2	-1.672^{***} (0.415)
samegoal1	-0.594^{***} (0.216)
Constant	0.910 (0.997)
Observations	1,188
Log Likelihood	−395.733
Akaike Inf. Crit.	1,031.465
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 8: Male individual fixed effect (iid as factor) under Teamwork Model

	Dependent variable:
	Match
Rating of Attractiveness	0.344^{***} (0.071)
Rating of Ambition	-0.237^{***} (0.066)
Rating of Shared Interests	0.151^{***} (0.051)
Overall Score	0.653^{***} (0.092)
Importance of Race	-0.138^{***} (0.053)
Importance of Religion	0.069 (0.051)
date	-0.218^{***} (0.078)
iid8	5.322^{***} (1.491)
iid9	4.142^{***} (1.438)
iid22	3.529^{**} (1.478)
iid23	2.491^{*} (1.435)
iid29	3.876^{***} (1.449)
iid34	3.099^{*} (1.607)
iid35	2.752^{**} (1.389)
iid36	4.328^{***} (1.601)
iid37	3.139^{*} (1.613)
iid39	4.825^{***} (1.769)
iid61	3.993^{***} (1.459)
iid63	2.968^{*} (1.671)
iid64	3.548^{**} (1.642)
iid77	2.408^{*} (1.288)
iid78	3.163^{*} (1.676)
iid86	3.665^{***} (1.313)
iid87	3.495^{*} (2.057)
iid92	2.276^{*} (1.341)
iid242	4.577^{***} (1.527)
iid252	3.848^{**} (1.565)
iid253	3.542^{**} (1.379)
iid256	2.613^{**} (1.322)
iid260	4.456^{***} (1.369)
iid261	3.688^{***} (1.353)
iid263	2.550^{*} (1.364)
iid264	2.868^{**} (1.287)
iid266	4.223^{***} (1.575)
iid269	2.407^{*} (1.277)
iid270	2.870^{**} (1.306)
iid380	3.935^{***} (1.469)
iid382	3.756^{**} (1.521)
iid384	3.176^{**} (1.497)
iid387	3.089^{**} (1.368)
iid396	3.028^{**} (1.437)
iid416	5.579^{***} (1.683)
iid431	4.158^{***} (1.540)
iid432	2.503^{*} (1.337)
iid433	3.238^{**} (1.482)
iid438	2.584^{*} (1.448)
Same Goal yes	-0.702^{***} (0.267)
Constant	-12.617^{***} (1.815)
Observations	1,188
Log Likelihood	−283.583
Akaike Inf. Crit.	809.167
Note:	*p<0.1; **p<0.05; ***p<0.01