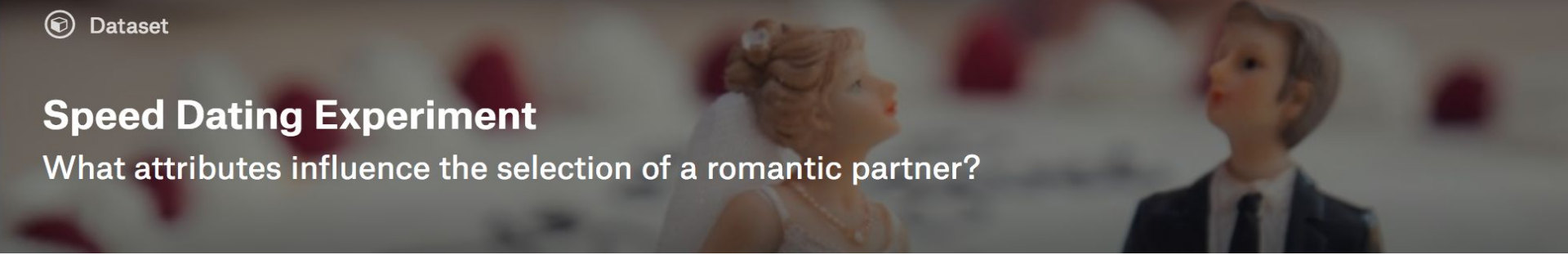


Speed Dating

UCLA Stat 402

Group 4
Tanner Koscinski, Richard Tran, James Tang, Shu Jiang

Background

- Data gathered from participants in experimental speed dating events from 2002-2004
- Over 3000 men and women
  - Drawn from graduate and professional schools at Columbia University
- Speed dates:
  - 4 minutes long
  - Person of the opposite sex
- Questionnaire given after each speed date:
  - Would you like to see your date again?
  - Rate your date on a scale from 1 to 10 on:
    - Attractiveness
    - Sincerity
    - Intelligence
    - Fun
    - Ambition
    - Shared Interest

Final Model for Males

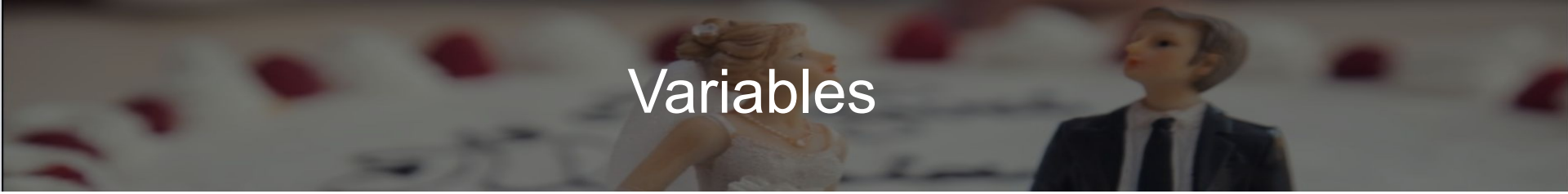
The final model for males shows that the attractiveness of the female has the strongest influence on a man wanting to go on a second date.

A one point increase in the attractiveness of the female increases the odds of the male wanting to see the female again by 2.1 times, which is both statistically and practically significant.

Other important predictors include fun, shared interests, importance of race, and dating frequency.

Abstract - Research Question

The purpose of this study is to identify attributes that influence the selection of a romantic partner (for men and women). This study uses data gathered from participants in experimental speed dating events from 2002-2004. The predictors include attractiveness, personality traits, interests, and goals. Questionnaires were given to each participant to rate their dates and ask whether they would like to see their dates again. Our logistic regression analyses showed that the most significant predictor of a person wanting to go on a second date is the attractiveness of the date.



Variables

dec - Outcome Var - Decision (yes/no) to see date again

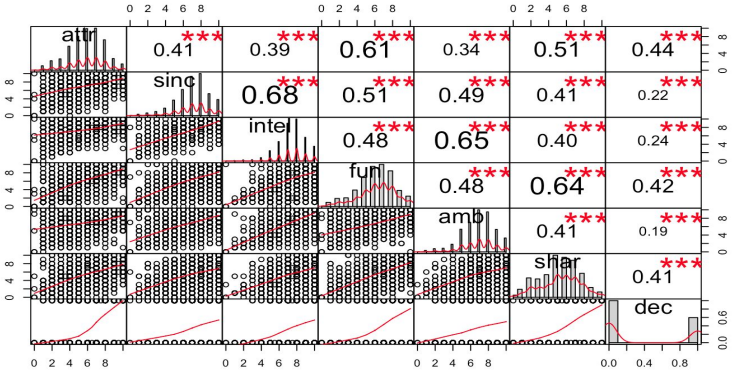
1. age - age
2. years_older - how much older is the person than his/her date?
3. attr - (1-10) Attractiveness rating for date
4. sinc - (1-10) Sincere rating for date
5. intel - (1-10) Intelligence rating for date
6. fun - (1-10) Fun rating for date
7. amb - (1-10) Ambition rating for date
8. shar - (1-10) Shared interests/hobbies rating for date
9. samerace - Both partners are the same race. 1 = yes, 0 = no
10. imprace - How important is it to you (on a scale of 1-10) that a person you date be of the same racial/ethnic background?
11. date - In general, how frequently do you go on dates?
    1 = Several times a week
    2 = Twice a week
    3 = Once a week
    4 = Twice a month
    5 = Once a month
    6 = Several times a year
    7 = Almost never
12. goal - primary goal for participating in this event
    1 = Seemed like a fun night out
    2 = To meet new people
    3 = To get a date
    4 = Looking for a serious relationship
    5 = To say I did it
    6 = Other

	Gender	
Decision	**Female**	**Male**
No	2,108 (63%)	1,776 (51%)
Yes	1,258 (37%)	1,705 (49%)

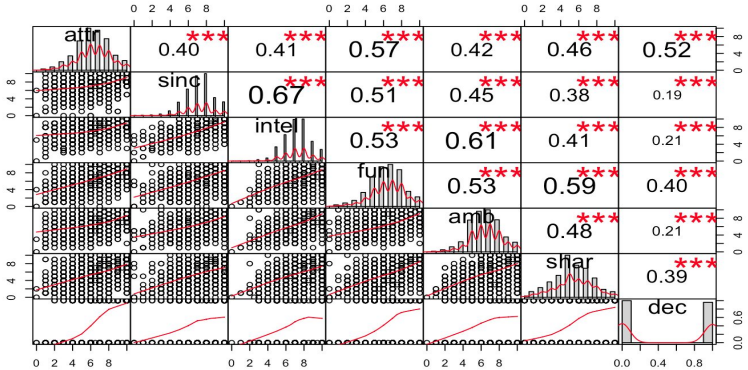
Men are more likely to say yes than women (49% vs 37%)

Correlation Matrices

Women



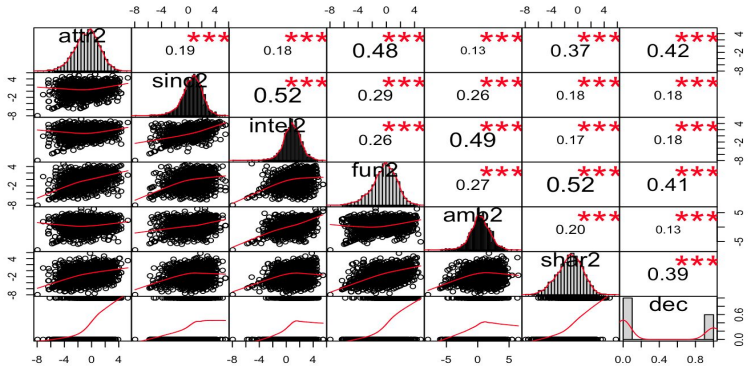
Men



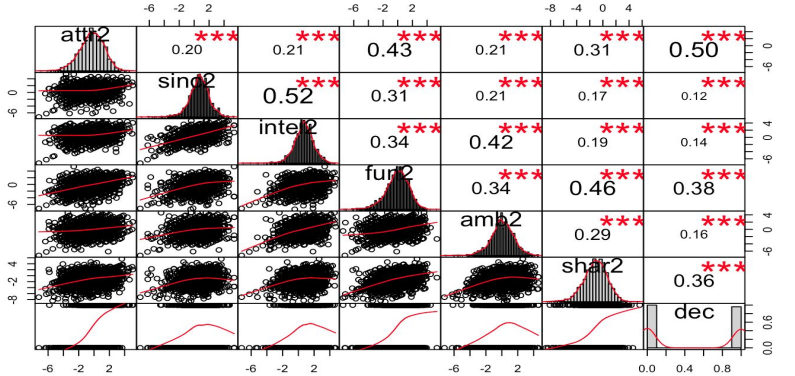
- Moderately high correlation between all predictor variables

Correlation Matrices - After Transformation

Women



Men



- Transformation: we center each person's date ratings to have a mean of 0.
- This transformation reduces the correlation between predictor variables significantly
- After the transformation, we can see there is a moderately high correlation between attractive and fun, fun and shared interests, sincerity and intelligence, intelligence and ambition
- attractiveness, fun, and shared interests are all moderately correlated with the outcome (dec).

Exploratory Data Analysis

We conducted exploratory data analysis on our variables, and based on visual inspection of the frequency tables, boxplots, and histograms, we concluded that the assumptions for logistic regression were met. For example, there was not any extreme skew in our categorical variables and there was relatively equal variance in our numeric variables relative to our binary outcome variable.

(see Appendix for a link to our code for our exploratory data analysis)

Table 1: Females vs Male Initial Model

	Dependent variable:	
	Decision	
	Females	Males
	(1)	(2)
Age	0.976 (0.942, 1.009)	0.971 (0.938, 1.004)
Years Older	1.007 (0.981, 1.033)	0.998 (0.975, 1.021)
Attractiveness	1.536*** (1.477, 1.596)	2.122*** (2.053, 2.191)
Sincerity	1.087* (1.013, 1.160)	0.984 (0.906, 1.061)
Intelligence	1.144** (1.053, 1.235)	1.010 (0.922, 1.099)
Fun	1.430*** (1.361, 1.499)	1.349*** (1.279, 1.420)
Ambition	0.942 (0.872, 1.012)	0.944 (0.874, 1.015)
Shared Interests	1.403*** (1.347, 1.460)	1.392*** (1.337, 1.448)
Same Race	0.809 (0.511, 1.107)	0.704* (0.416, 0.991)
Importance of Race	0.888*** (0.848, 0.928)	0.846*** (0.802, 0.891)
Goal: Seemed like a fun night out	1.621* (1.174, 2.068)	1.192 (0.801, 1.582)
Goal: To meet new people	2.469*** (2.024, 2.913)	1.080 (0.682, 1.479)
Goal: To get a date	2.302** (1.694, 2.910)	0.954 (0.508, 1.400)
Goal: Looking for a serious relationship	3.106*** (2.443, 3.768)	2.256** (1.685, 2.827)
Goal: To say I did it	1.840* (1.295, 2.385)	0.764 (0.230, 1.298)
Date Frequency	0.996 (0.933, 1.059)	0.883*** (0.823, 0.943)
Same Racex Importance of Race	1.109*** (1.051, 1.166)	1.094** (1.029, 1.160)
Constant	1.143 (0.060, 2.225)	9.355*** (8.343, 10.368)
Observations	3,366	3,481
Log Likelihood	-1,631.666	-1,682.203
Akaike Inf. Crit.	3,299.333	3,400.407

Note: *p<0.05; **p<0.01; ***p<0.001

Initial Model

Both the full models for females and the full model for males show that the attractiveness of the date has the strongest influence on a person wanting to go on a second date.

We can also see that people who are looking for a serious relationship are more likely to agree to a second date in general.

We will remove some of the less significant predictors and see how our results change.

Initial models - Confusion Matrices

Women

	Reference	
Prediction	**No**	**Yes**
No	**1805**	**479**
Yes	**303**	**779**
Accuracy	**76.77%*** **	**(75.30% - 78.19%)**

Men

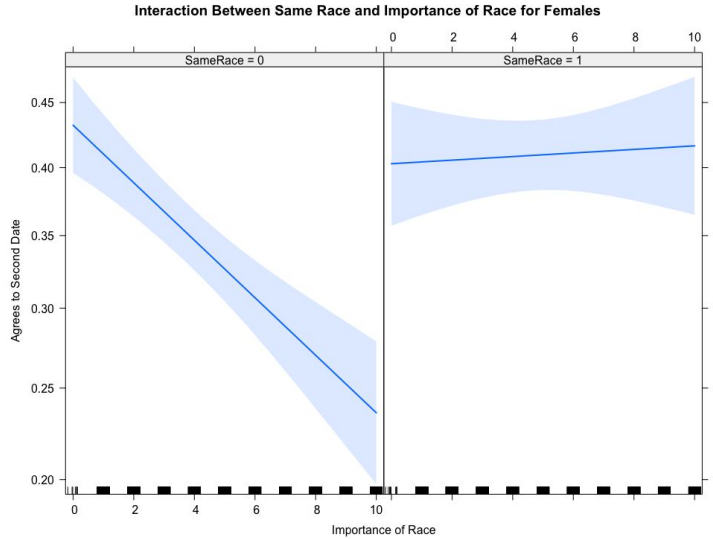
	Reference	
Prediction	**No**	**Yes**
No	**1347**	**387**
Yes	**429**	**1318**
Accuracy	**76.56%*** **	**(75.11% - 77.96%)**

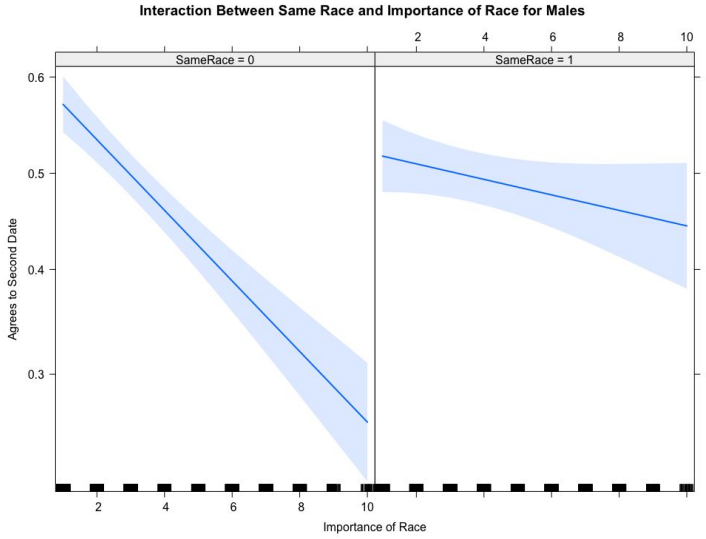
- 76% accuracy for both men and women
  - Intercept-only model accuracy: 62% for women and 51% for men.

- Goal for our final model
  - Simplify our model by removing the less significant predictors
  - Backward selection BIC (Bayesian Information Criterion)

Interaction effect (Same Race x Importance of Race)

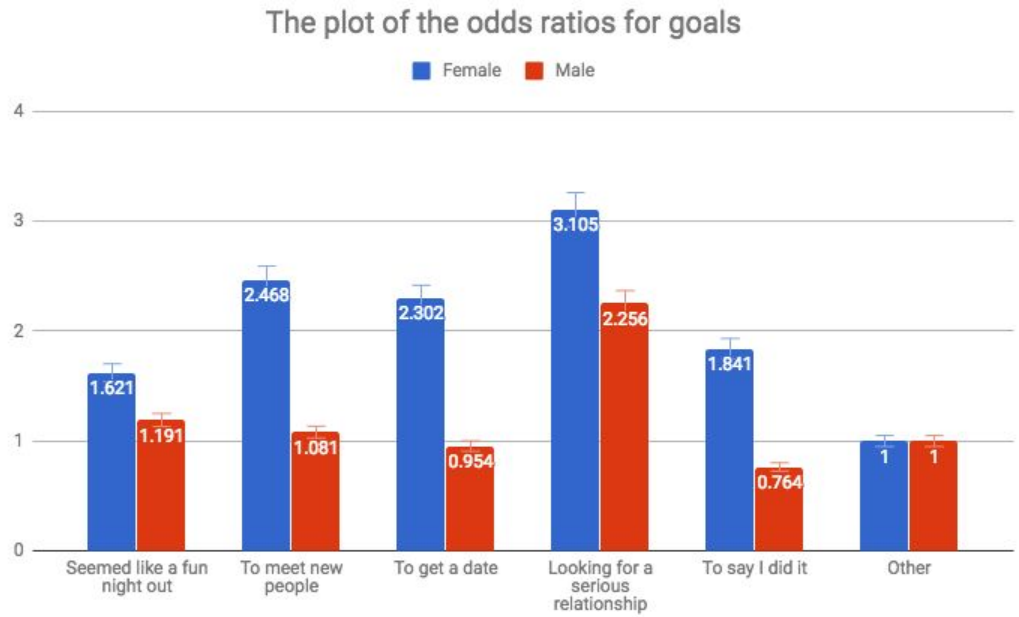
- When the dates were not the same race, women and men were less likely to say yes to a second date if the importance of race was higher.
- When the dates were the same race:
  - The women's decision for a second date did not change based on their importance of race.
  - The men's decision decreased slightly as the importance of race increased.





Plot of odds ratios for variable “goal”

Goal: What is your primary goal for participating in this event?



Takeaway: For both men and women, “Looking for a serious relationship” was the goal with the largest positive association with participants saying yes to a second date

Table 1: Females vs Male Final Model

	Dependent variable:	
	Decision	
	Females	Males
	(1)	(2)
Attractiveness	1.540*** (1.481, 1.599)	2.109*** (2.041, 2.177)
Intelligence	1.159*** (1.084, 1.234)	
Fun	1.413*** (1.346, 1.480)	1.334*** (1.268, 1.400)
Shared Interests	1.400*** (1.344, 1.455)	1.376*** (1.322, 1.430)
Same Race	0.825 (0.531, 1.118)	
Importance of Race	0.900*** (0.862, 0.938)	0.876*** (0.844, 0.909)
Same Race x Importance of Race	1.105*** (1.048, 1.161)	
Date Frequency		0.883*** (0.824, 0.941)
Constant	1.110 (0.910, 1.309)	4.168*** (3.842, 4.493)
Observations	3,366	3,481
Log Likelihood	-1,652.523	-1,699.411
Akaike Inf. Crit.	3,321.046	3,410.821
Note:		*p<0.05; **p<0.01; ***p<0.001

Final Model for Females

The final model for females shows that the attractiveness of the male has the strongest influence on a woman wanting to go on a second date.

A one point increase in the attractiveness of the male increases the odds of the female wanting to see the male again by 54%, which is both statistically and practically significant.

Other important predictors include shared interests, fun, intelligence, same race, importance of race, and the interaction between same race and importance of race.

Final models - Confusion Matrices

Women

	Reference	
Prediction	**No**	**Yes**
No	**1786**	**492**
Yes	**322**	**766**
Accuracy	**75.82%*** **	**(74.33% - 77.26%)**

- Final Model
  - Attractiveness
  - Fun
  - Intelligence
  - Shared Interests
  - Same Race
  - Importance of Race
  - Same Race x Importance of Race
- Dropped
  - Age
  - Years older
  - Sincerity
  - Ambition
  - Date Frequency
  - Goals

Men

	Reference	
Prediction	**No**	**Yes**
No	**1346**	**383**
Yes	**430**	**1322**
Accuracy	**76.64%*** **	**(75.20% - 78.04%)**

- Final Model:
  - Attractiveness
  - Fun
  - Shared Interests
  - Importance of Race
  - Date Frequency
- Dropped
  - Age
  - Years older
  - Sincerity
  - Ambition
  - Intelligence
  - Same Race
  - Goals

Cross Validation

For both men and women, our final models outperform our initial models in terms of cross validation prediction accuracy.

Men

Initial Model: 74.61% Accuracy

Final Model: 75.13%

Women

Initial Model: 72.86%

Final Model: 73.69%

Removing Variables and Checking Accuracy for Women

The full model predicting willingness to go on a second date for women includes:

Attractiveness + Shared Interests + Fun + Importance of Race + Ambition + Age + Dating Frequency + Same Race + Goal + Years Older + Sincerity + Intelligence

The full model has an accuracy of 75.82%.

Attractiveness + Shared Interests + Fun + Importance of Race + Ambition + Age + Dating Frequency + Same Race + Goal + Years Older + Sincerity + Intelligence

This reduced model has an accuracy of 75.88%.

62.63% of females say no to a second date.

Removing Variables and Checking Accuracy for Men

The full model predicting willingness to go on a second date for men includes:

Attractiveness + Shared Interests + Sincerity + Importance of Race + Fun + Age + Dating Frequency + Intelligence + Same Race + Years Older + Goal + Ambition

The full model has an accuracy of 76.64%.

Attractiveness + Shared Interests + Sincerity + Importance of Race + Fun + Age + Dating Frequency + Intelligence + Same Race + Years Older + Goal + Ambition

This reduced model has an accuracy of 76.64%.

51.02% of males say no to a second date.

Shortcomings and Recommendations for Improvement

It turns out attractiveness alone can be a satisfactory predictor of a person's willingness to go on a second date, so while many variables are statistically significant, the improvement in the accuracy of the models is not practically significant, and for convenience, it may just be easier to look at how a person rates his or her date's physical appearance to judge the likelihood of wanting to go on a second date.

Shortcomings and Recommendations for Improvement

It is possible that 4 minutes is too short of a time for someone to accurately assess the important personality traits of his or her date, but 4 minutes is enough time to assess attractiveness, so this logic could help explain why attractiveness is favored so significantly.

It would be interesting to see a similar study conducted with longer dates. We may find that attractiveness becomes less significant the longer the dates last, and important personality traits such as shared interests and fun may become more significant deciding factors.

Conclusion

Just be