

# Logistic Regression

This assignment will focus on the aggregated speed dating data, containing information about gender, race, activity participation, and career.

When doing the following, examine and interpret the coefficients of each model. Also, examine the area under the ROC curve as well as the shape of the ROC curve itself.

- Use logistic regression with the `glm()` function. In order for it to work, you need to either have a binary variable (taking on values 0 or 1) as the target variable, or convert the thing that you want to predict into a factor variable, *e.g.*,

```
df$gender = factor(gender)
m = glm(gender ~ . , family = "binomial")
```

- Predict gender in terms of self-rated activity participation.
- Restrict to the subset of participants who indicated career code 2 (academia) or 7 (business / finance). Use logistic regression to differentiate between the two.
- Restrict to the subset of participants who indicated being Caucasian (race = 2) or Asian (race = 4). Use logistic regression to differentiate between the two.