Group Beta

Jaya Jha, Jieun Lee, Tianbi Men, Qi Shi,
Xinling Xu, Henry Wyneken

May 1st 2013

Outline

- Introduction
  - Problem Statement
  - Data Exploration
- Model Selection
- Data Analysis
  - Results
- Conclusion
- Further Study

Introduction

Problem

Does an instructor discriminate among his students based on their gender and/or clothing?

Introduction

Data Collection:

– Video recording
– Two evaluators

Population:

– Male and female students
– Introductory class

Sample Size:

– 231 students

Introduction

Variables:

Instructor-Student interaction: Positive/Negative

Gender: Male/Female

Clothing Type: Unisex/Standard/Other

Introduction: N/A vs. Zero

	Positive Interaction	Negative Interaction
N/A	23	161
Zero	7	1

Total	30	162

Objective

Is there evidence of discrimination?

Data Exploration

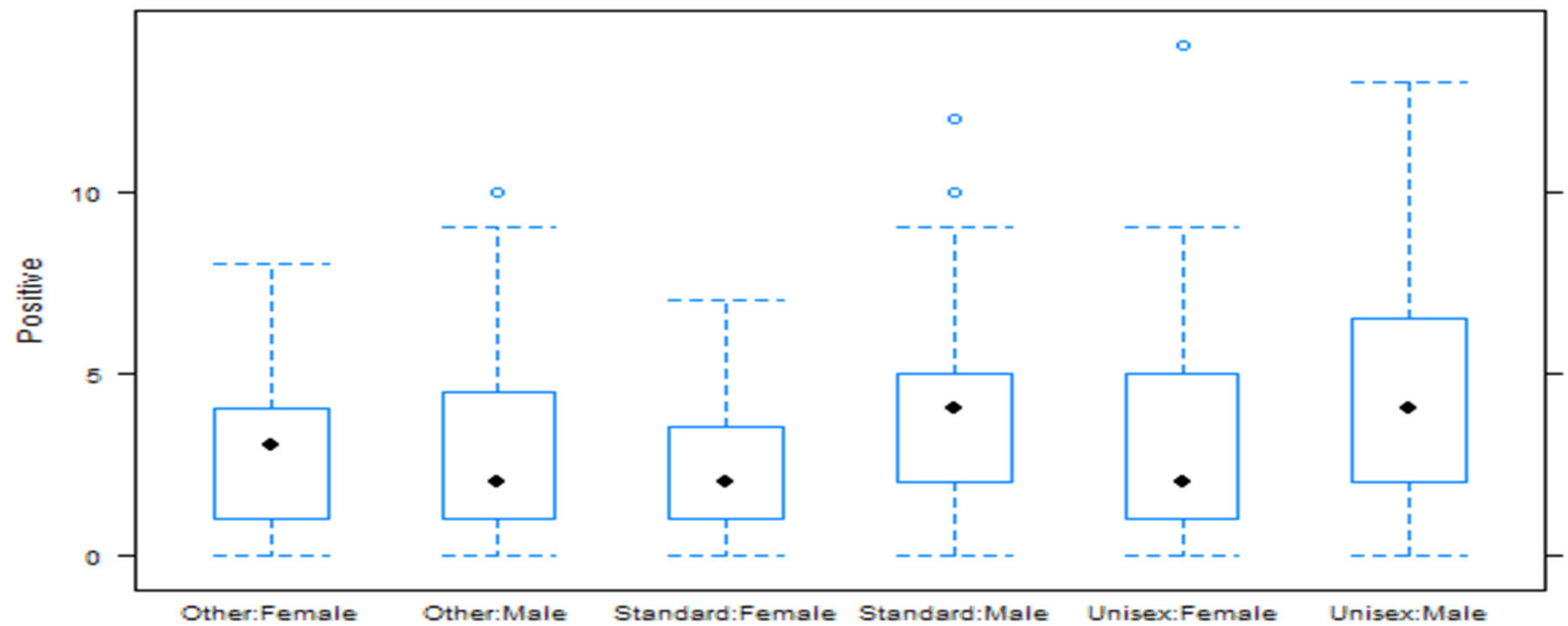
Data Summary: Sample

Female	111	48.1%
Male	120	51.9%
Total	**231**	**100%**
Unisex	54	23.4%
Standard	72	31.2%
Other	105	45.4%
Total	**231**	**100%**

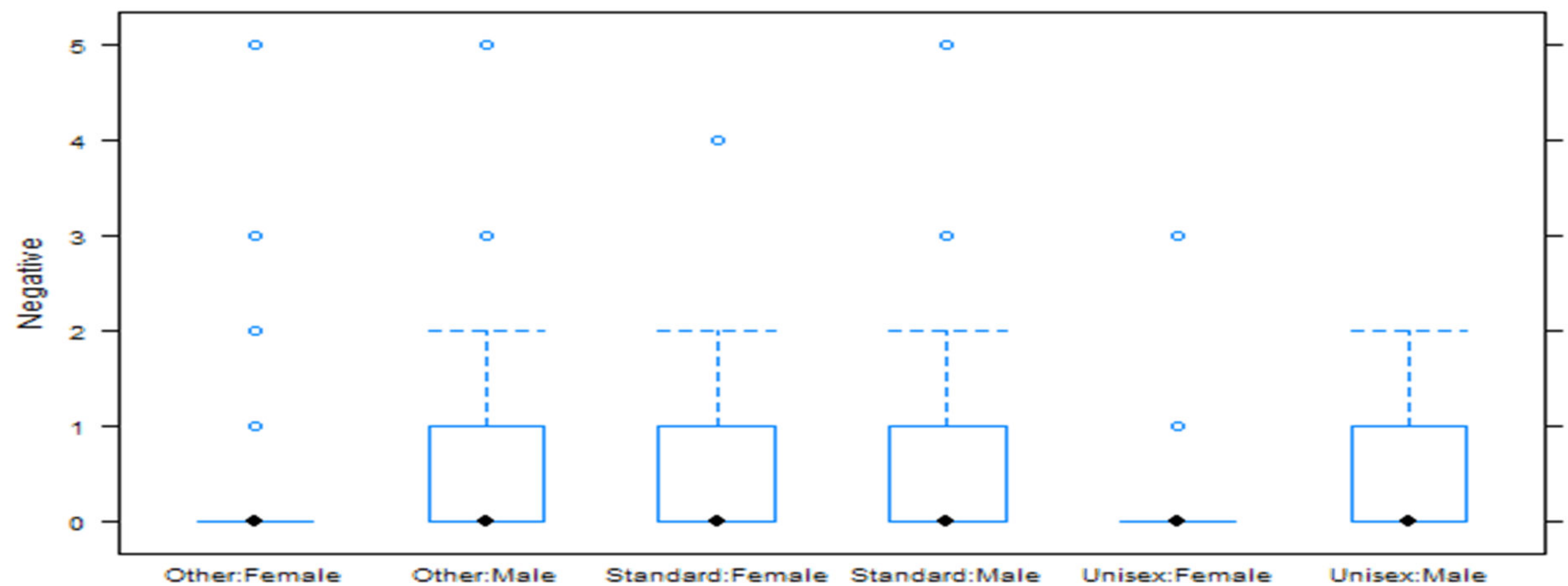
Data Summary: Sample

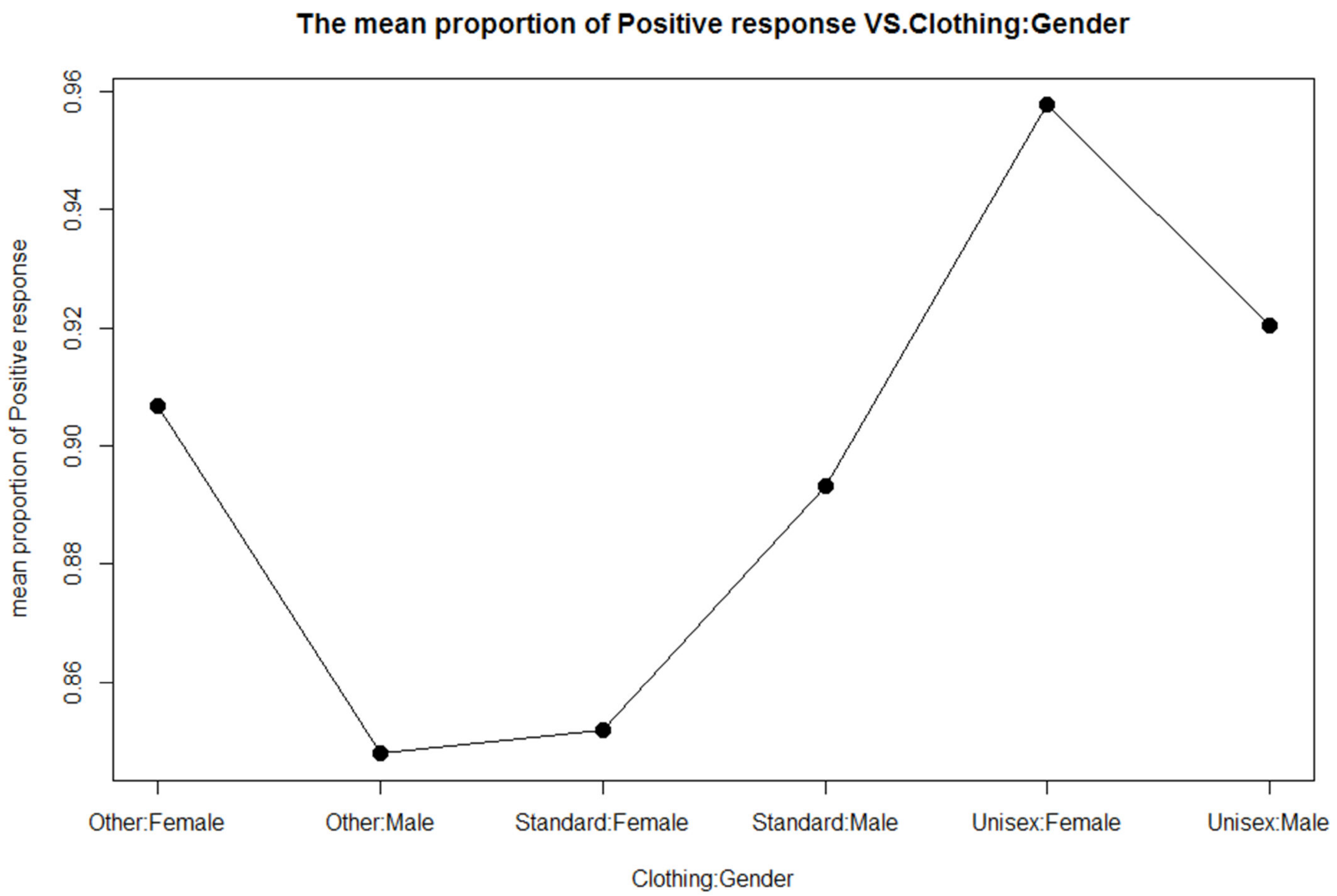
Unisex Female	19	8.2%
Unisex Male	35	15.2%
Standard Female	39	16.9%
Standard Male	33	14.3%
Other Female	53	22.9%
Other Male	52	22.5%
Total	**231**	**100%**

The distribution of Positive response VS Clothing:Gender



The distribution of Negative response VS Clothing:Gender




The mean proportion of Positive response VS.Clothing:Gender
mean proportion of Positive response
0.86
0.88
0.90
0.92
0.94
0.96
Other:Female
Other:Male
Standard:Female
Standard:Male
Unisex:Female
Unisex:Male
Clothing:Gender

Model Selection

Candidate Models

1. Poisson Model
2. Zero-inflated Poisson Model
3. Negative Binomial Model
4. Binomial Model
5. Multinomial Model

Candidate Model: Poisson

- Motivation
  - Count data, non-negative integers
- Assumptions

$$y_i \sim \text{Poisson}(\mu_i)$$
$$\mu_i = Var(y_i)$$

- Concerns
  - Highly skewed
  - mean < variance (too many zeros)

Candidate Model: Poisson

- Models

ln(Positive)=Clothing*Gender

ln(Positive)=Clothing+Gender

ln(Negative)=Clothing*Gender

ln(Negative)=Clothing+Gender

Candidate Model: ZI Poisson

- Motivation
  – Count Data
  – Many zeros, especially for Negative Feedback
- Assumptions
  – Some Zero All zero
  – Some Count Poisson process
- Concerns
  – Too few predictors (Gender & Clothing)

Candidate Model: ZI Poisson

- Model

  Positive~Clothing*Gender|1

  Negative~Clothing*Gender|1

  Positive~Clothing*Gender|Clothing*Gender

  Negative~Clothing*Gender|Clothing*Gender

Candidate Model: Negative Binomial

- Motivation:
  - Count Data
  - Overdispersion
- Assumptions

$$y_i \sim Negbin(\mu_i)$$
$$\mu_i = \phi Var(y_i)$$

- Limitations
  - Fit Positive feedback and Negative feedback separately

Candidate Model: Negative Binomial

- Models

ln(Positive)=Clothing*Gender

ln(Positive)=Clothing+Gender

ln(Negative)=Clothing*Gender

ln(Negative)=Clothing+Gender

Data Analysis

Final Model: Binomial

- Motivation:
  - Interaction=Bernoulli Experiment
  - Simplicity
  - Negative and Positive in a Single Model
- Assumptions

$$y_i \sim Bin(n_i, p_i)$$

$$y_i = \#\ Positive\ Interaction$$

$$n_i = \#\ Total\ Interaction$$

Final Model: Binomial

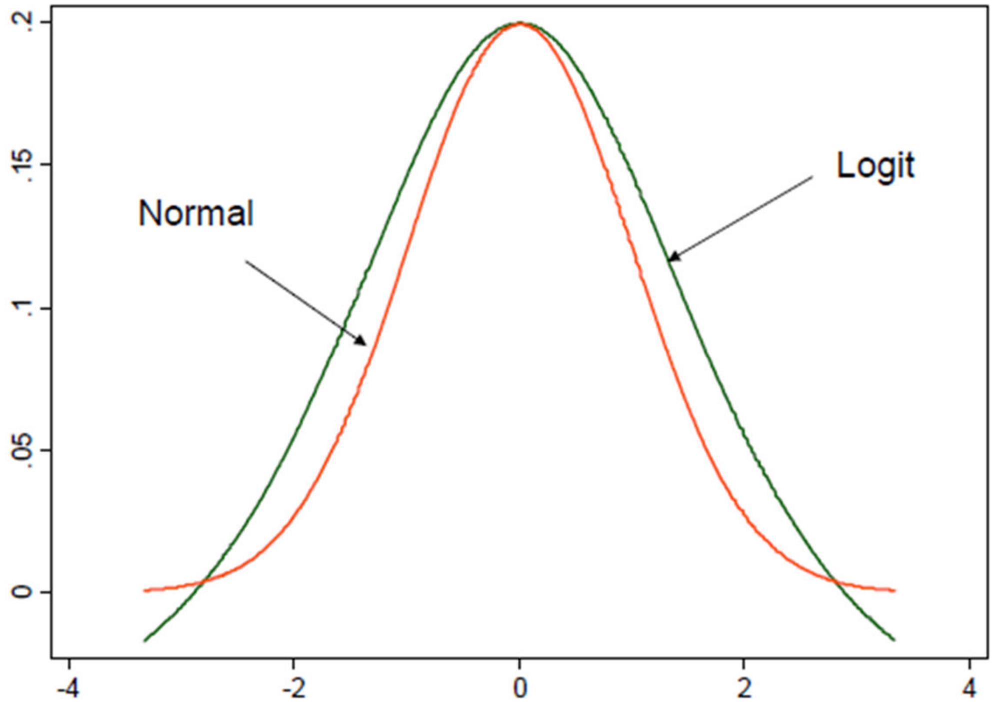
- Data Deletion:
  - 26 observations with no interaction

- R Function

glm(cbind(Positive,Negative)~Gender+Other+Unisex,

family=binomial(link=logit),data)

Final Model: Binomial



- Logit: response=log(p/(1-p))
- Probit: response=$\Phi^{-1}(\mathrm{p})$, where $\Phi^{-1}$is the inverse normal cumulative distribution function

Final Model: Binomial

- Final Model:

$$Ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = 1.72 + 0.82Unisex$$

- Model Indication

- $p_{unisex} = 92.7\% \;\; vs \;\; p_{non-unisex} = 84.8\%$

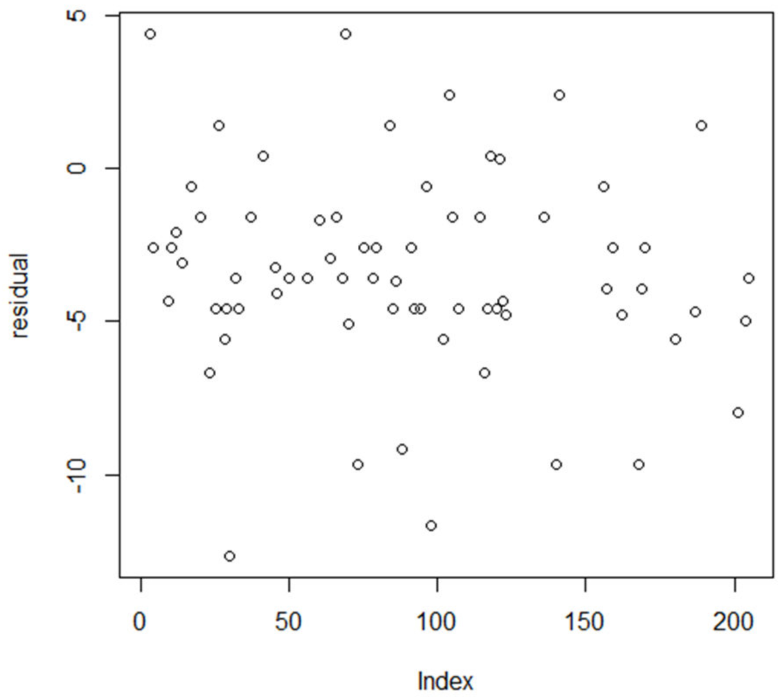
- Gender not statistically significant

Final Model: Binomial

- Limitations:
  - Low deviance explained

Null deviance: 243.51 on 204 degrees of freedom
Residual deviance: 232.72 on 203 degrees of freedom

  - Poor residual plot



Final Model: Multinomial

- Consider a restatement of the problem
- For each student, there are three possibilities
  - Only positive interactions (somePos)
  - Only negative interactions (someNeg)
  - Both positive and negative interactions (Both)

Final Model: Multinomial

- Do Gender and Clothing matter?
- No interactions: 26 Students
- Likelihood-ratio tests: Gender matters

Final Model: Multinomial

- Let Base be the base group
- Let j be the jth group
- Let x be a predictor
- Under the multinomial model:

$$\log\left(\frac{p_j}{p_{Base}}\right) = \beta_{0j} + \beta_1 jx$$

- Base group in our model: *Both*

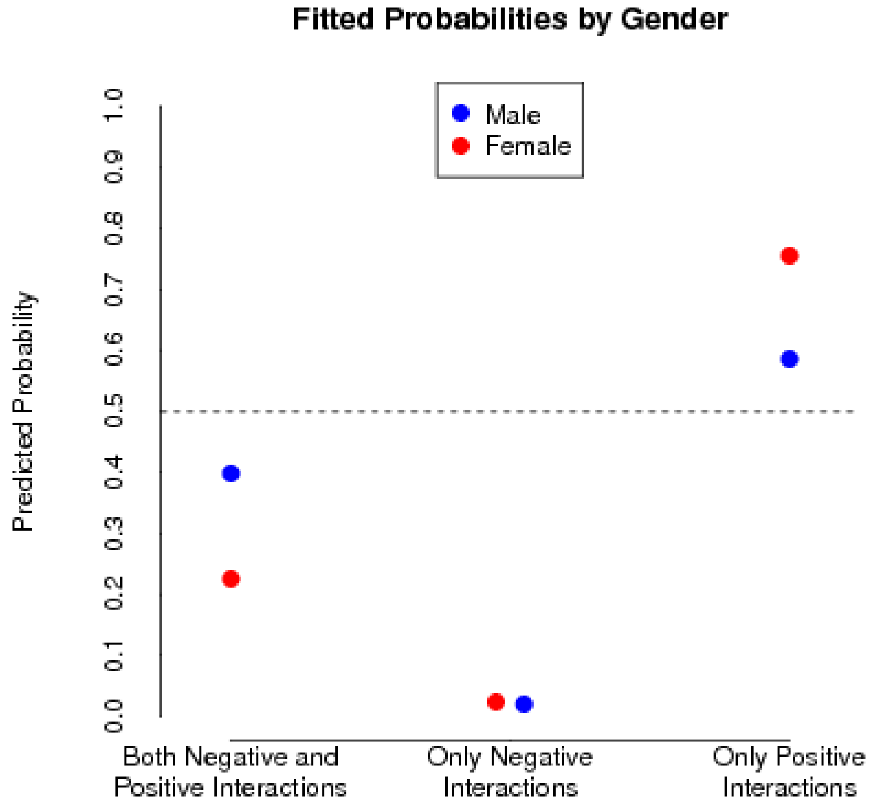
Final Model: Multinomial

$$log\left(\frac{\hat{p}_{someNeg}}{\hat{p}_{Both}}\right) = -2.35 - 0.73GenderMale$$

$$log\left(\frac{\hat{p}_{somePos}}{\hat{p}_{Both}}\right) = 1.22 - 0.83GenderMale$$

- Only the GenderMale for somePos was significant

Final Model: Multinomial



Final Model: Multinomial

- Only Positive is the most likely category
- Only Negative is the least likely category
- 20% gap for males

Conclusions

Conclusions

- Different results in the final models

  Choice of response matters

- We can measure associations, not discrimination

- Statistical significance does not equal practical importance

Further Study

To improve the study:

- Student’s academic performance (i.e. GPA)
- Student’s major
- Clearer definitions of clothing type
- More observers
- Semester evaluation by students
- Interview the four students (only negative)
- Do this study at the first week of school

Afterword: All Zeroes

	Unisex	Standard	Other	**Total**
Female	3	6	8	**17**
Male	2	2	5	**9**
Total	**5**	**8**	**13**	**26**