Final Project: Evaluating Christine L Exley and Judd B Kessler's paper "Gender Gap in Self Promotion"

Judy Xie (jx2433)

Part One: A Paper Summary

Research Goals

"The Gender Gap in Self-Promotion" by Christine L. Exley and Judd B. Kessler examines the gender gap in self-promotion and its potential effects on career advancement. Evidence of gender gap in annual earnings in financial and corporate sectors along with science, technology, engineering, and math specific fields has inspired a rich literature on factors that can help explain this difference in earnings. This article is motivated by observing whether or not individuals' evaluation of their own performance (often communicated to others through applications, interviews, performance reviews, persenations, meetings etc) may influence others assessment of them, influencing whether or not they are hired for a job or receive a promotion. Thus, the research will attempt to uncover how women and men evaluate their own performance on a math based assessment compared to how well they actually performed.

Research on how people describe their performance face challenges including 1) it can be challenging to evaluate subjective descriptions because they are qualitative in nature and difficult to measure 2) comparing the subjective descriptions of men and women who perform equally well requires observing their descriptions of a well-defined performance that can be accurately assessed 3) it is difficult to examine the underlying factors that influence subjective descriptions in settings where the environment cannot be easily controlled.

Therefore, in order to accurately document any gender gaps in subjective descriptions of performance, this study will be conducted in a carefully controlled experimental setting. By using self-evaluation questions, they were able to compare the subjective descriptions of men and women who perform equally well on a stereotypically male-typed task. The goal is to identify the underlying factors that may contribute to any observed gender gaps in self-evaluations of performance.

Introducing the Data

The data used for this research come from a series of controlled experiments. The collected data includes responses and information involving around 4,000 online participants and over 10,000 aged school youth. For the scope of this project, I will be focusing only on the 3,982 online participants. Participants were from online labor markets—Amazon's Mechanical Turk (MTurk) and Prolific— and took part in one of seven versions of our study across five waves of data collection.Each participant was guaranteed a completion fee and had the chance to receive a bonus payment from one part of the study that was chosen at random. After completing all parts of the study, participants took a short follow-up survey that collected demographic information, including gender. Gender was not mentioned prior to the follow-up survey, so participants were not influenced to consider gender when answering self-evaluation questions.

Models Used

To conduct the study, they used a variety of methods to analyze the data, including regression analysis and controlled experiments. The researchers also used a behavioral model to better understand the underlying psychological mechanisms that may contribute to the gender gap in self-promotion.

In terms of models used, they first ran three simple linear regressions using the entire dataset (not divided into waves) to find how gender affected 1) the performance (number correct answers from the math based test) 2) their belief (how well participants believed they did on the test) and 3) the belief gap (the difference between how well they believed they did and their actual performance). The coefficients helped them determine if gender had an affect on one's perception of their performance. They used their results from this regression as controls for their design.

Next, to account for the persistence of the gender gap across study versions and because of their desire to test the boundaries of this gap. They ran linear regression models for every wave to see how gender affected the following categories 1) performance, 2) performance bucket (how well they think did on a Likter scale), 3) willingness to apply (rating on a scale the answer to this statement: "I would apply for a job that required me to perform well on the test I took in part 1" and 4) succeed (rating on a scale their answer to this statement: "I would succeed in a job that required me to perform well on the test I took in part 1.) Finally, robustness tests were then done to confirm and complicate their findings.

Conclusions and Key Results

The main findings of the paper indicate that there is indeed a gender gap in self-promotion, with men being more likely to engage in self-promotion than women. This could be seen in the statistically significant negative beta coefficients for the gender variable across almost all the models when it came to how women undervalued themselves. The results indicated that women would generally under-estimate their performance compared to men, no matter the condition- whether they were informed about the average score or not or even in the face of incentives. The only exception was found in the nature of the task- that is, there was no statistically significant effect when it came to the verbal (or non-stem) evaluation. The authors also found that this gap is linked to gender differences in risk-taking and negotiation behavior. Additionally, the researchers found that the gender gap in self-promotion can have negative consequences for women's career advancement.

The paper concludes that the gender gap in self-promotion is an important issue that needs to be addressed in order to promote gender equality in the workplace. The authors suggest that interventions aimed at reducing the gender gap in self-promotion, such as training programs that teach women how to negotiate and take risks, could help to level the playing field for women in the workforce.

Part Two: An Introduction to my Report

For this report, I wanted to evaluate one of the paper's first findings, where they explored the relationship between gender and belief_gap. I reproduced their linear regression model and then decided to evaluate the model by adding education number as a feature and re-running it on an edited dataset. This dataset had a more equal distribution of age amongst the two different genders because I removed outliers for ages where men were overrepresented. My evaluations revealed that when making these adjustments to the model and the data, the beta-coefficient of females was reduced thereby weakening its relationship with belief_gap. I also ran cross-validations on all the models to determine how well they fit the data and found complications for the original model in the cross-validation coefficient and my "R^2" value, which indicated that perhaps the model was not fitting the data well. By fine tuning the data and model, I was able to reduce the beta coefficient in both cases, this complicates the paper's research by indicating that the relationship between gender and beleif_gap, while still significant, is perhaps not as significant when we consider more factors.

Part Three: A description about the data that is used

The Data Set at Hand: Can See Visuals in my Appendix

To better understand the data that was collected, I decided to take a closer look at the 3,928 participants as a group. Since we are concerned about gender's effect on self evaluation, I wanted to see if there was a relatively equal amount of men and women in my data i.e. not one group being over represented. I found that there were 1659 women and 2233 men (A1). I then took a look at the distribution of age in the data and found the graph to be skewed to the right, which made sense as the workforce does tend to be younger from 18-40 accurately representing the population (A2). Next, I looked at the distribution of the deservingness measure to see if our participants had certain perceptions of self before the experiment and it was multimodal spiking at a 100% deservingness measure (A3). This made me anxious so I decided to see if at least this deservingness measure was equal across men and women, while I found this to be mostly true, I did see a considerable difference for men and women at a 100 for the deservingness measure (A4).

Next, I looked at the key variables that I would be evaluating, performance, belief, and belief gap. I created bar graphs that compared gender for each of these variables (A5, A6, A7). For example, to see how "females" did on performance my bar graph indicated on the x-axis the number correct and the y-axis showed the number of females and males performance for each number. The graphs indicated that the distribution of scores across performance, belief and belief gap were all pretty symmetrical. I then created a boxplot that compared female and males belief scores (a negative indicates you undervalued yourself, a 0 indicates that you believed you did exactly as well as you performed), and a positive score indicated that you overvalued your performance) (A8) . The boxplot indicated that the general distribution of females' belief score was lower with an average of -1.159735 compared to their male counterpart of 1.536498.

My Ideal Dataset

My ideal dataset would have men and women more evenly distributed across certain variables. For example, there would be a more equal number of men and women across the different ages. Further, I would like if my dataset had more variables for the exploration of complex relationships between variables. For example, variables that accounted for the demographic that is more information on participants' social-economic status and upbringing as well as other variables like race or geographical region. Furthermore, I would also be interested in participants' careers, a score of whether or not they view their career as one that is based in stem. These additional variables would give me a better picture of the dataset , to discern if it is representative of the population being studied, so that the results of the analysis can be generalized to the broader population.

Implication on The Research Question

The demographic variables in this data set only include age, gender, education number, as well as republican leaning. The lack of context regarding participants careers as well as other demographic variables could affect the result of the research question because we would potentially be working with a biased dataset. For example, if most of the women within this dataset were in fields that required performing more verbal tasks over stem tasks from a day-to-day basis, the results of the study would not be generalizable. Furthermore, self-evaluation could also be affected by many other factors including one's socio economic status growing up, upbringing in general, as well as race, having information regarding these variables could help create a better model for this study.

Part 4: Reproducing the results

For this project, I decided to focus on reproducing the CDFs graphs from Figure One of the paper as well as its corresponding linear regression table (B1, B2, B3). The table indicates the result of three different linear regression models that evaluate genders effect on num_correct ie Performance, abs_belief ie a Belief Score, and lastly the belief_gap which indicates the difference between the num_correct- (how well an individual did) and their belief (how well they believed they did) . Out of all the results in the paper, I chose to study these three results because they are the basis for the design of the study. These results also included data from all the participants, and so it was drawing from the largest sample size, thus it would increase the power of the study, which is the probability of detecting an effect gender might have. Lastly, the other regression results in the paper underwent many robustness tests while these results were left generally alone. For these reasons, I decided to move forward with my project by analyzing this section.

Regression Results

	Dependent variable:		
	num_correct	abs_belief	belief_gap
	(1)	(2)	(3)
female	0.598***	-2.287***	-2.885***
	(0.132)	(0.141)	(0.176)
Constant	9.344***	11.053***	1.709***
	(0.086)	(0.091)	(0.114)
Observations	3,587	3,587	3,587
R2	0.006	0.068	0.070
Adjusted R2	0.005	0.068	0.070
Residual Std. Error (df = 3585)	3.911	4.165	5.196
F Statistic (df = 1; 3585)	20.394***	263.402***	269.317***
Note:	*p<0.1; **p<0.05; ***p<0.01		

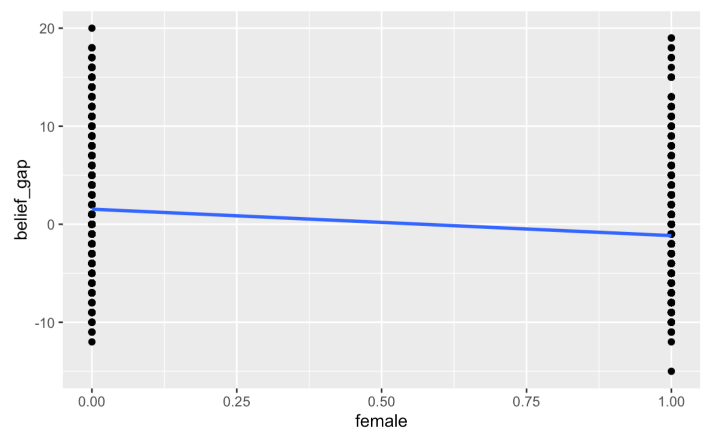
From there, I decided to hone in on the belief_gap model because out of these three models, I felt like it best represented the goal of the research paper as it finds the difference between performance and believed performance. The beta coefficient was statistically significantly at -2.8846 and the linear regression model indicated that as belief_gap increased "femaleness" or the probability that you are a women would decrease.

```
Call:
lm(formula = belief_gap ~ female, data = new.main)

Residuals:
     Min       1Q   Median       3Q      Max 
-13.8247  -3.7093  -0.7093   2.2907  20.1753 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   1.7093     0.1139   15.01   <2e-16 ***
female       -2.8846     0.1758  -16.41   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.196 on 3585 degrees of freedom
Multiple R-squared:  0.06987,	Adjusted R-squared:  0.06961 
F-statistic: 269.3 on 1 and 3585 DF,  p-value: < 2.2e-16
```

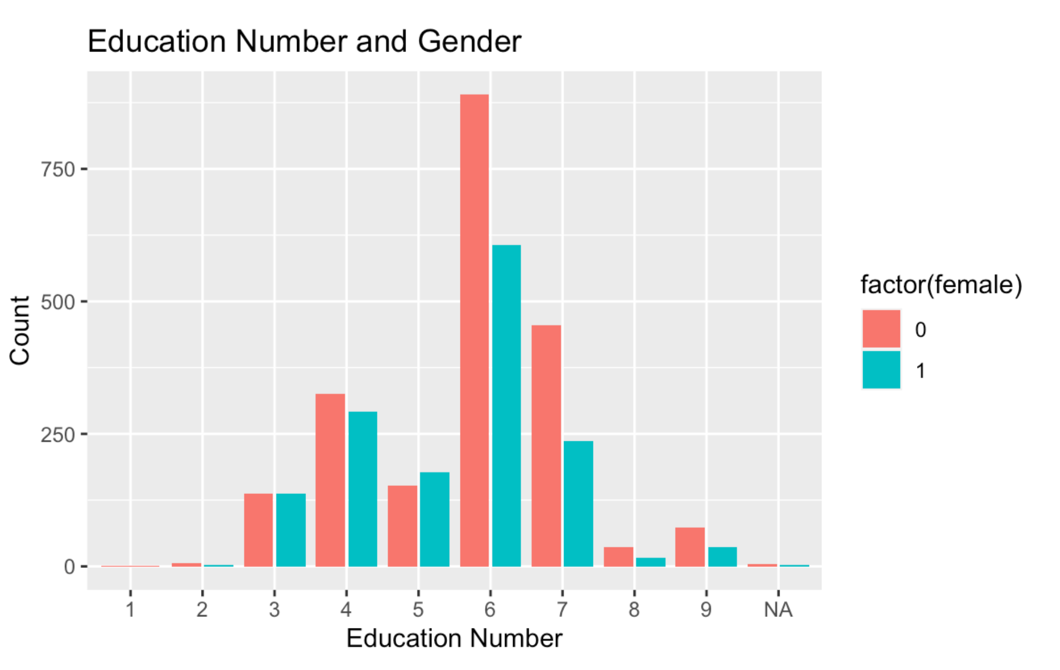


Part 5: Evaluating the Result

To evaluate my result I wanted to test the sensitivity to model choice and then test the result's sensitivity to data.

Testing sensitivity to model choice

I wondered why this particular model didn't have a confounder for the education number of the individuals in the data-set. I felt that someone's education would affect their belief_gap given that intuitively it seems to make sense that a more educated individual might feel more confident in their performance or have a better sense of their own abilities and thus would have a belief gap that is either 0 or positive, respectively. It also appears that education numbers are not evenly distributed between genders, i.e. there is not the same amount of men and women for each education number group.



Therefore, I thought it would make sense to change my model so that it had an education number as a confounder.This way, I can help control for the effects of educ_num on the relationship between gender and belief_gap in the model. This can provide a more accurate representation of the underlying relationship between my x and y. My results indicated that the educ_num coefficient was 0.77780 which was statistically significant in having an effect on my outcome. Moreover, by adding a confounder for education number we saw that our beta coefficient for females dropped from 2.8846 to 2.65216, and thus, reducing its overall relationship with beleif_gap.

```
Call:
lm(formula = belief_gap ~ female + educ_num, data = new.main)

Residuals:
     Min       1Q   Median       3Q      Max
-13.1133  -3.4611  -0.6689   2.7611  19.7611

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.77574    0.37303  -7.441 1.25e-13 ***
female      -2.65216    0.17314 -15.318  < 2e-16 ***
educ_num     0.77780    0.06179  12.589  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.085 on 3578 degrees of freedom
  (6 observations deleted due to missingness)
Multiple R-squared:  0.1092,    Adjusted R-squared:  0.1087
F-statistic: 219.4 on 2 and 3578 DF,  p-value: < 2.2e-16
```

To further evaluate my model, I decided to use-cross validation to better understand how it was fitting to my data. After cross validation, I found the Cross-Validation Correlation, an “R^2” value, and the shrinkage.

cor <dbl>	R2 <dbl>	shrinkage <dbl>
0.2595942	0.06738914	-0.006962344

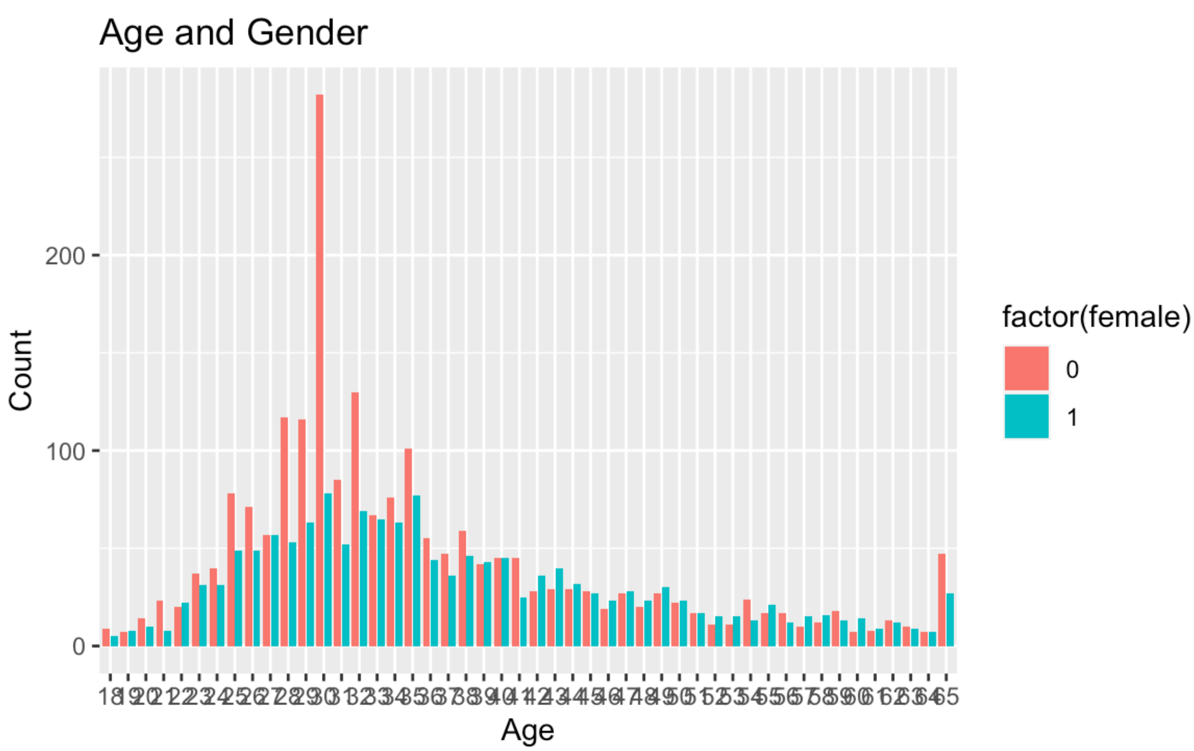
- The variable cor is the Cross-Validation Correlation = correlation between prediction and actual values. My cross-validation correlation was 0.258 which indicates that the model is not making very accurate predictions and may not be a good fit for the data. Given the context of my data set this would make sense because predicting for something like the belief gap number would be quite difficult to do because of the many factors that could lead to a person's evaluation of themselves.
- The variable R2 is found by Squaring the Cross-Validation Correlation which acts like an "R^2" value for the testing data set. My "R^2" is very low at 0.067, this means that 6% of the variance in the

dependent variable (belief_gap) is explained by the model which indicates that our model is not accurately capturing the underlying relationships in the data.

- The variable shrinkage is the difference between the "r^2" for the training data and the squared cross validation correlation. The shrinkage value is the measure of how much the performance of a model is reduced by regularization. Regularization is a technique used to prevent overfitting in a model by penalizing large coefficients, my coefficient is very small so it makes sense that my shrinkage value of 0.0069 would indicate that the regularization has had a relatively small impact on the performance of the model. This means that my model was not at risk of over fitting, which we found to be true earlier.

Testing sensitivity to data

Next, I tested the results sensitivity to data. In my data there seemed to be a large amount of men at age 31. In general there were a lot more younger men compared to women - this might affect the belief_gap as younger individuals might be less adequate at judging their own performance and might over-value their performance which would explain the -correlation in gender (x) when seeing how it might affect belief gap (y). After doing t.tests, I also saw that there is a statistical significance on gender and age (clear difference in the data) as well as belief_gap and age.



To test my result's sensitivity to data, I removed where men were over-represented in certain ages, specifically the ages 28,29,30 and 32. It appeared that age was statistically significant when determining the outcome: belief gap and my dataset indicated that there were a much larger number of men for certain ages. I decided to edit my dataset so that it randomly removed data of men who were aged 28,29,30 and 32. I ran another regression model with my edited data and found a new beta coefficient for female -2.2267. This was 0.6578288 less than my original beta. This is a 22.8 % change from the original beta and it seems quite large considering the original beta coefficient is already a very small number.

This tells me that once we account for the over-representation in men of certain ages in the dataset the slope coefficient decreases. Therefore, indicating that the line of best fit becomes flatter, which indicates a weaker relationship between the two variables. This made made me concerned about original my model potentially overfitting and what that would say about my beta coefficient result, and so I wanted to compare the mean of the squared residuals between training and test data sets for each model (my original model and the one I derived from my changed data set) to see if my new model ie the one with edited dataset has a better fit.

After doing Cross validations with 17 and 19 k fold tests for my original model and my edited model respectively, I found both models seemed to be fitting well. However, the difference between my training error and my testing error was smaller for my edited model (-0.03665076) compared to my original model with a difference of (-1.949176). This indicated that my edited model had a better fit to the data.

Summary of my evaluation and its implication on the model of the problem

The evaluation of my results revealed that when testing the sensitivity of the model to other features, I found that by adding the feature educ_num my beta coefficient for females decreased, reducing its overall relationship with beleif_gap. The weakening relationship, indicates that by considering relevant confounders we might have a better sense of the relationship between gender and beleif_gap . Without confounders, the resulting beta coefficient from the original model might be over-estimating gender's effect on beleif_gap and perhaps attributing a much stronger relationship between the two than what actually exists (at least for this particular experiment).

When I tested the sensitivity of my model to the data, I found that after changing the data my new linear regression model adjusted the female beta coefficient by reducing it by 0.6578288. This indicated that my data might be biased to help add significance for the alternative hypothesis- gender has an affect on belief_gap- because when I adjusted the data to reduce overrepresentation of men for certain ages my coefficient shrunk showing that the female's coefficient had less of an effect on the belief_gap, but it was still statistically significant. Furthermore, when I did cross validation tests to see which model had a better fit to its respective data, I found that my altered model and changed dataset performed better on not over-fitting as its training and test errors had a smaller difference. The difference between these two models derived from different datasets is important to consider because it might affect how applicable the paper's result is to the general population as the data and therefore the model might not truly represent the population.

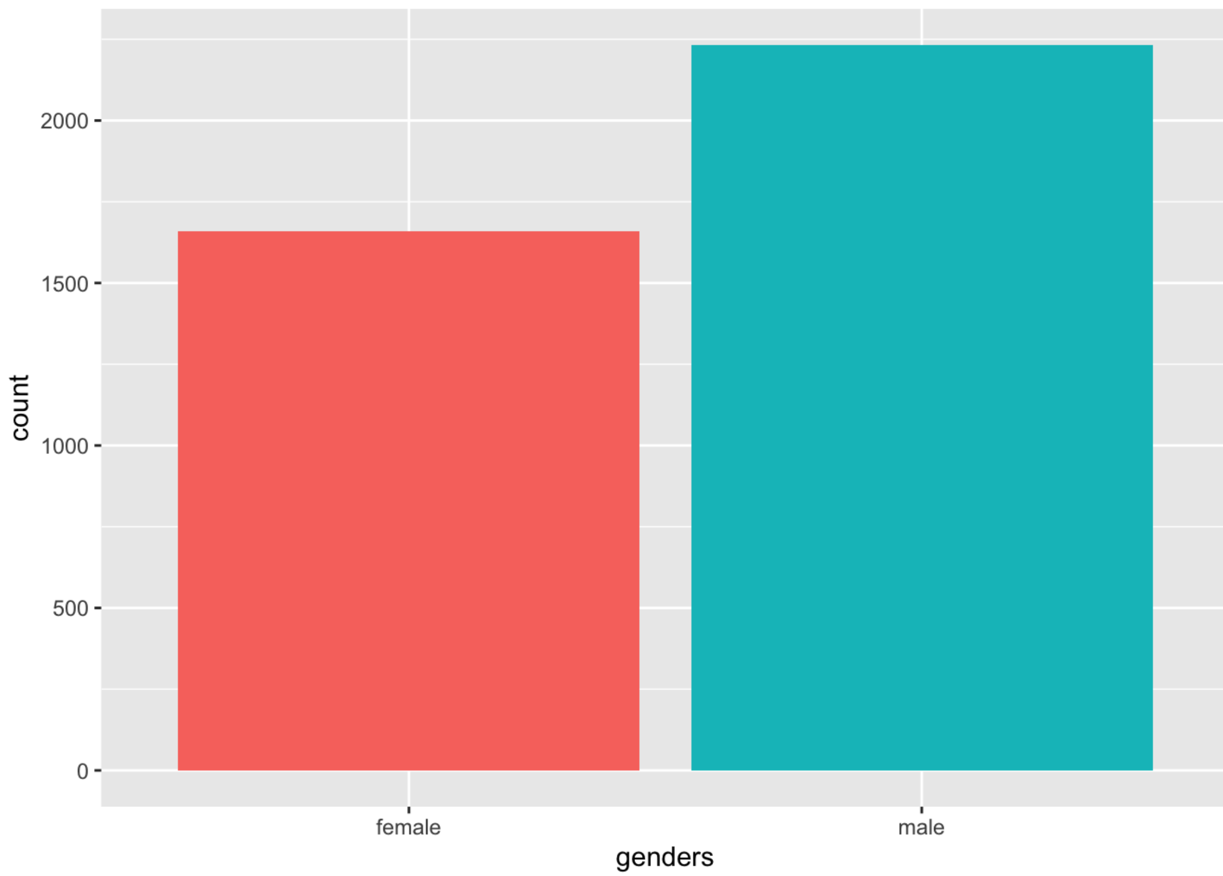
Part 6: Conclusion

To conclude, I think the findings of the paper do indicate that there exists some relationship between gender and someone's self evaluation of their performance. However, it is difficult to draw any true conclusions because there are many factors that might be cross-correlated with gender. Furthermore, my evaluation of the model within my report indicates that there are potentially many limitations to this study as well as its model. My evaluation revealed that while my model performed well and had similar errors for the training and testing datasets during cross validation, it still had a low R^2 value and weak correlations between the predicted values and the true values. Furthermore, I saw rather significant changes in its beta coefficient when I added features or changed the dataset. Therefore, considering the scope of this paper's goal, a linear regression model might be too simple to account for all the necessary variables and factors that might arise when measuring and evaluating the relationship between gender and "self- promotion".

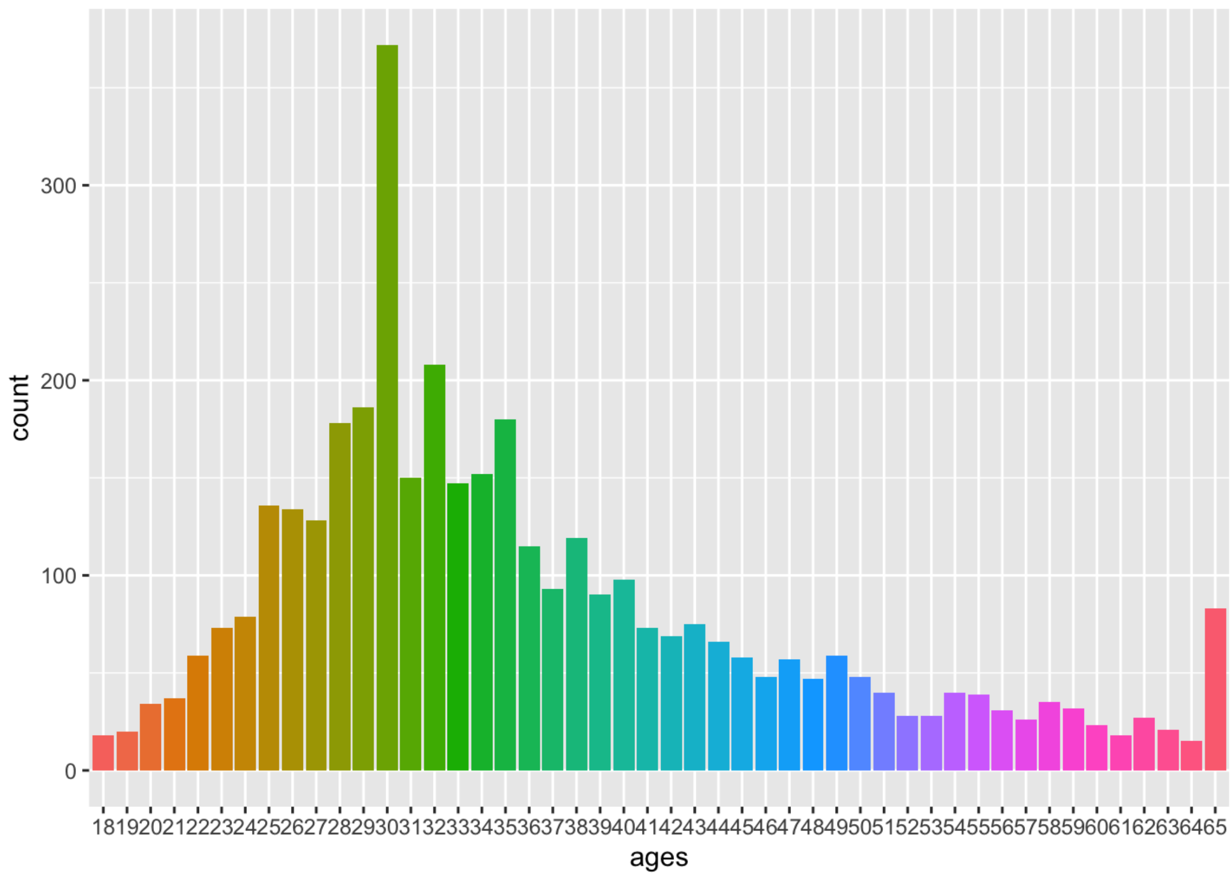
Appendix

Data Visualizations

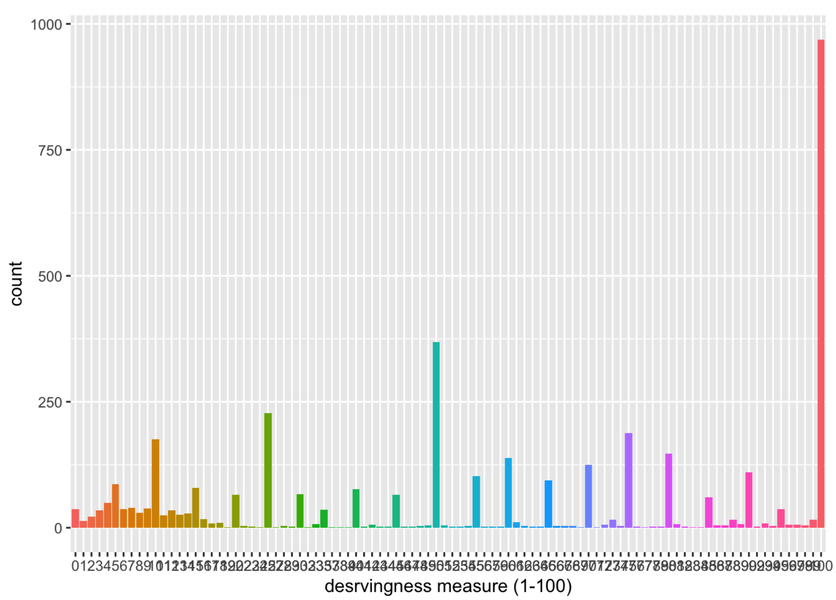
A1



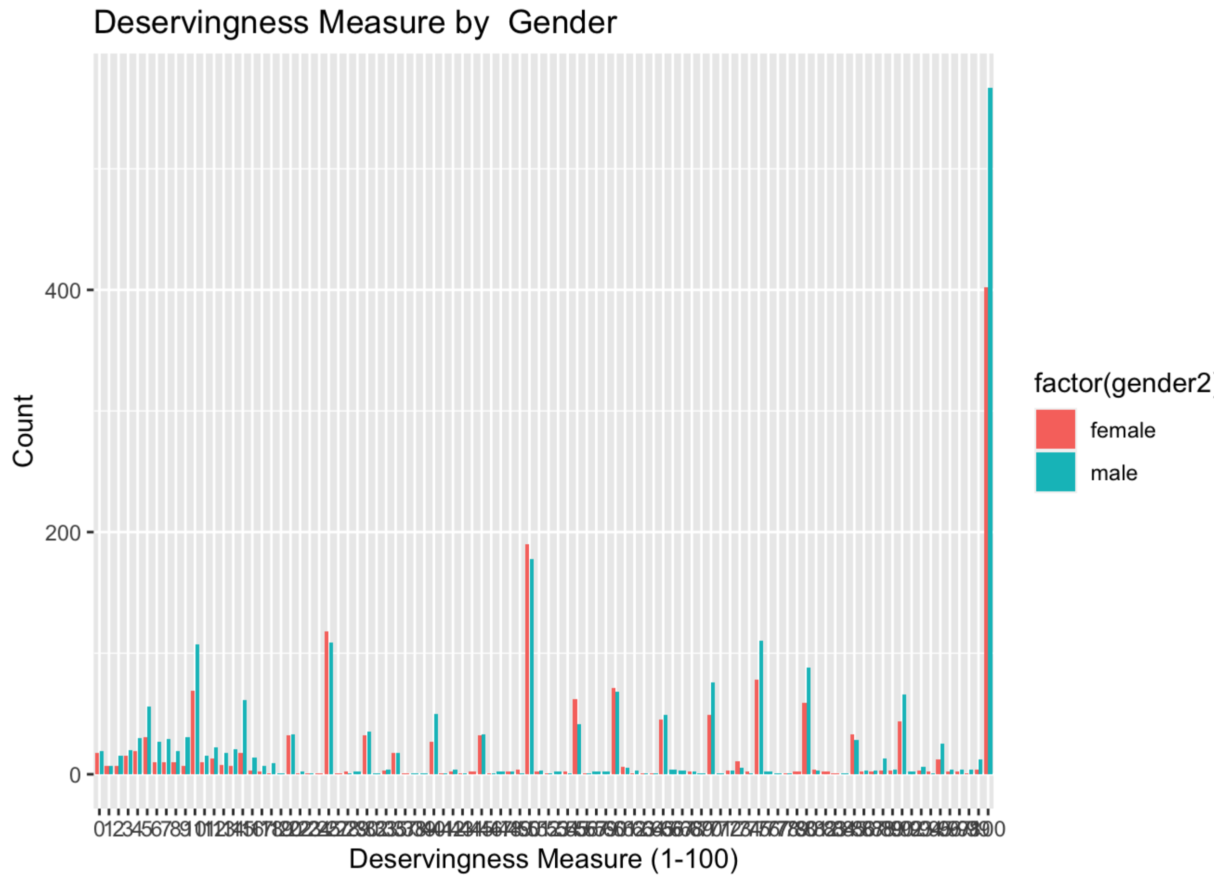
A2


count
300
200
100
0
18192021222324252627282930313233343536373839404142434445464748495051525354555657585960616263646465
ages

A3


1000
750
500
250
0
count
desrvingness measure (1-100)

A4

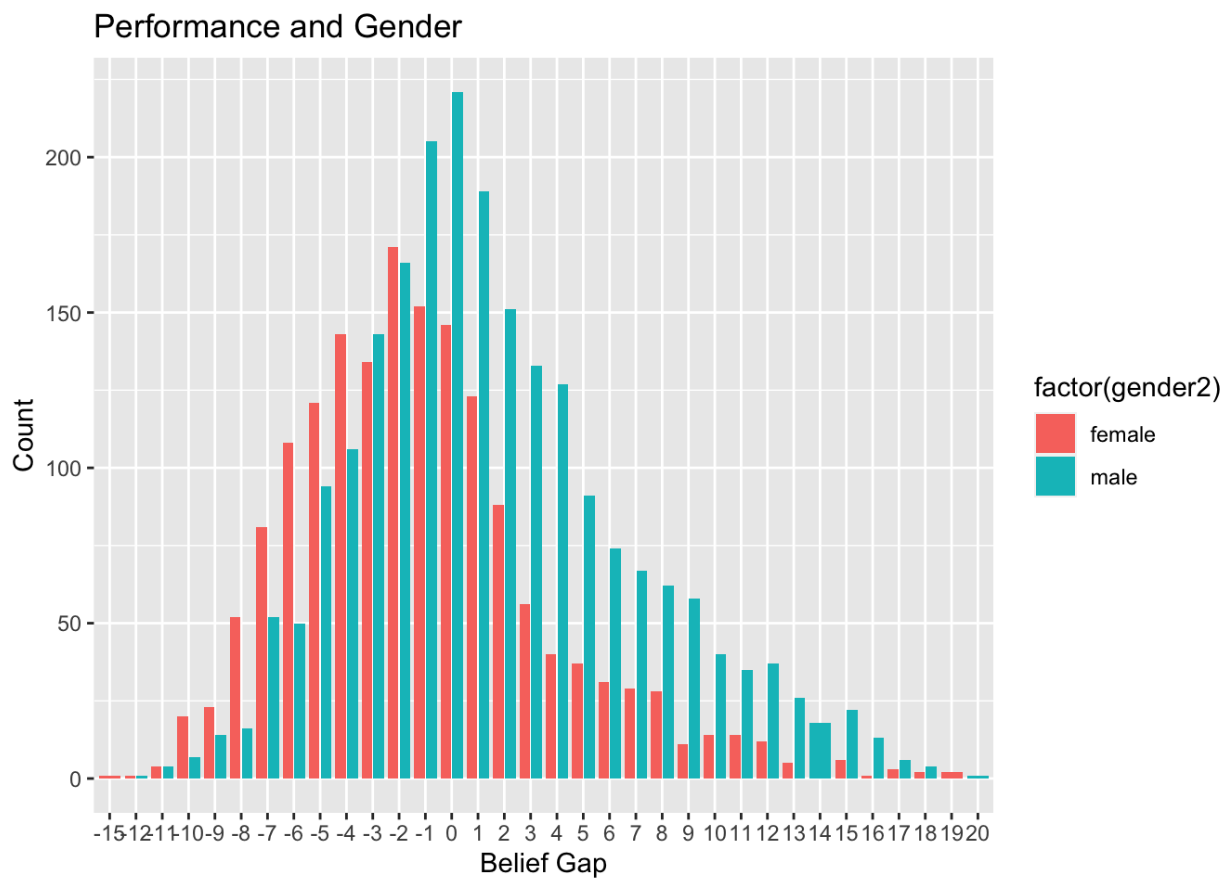


Deservingness Measure by Gender
Count
400
200
0
Deservingness Measure (1-100)
factor(gender2)
female
male

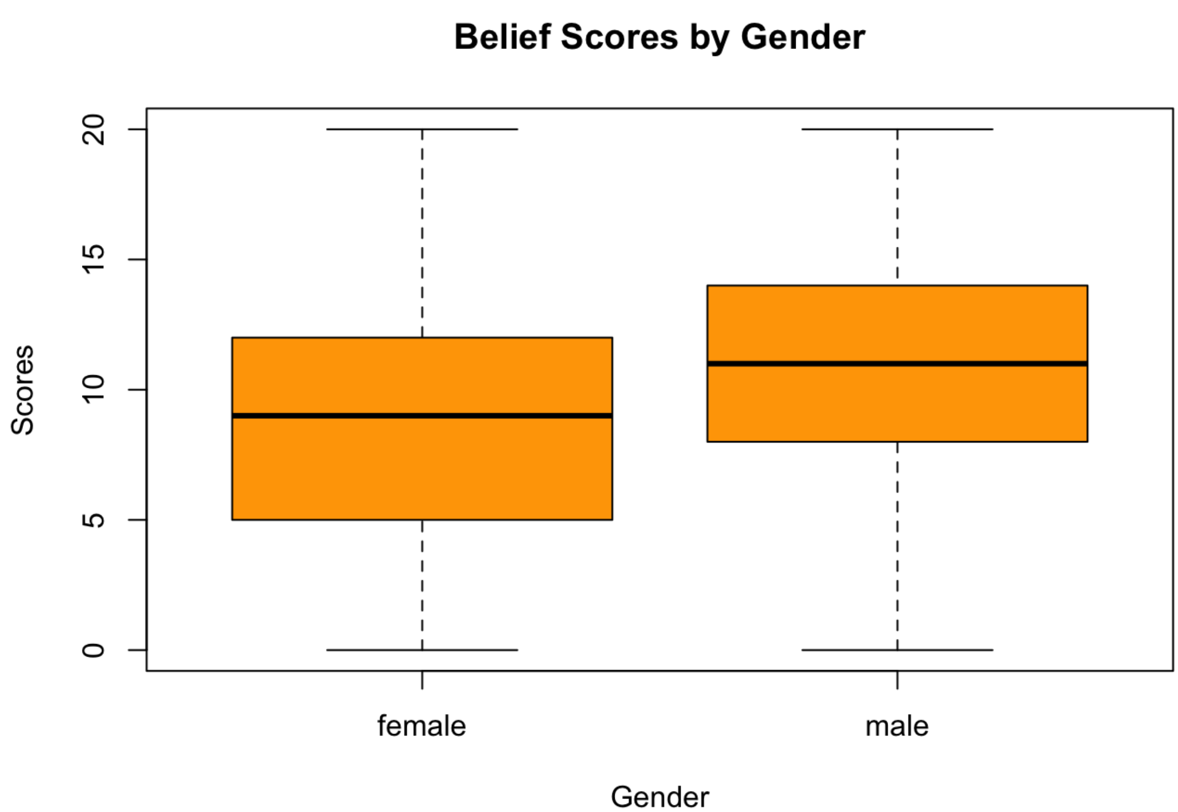
A5

Belief and Gender
Count
400
300
200
100
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Belief
factor(gender2)
female
male

A6

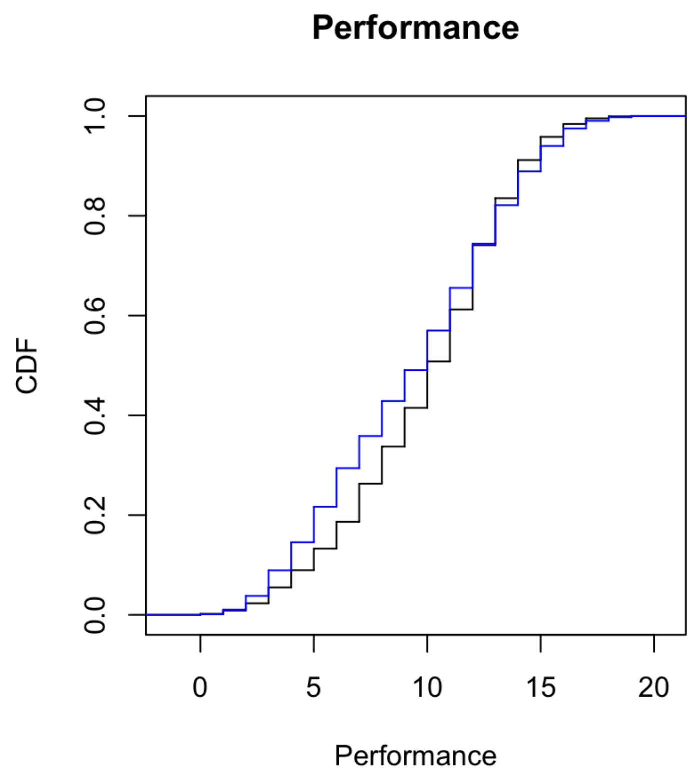

Performance and Gender
Count
Belief Gap
factor(gender2)
female
male
200
150
100
50
0
-15 -12 -11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

A7

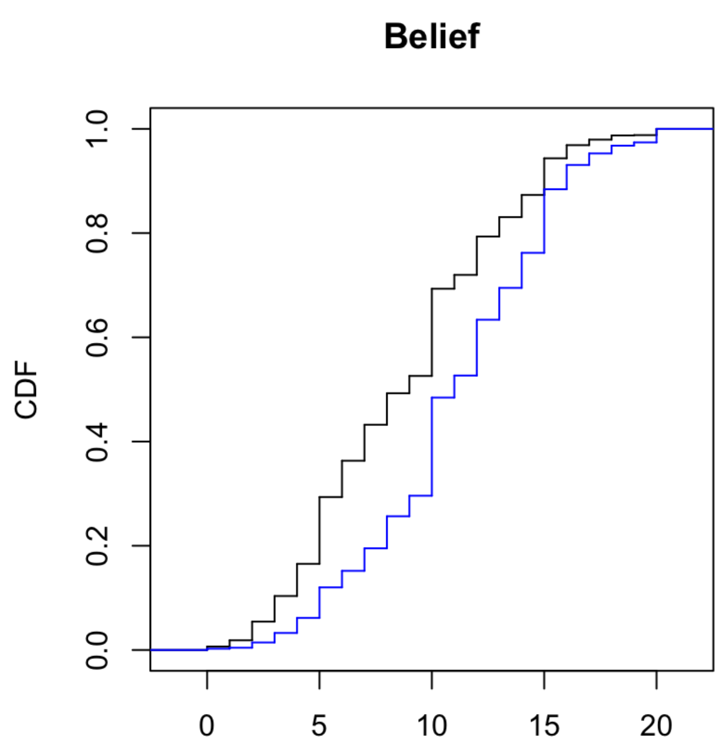

Belief Scores by Gender
Scores
Gender
female
male
0
5
10
15
20

Reproducing Results

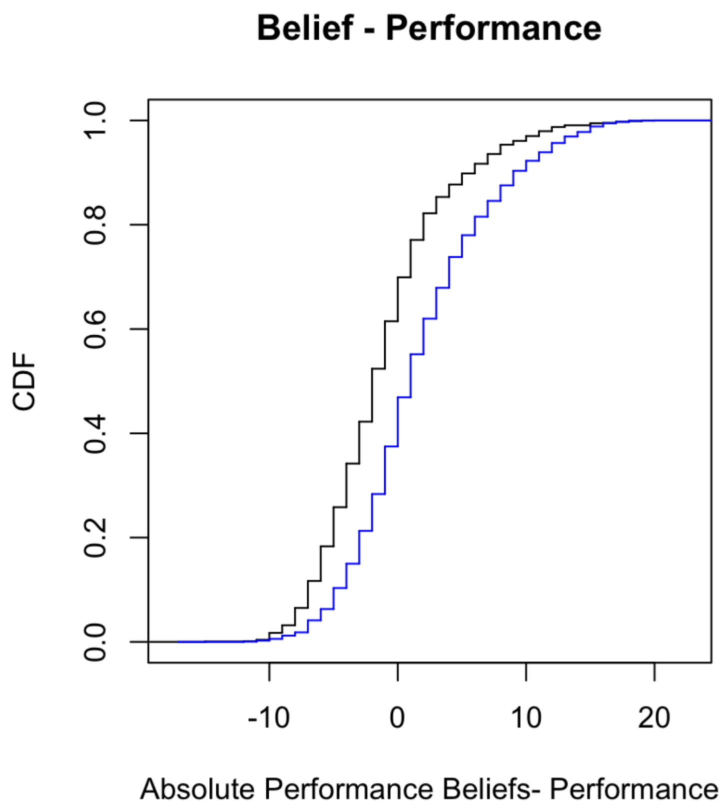
B1



B2



B3


Belief - Performance
CDF
0.0
0.2
0.4
0.6
0.8
1.0
-10
0
10
20
Absolute Performance Beliefs- Performance