An Investigation into Imposter Syndrome

Najia Pan, Neftali Lemus, Orion Shi, Sebastian Rivera-Chepetla, Yao Zhang, and Yilin Xu

Definitions of Imposter Syndrome

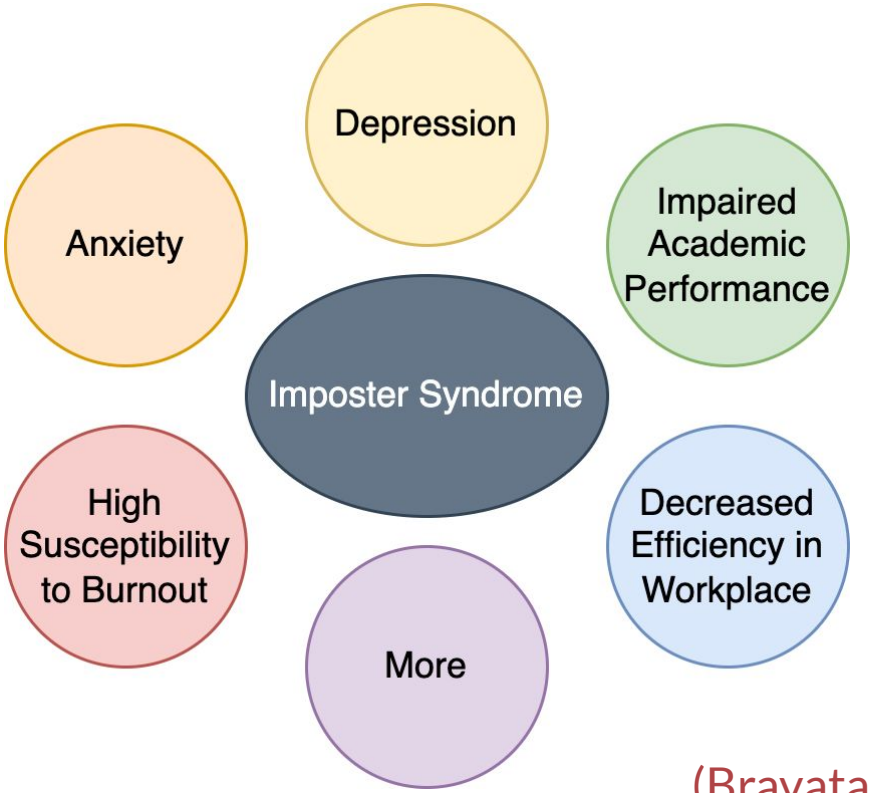
From Literature

- Experience no internal sense of success while being objectively high-achieving (Clance and Imes, 1978).
- A set of self-referential ideas centered around perceived fraudulence (Kolligian and Sternberg, 1991).
- Feelings of inauthenticity is the singular driving force behind imposter syndrome (Leary et al., 2000).

Our Definition

Imposter syndrome is a phenomenon that a person believes they are not as successful as others think they are and feels they do not deserve the position or role they currently have.

Why Should We Care?



(Bravata et al., 2020; Weir, 2013)

Research Questions

Can we predict a person's imposter syndrome score from existing scales using a newly developed scale that contains statements pertaining specifically to their experiences at UCLA?

Research Questions

Can we predict the odds of a person being impacted by imposter syndrome using statements from the newly developed scale that contains statements pertaining specifically to their experiences at UCLA?

Research Questions

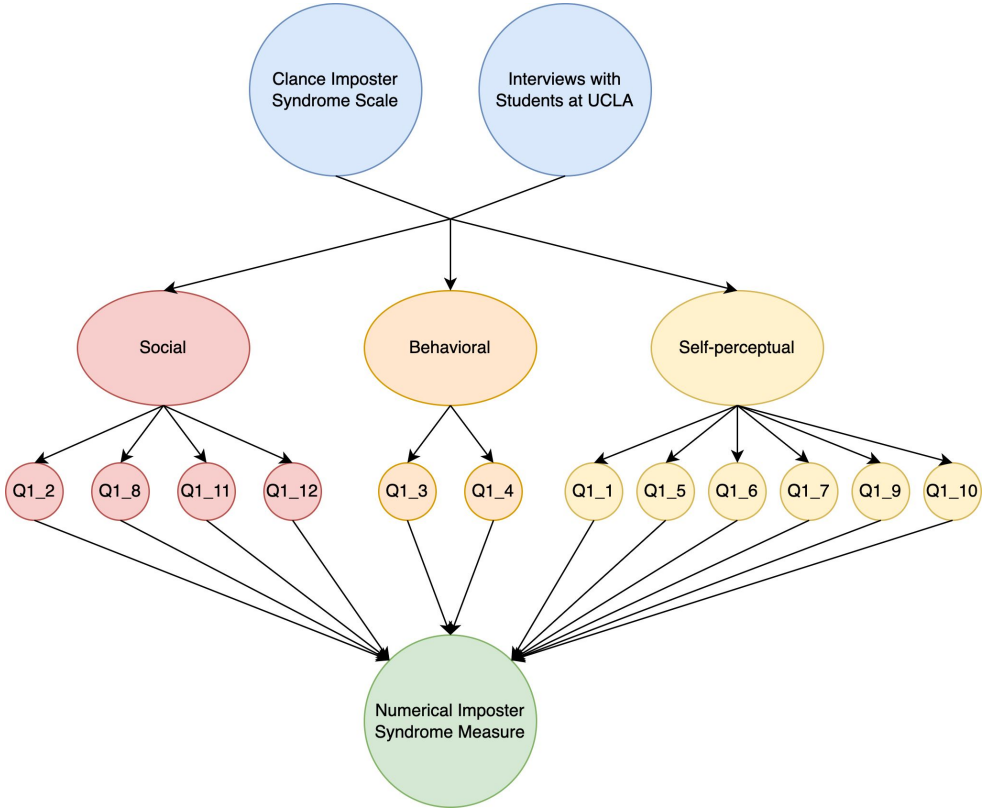
Does the design of our project have high ecological validity? More specifically, do the results from our study match with what people think of imposter syndrome in reality?

Survey Distribution

- The survey was distributed to students enrolled in STATS 140XP, PSYCH 124C, and SOC 101.
- We collected 89 complete responses.
- Participation was voluntary, and participants was able to leave some questions blank if they didn't wish to answer.

Numerical Imposter Syndrome Measure

- 12 statements were included for calculating a numerical measure of imposter syndrome.
  - Consists of statements paraphrased from the Clance Imposter Phenomenon Scale (CIPS; Clance, 1985) and those summarized from interviews with students at UCLA.
  - Contains statements from three dimensions: social, behavioral, and self-perceptual.
  - Each statement is rated from 1 to 7, where 1 indicates strongly disagree and 7 indicates strongly agree.
- Final score is calculated by summing scores on all 12 questions
  - Final score ranges from 7 to 84. A higher score means a person is more severely affected by imposter syndrome.



Example Statements from the Survey

- “I feel anxious about being judged by others.”
  - Example from the social cohort.
- “I tend to avoid challenging tasks.”
  - Example from the behavioral cohort.
- “I have high expectations of myself.”
  - Example from the self-perceptual cohort.

Scoring Scale from the Survey

Strongly disagree	Disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Agree	Strongly agree
1	2	3	4	5	6	7

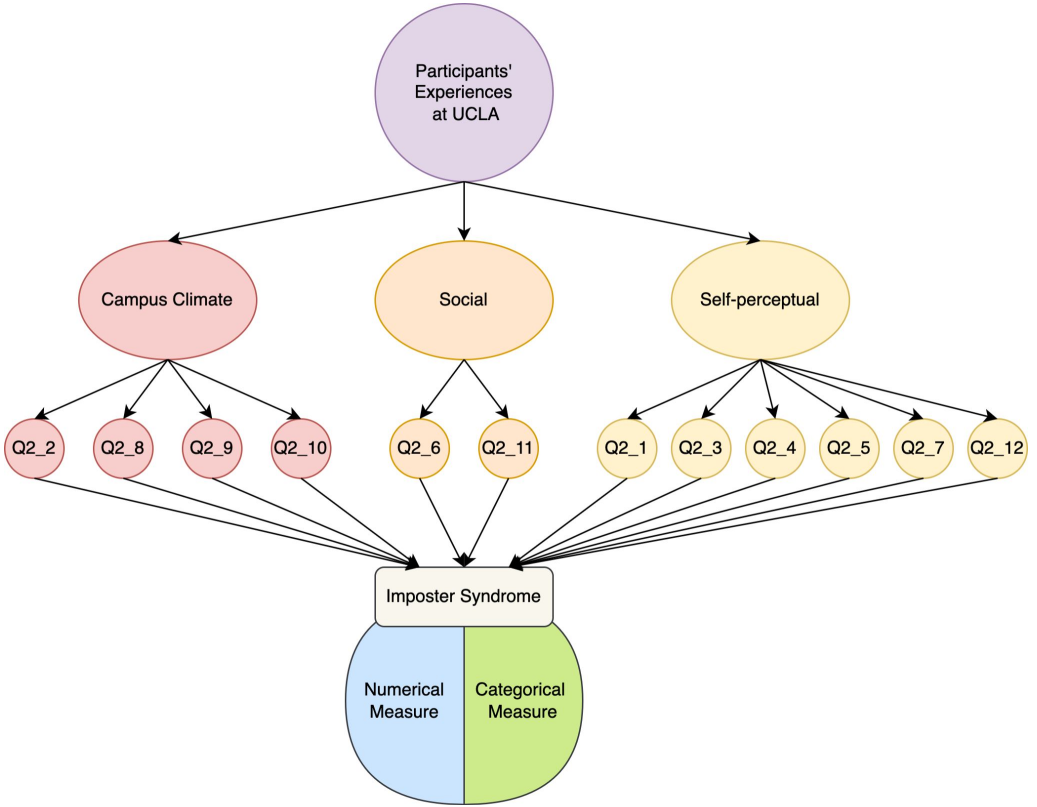
Categorical Imposter Syndrome Measure

“Do you think you are impacted by imposter syndrome?”

- A self-reported variable with three levels: yes, no, and maybe.
- More on how we used the responses from this questions later.

Predictor Statements

- 12 statements were included to serve as predictors.
  - Designed specifically to capture elements of a person's experiences at UCLA.
  - Contains statements from three dimensions: campus climate, social, and self-perceptual
  - Each statement is rated from 1 to 7, where 1 indicates strongly disagree and 7 indicates strongly agree.
- Statements were selected to avoid direct overlap with statements from the existing scales.
  - If an element is too important to be removed because of overlap, it is either modified to be an extension or a more specific version of the previous statement.



Example Statements from the Survey

- “I feel people like myself are underrepresented at UCLA.”
  - Example from the campus climate cohort.
- “I think the appraisals I get from others are insincere.”
  - Example from the social cohort.
- “I think I am not academically competent.”
  - Example from the self-perceptual cohort.

Example Statement Modification

“I attribute my success to luck rather than hard work.”

Luck, or locus of control, is the element of imposter syndrome captured here.

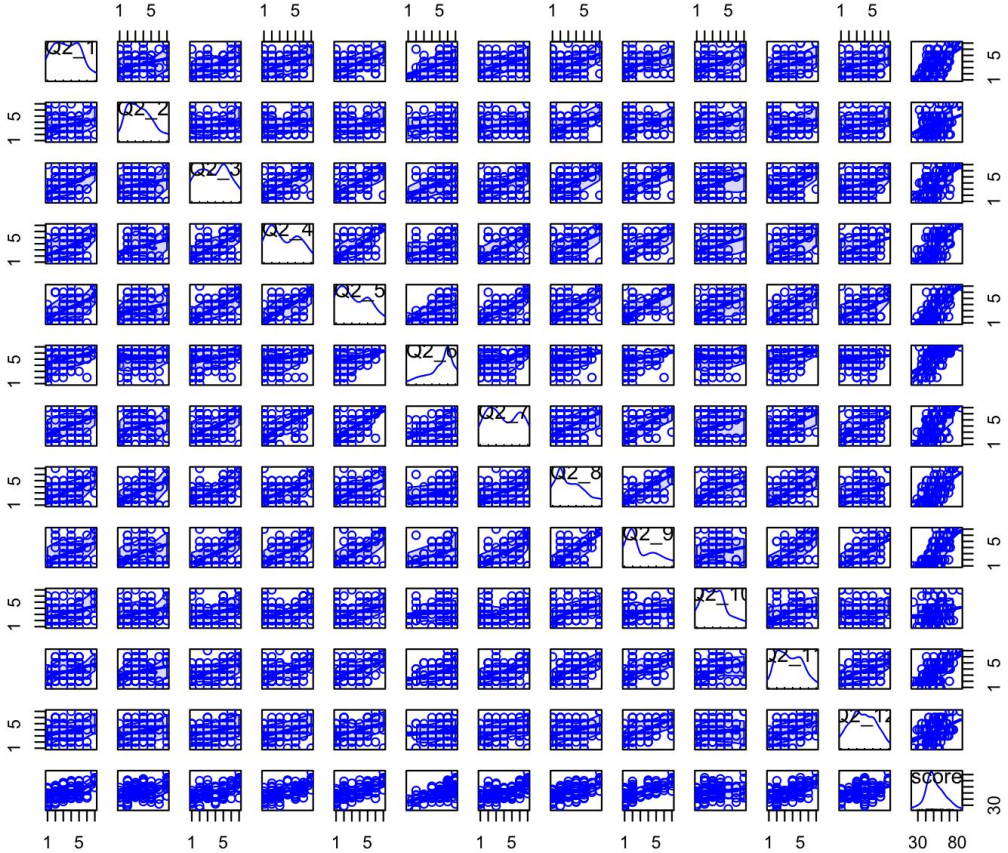
“I doubt if I can repeat my academic successes.”

Scoring Scale from the Survey

Strongly disagree	Disagree	Somewhat disagree	Neither agree or disagree	Somewhat agree	Agree	Strongly agree
1	2	3	4	5	6	7

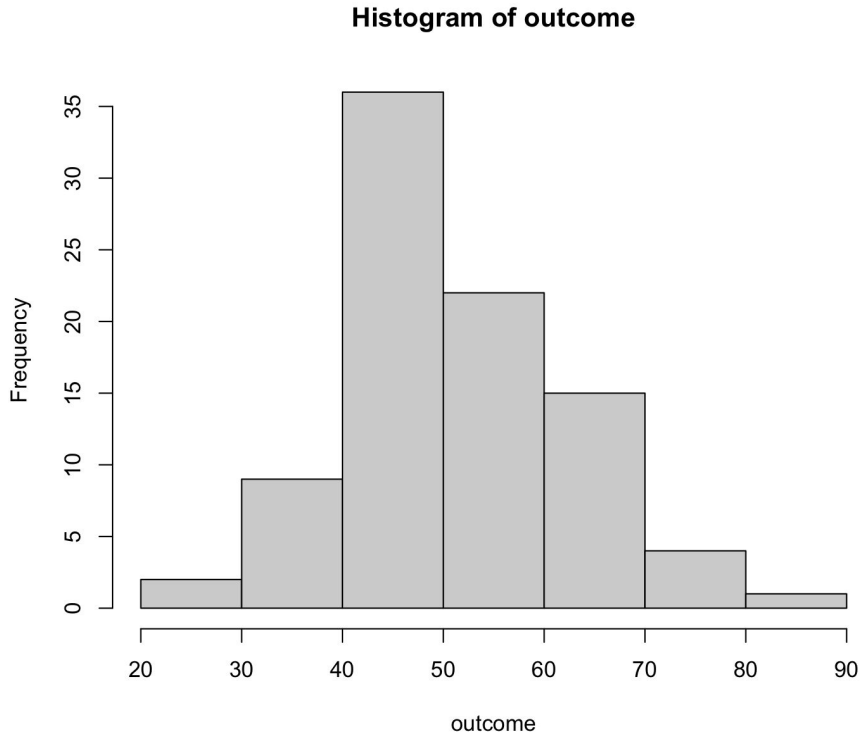
Exploratory Data Analysis

Scatterplot Matrix



- Given that our sample size is relatively small, we expected most numerical predictors to not be perfectly normally distributed.
- Most predictors appear not to deviate too much from normal distribution.
- All predictors seem to be linearly correlated with the numerical outcome variable, indicating that a linear model might fit the data well.

Numerical Imposter Syndrome Score (Outcome)



The outcome variable looks normally distributed. The conclusion is also support by the Shapiro-Wilk test for normality (p-value = 0.4021).

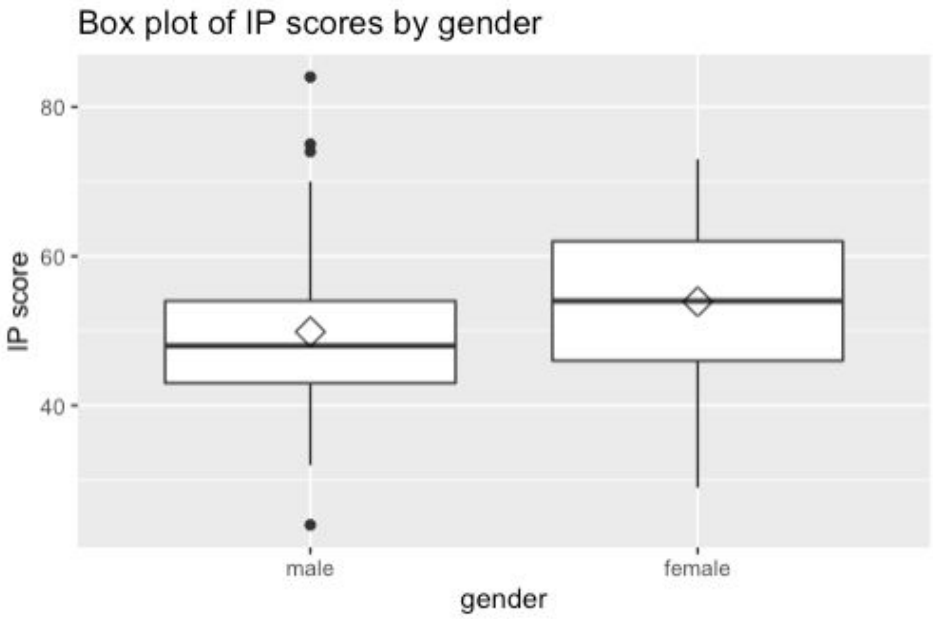
Grouping Variables: Gender

Frequency table of gender

	male	female
count	49	40

Balanced

Box plot with Group Mean



Females seem to have higher imposter syndrome scores than males, but the difference is not statistically significant **(p-value = 0.0991; SLR)**.

Grouping Variables: Sexual Orientation

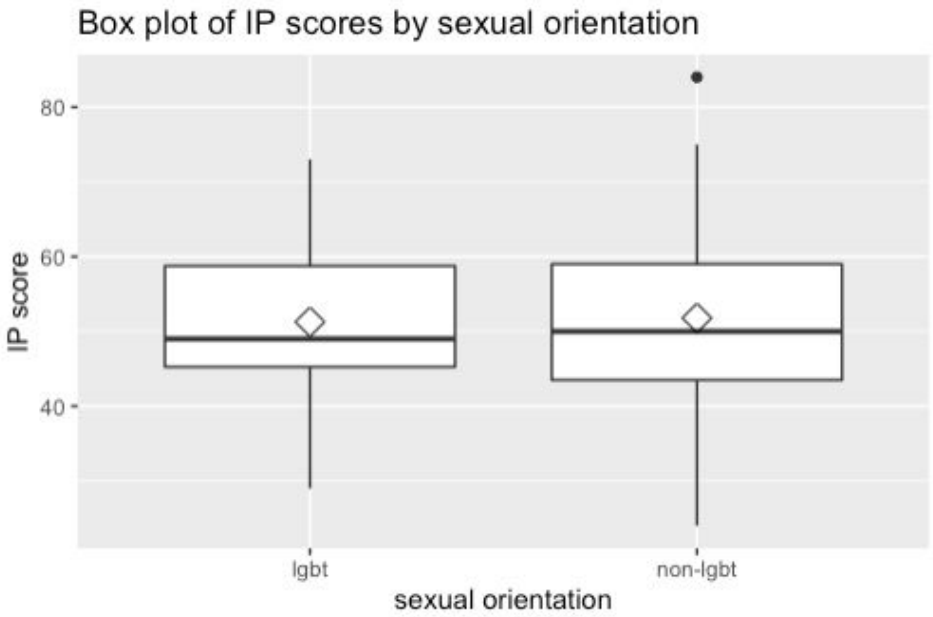
Frequency table of sexual orientation

	heterosexual	homosexual	bisexual	self-describe
count	71	6	8	4

	non-LGBT	LGBT
count	18	71

Imbalanced

Box plot with Group Mean



Imposter syndrome score does not differ by sexual orientation **(p-value = 0.867; SLR)**.

Grouping Variables: Ethnicity

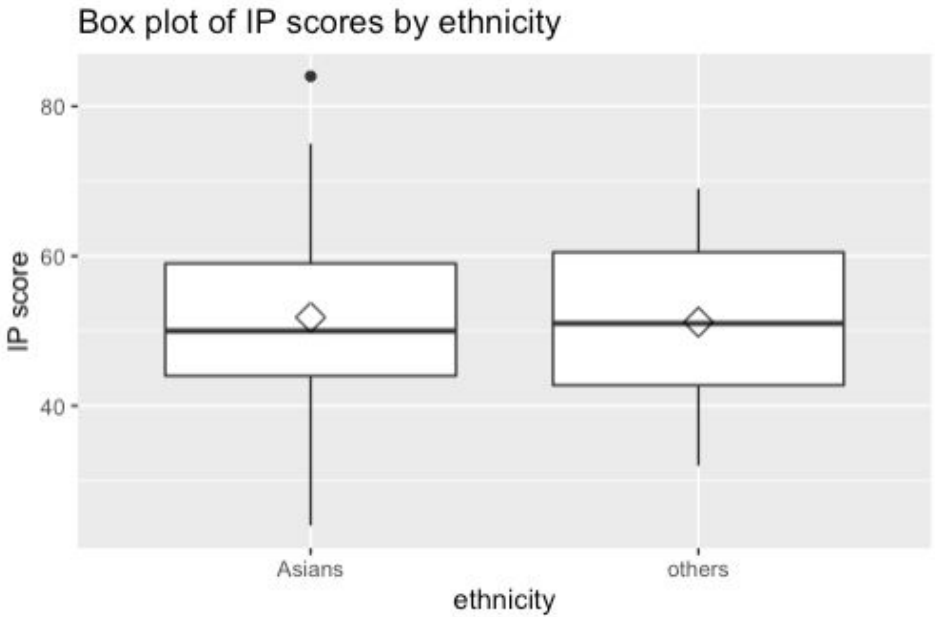
Frequency table of ethnicity

	White	Black	American Indian or Alaska Native	Asian	Native Hawaiian or Pacific Islander	Self-describe
count	8	3	2	71	0	5

	Asian	Others
count	71	18

Imbalanced

Box plot with Group Mean



Imposter syndrome score does not differ by ethnicity **(p-value = 0.831; SLR)**.

Grouping Variables: Age

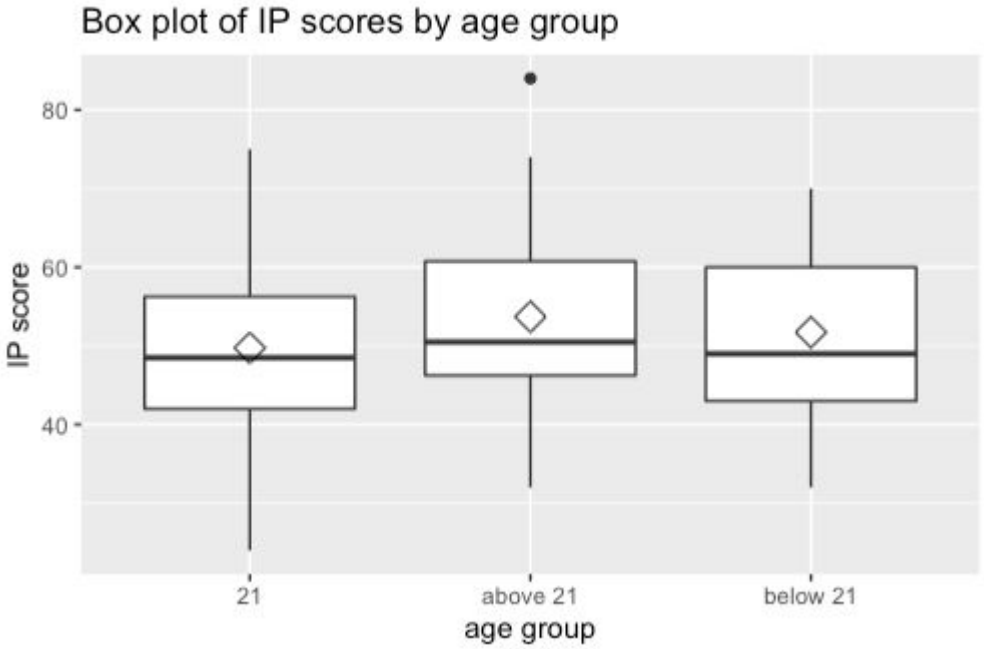
Frequency table of age

	18	19	20	21	22	23	24	25	26	28	29	31
count	3	6	10	36	17	5	4	3	2	1	1	1

	21	Below 21	Above 21
count	34	36	19

Balanced

Box plot with Group Mean



Imposter syndrome score seems to be higher for people above 21, but the difference is not statistically significant **(p-value = 0.355; one-way ANOVA)**.

Grouping Variables: Major

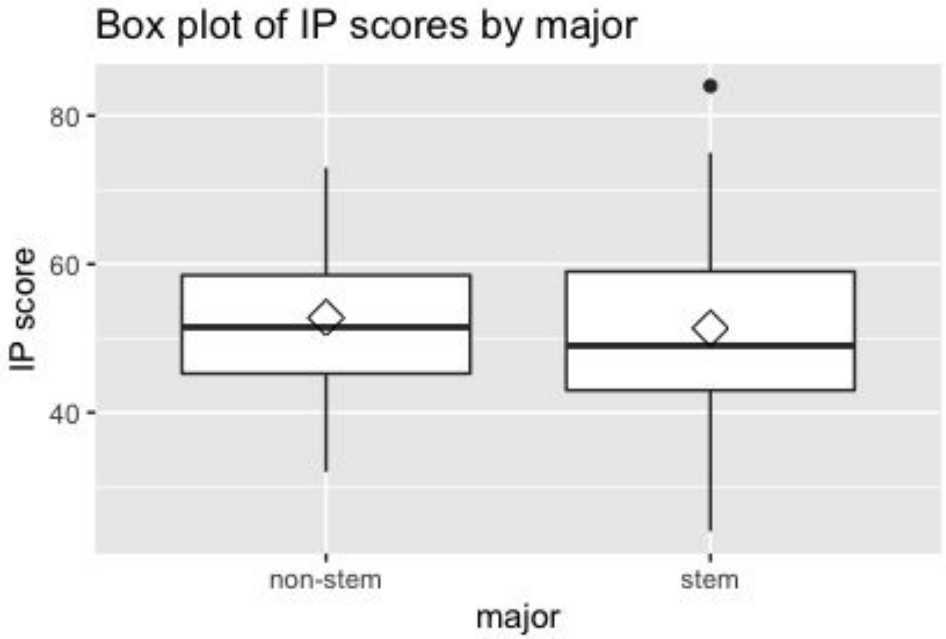
Frequency table of major

	non-STEM	STEM	both
count	12	69	10

	non-STEM	STEM
count	22	67

Somewhat imbalanced

Box plot with Group Mean



Imposter syndrome score does not differ between people majoring in non-STEM and those majoring in STEM **(p-value = 0.611; SLR)**.

Grouping Variables: GPA

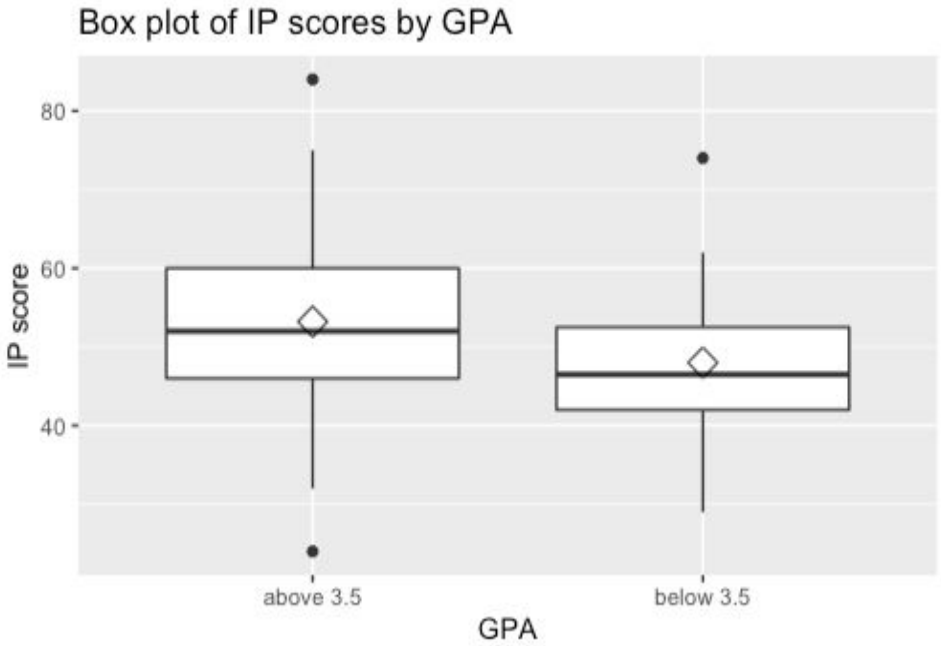
Frequency table of GPA

	Below 2.00	2.00 - 2.49	2.50 - 2.99	3.00 - 3.49	3.50 - 3.99	4.0
count	0	1	3	22	54	9

	Above 3.5	Below 3.5
count	63	26

Somewhat Imbalanced/balanced

Box plot with Group Mean



Imposter syndrome score seems to be higher for people with GPAs above 3.5, the difference is marginally significant **(p-value = 0.0503; SLR)**.

Models

Part I:
Numerical Outcome

Variable Selection

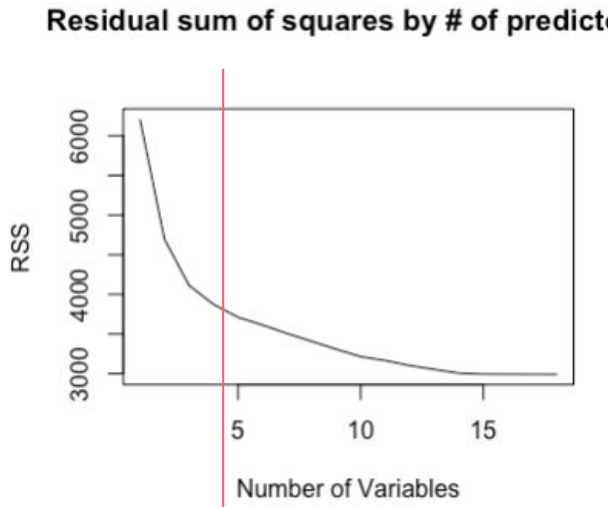
- Methods
  - Least absolute shrinkage and selection operator (LASSO)
  - Best subset selection
  - Forward stepwise selection
  - Backward stepwise selection
- Compare results across different methods and select the recurring important variables.
- Start with no interaction effects and add them in after we obtain a simplified linear model.

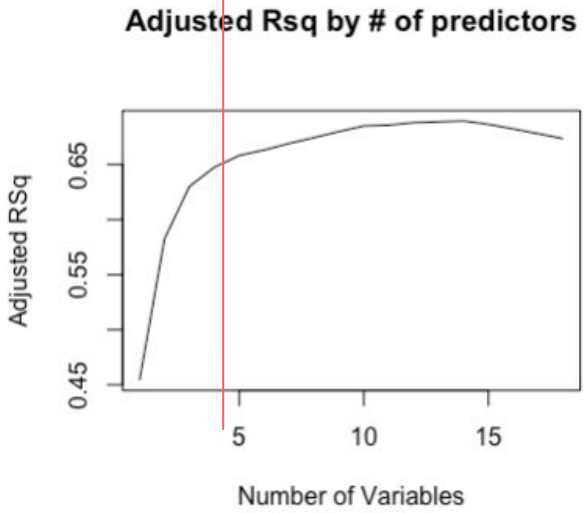
	s1
Intercept	24.093
Q2_1	0.650
Q2_2	0.453
Q2_3	.
Q2_4	0.728
Q2_5	0.821
Q2_6	1.501
Q2_7	0.488
Q2_8	.
Q2_9	1.404

	s1
Q2_10	0.650
Q2_11	0.453
Q2_12	.
Q5male	.
Q6non-lgbt	2.348
Q7others	1.833
Q8above21	1.004
Q8below21	.
Q10stem	0.237
Q11below	-3.861

Interpretation of LASSO coefficients

- Because we relied on LASSO for its feature selection function, we will focus more on the practical rather than the mathematical aspect.
- In addition, because our data set is relatively small (N = 89), we used LASSO mainly as a means to understand patterns within the data set instead of using it to construct models for future predictions.
- Coefficients are shown in the s1 column of the tables. A general rule of interpreting the coefficients is that, variables with greater absolute coefficients are judged by LASSO to be more important toward making the final prediction.
- LASSO shrinks the coefficients of unimportant variables to zero. In the tables shown, zero coefficient appears as a dot.
- Q2_4, Q2_5, Q2_6, Q2_9, sexual orientation, ethnicity, age, and GPA are the most important predictors of all.
    - Sexual orientation and ethnicity are imbalanced categorical variable, therefore we need to interpret this result with caution
    - We employed three more variable selection methods to validate the result from LASSO.





Interpretation of best subset selection

- In short, best subset selection fits all possible combination of predictors to select the best performing model
  - In our case, there would be $2^p = 262144$ models fitted.
- We can see from the plot that as the number of predictors entered in the model increases, residual sum of squares decreases. Also, as the number of predictors increases, adjusted R-squared shows an overall increasing trend as well. However, the increase stops after 14 or more predictors are entered.
- For simplicity and interpretability of our model, we conclude that the optimal number of predictors is 4.
  - This is because after 4 predictors, the decrease in RSS and the increase in adjusted R-squared slow down, indicating that the improvements become more and more trivial after this number.

Forward and Backward Stepwise Selection

Forward

	(Intercept)	Q2_4	Q2_6	Q2_9	Q11
coefficient	28.977	1.641	1.192	2.684	-3.945

Backward

	(Intercept)	Q2_4	Q2_6	Q2_9	Q11
coefficient	28.977	1.641	1.192	2.684	-3.945

Q2_4, Q2_6, Q2_9, and Q11 are consistently the most important predictors, therefore they will be entered in our linear regression model.

Model Output

	Estimate	Std. Error	t value	p value	Sig. Level
(Intercept)	28.9773	2.8022	10.341	<2e-16	***
Q2_4	1.6411	0.4567	3.593	0.000550	***
Q2_6	1.9156	0.5210	3.677	0.000415	***
Q2_9	2.6845	0.4664	5.756	1.37e-07	***
Q11below 3.5	-3.9450	1.7306	-2.280	0.025166	*

R-squared = 0.6636; adjusted R-squared = 0.6475

Model Output Interpretation

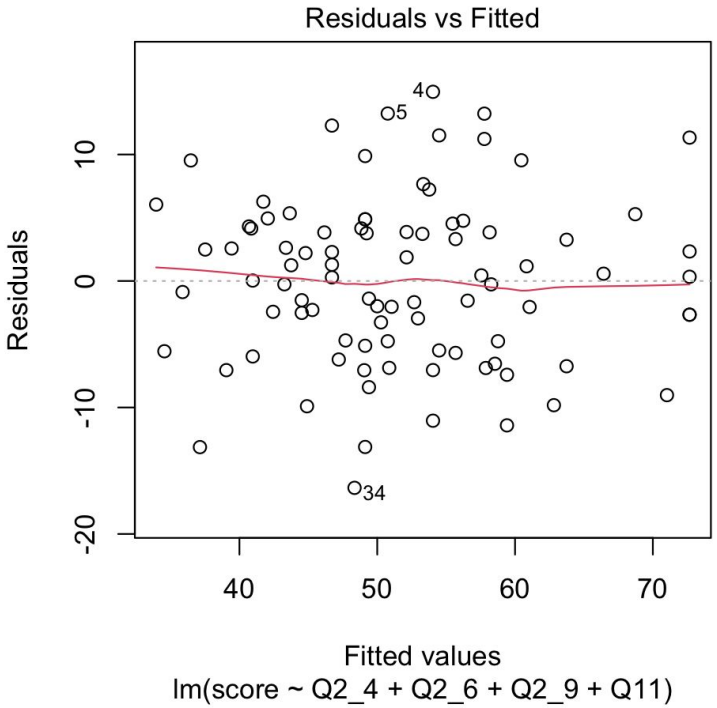
- Intercept: Imposter Syndrome Score
- Q2_4: “I doubt if I can repeat my academic successes.”
  - Feelings of fraudulence.
- Q2_6: “I tend to compare myself with others.”
  - Tendency to compare.
- Q2_9: “I do not think I belong at UCLA.”
  - Feelings of estrangement.
- Q11: GPA (baseline = above 3.5)

Multicollinearity

	VIF
Q2_4	1.529214
Q2_6	1.357551
Q2_9	1.572224
Q11	1.185658

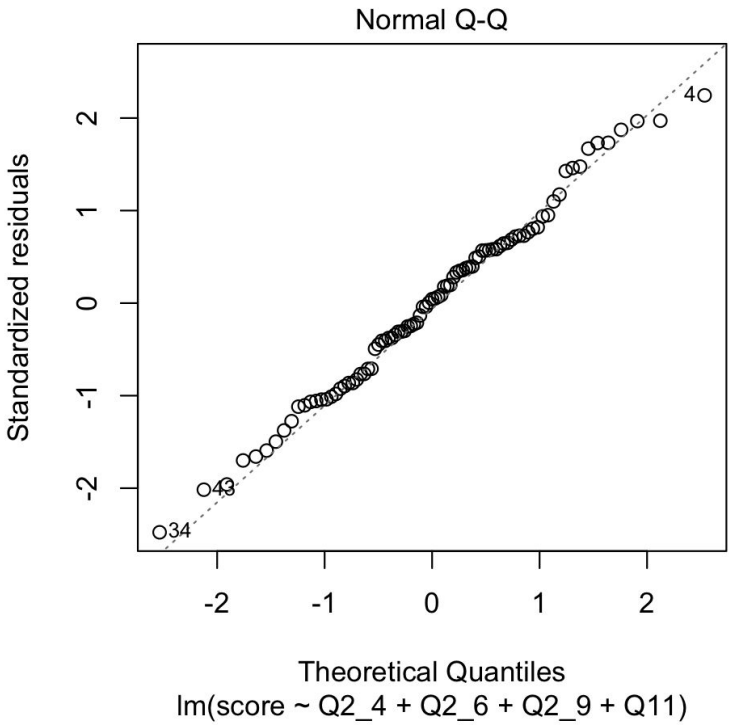
There is no issue with multicollinearity. The VIF value associated with each predictor is low, which indicates they are not correlated with one another.

Equality of Error Variance



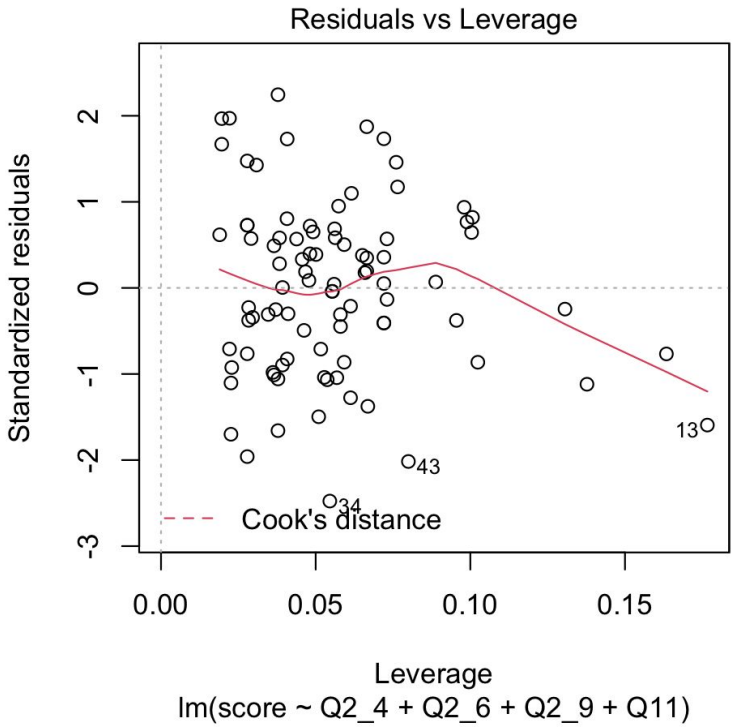
There is no pattern in the plot, which indicates that there is no correlation between the residuals and the predicted values.

Normality



Most observations fall in the the line, however, there are some misleading observations at the edges.

Outliers and Influential Points

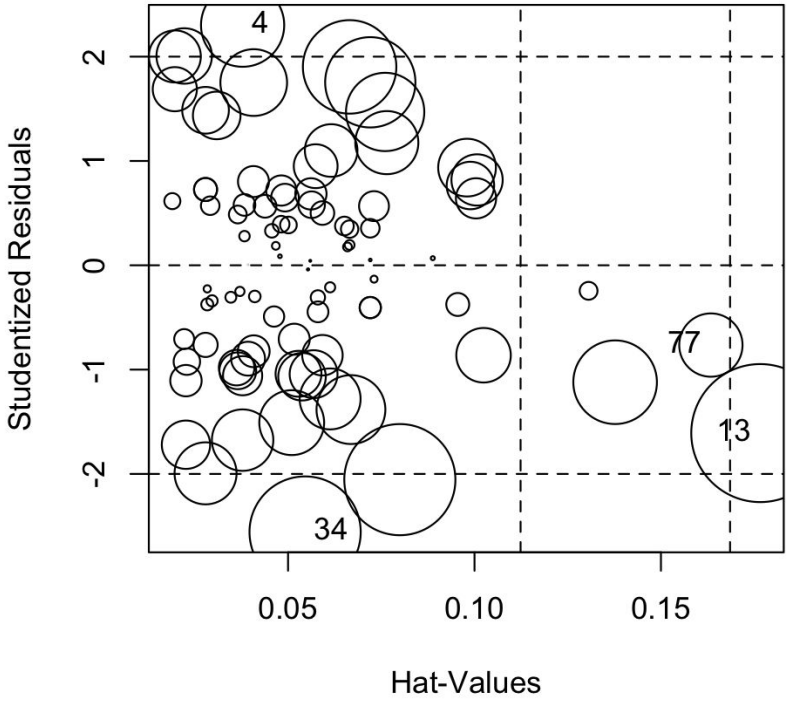


All points seem to be within the Cook's distance.

A small data set has a (-2,2) range for standardized residuals.

There are several outliers in the data.

Outliers and Influential Points Continued



	StudRes	Hat	Type
4	2.3018084	0.03780997	**No Leverage**
13	-1.6084707	0.17662806	Good leverage
34	-2.5570384	0.5456995	**Bad leverage**
77	-0.767405	0.16339302	Good leverage

Leverage cutoff: 4 / 89 = 0.04494382

We decided to remove observation #34 from the data.

Model with Interaction

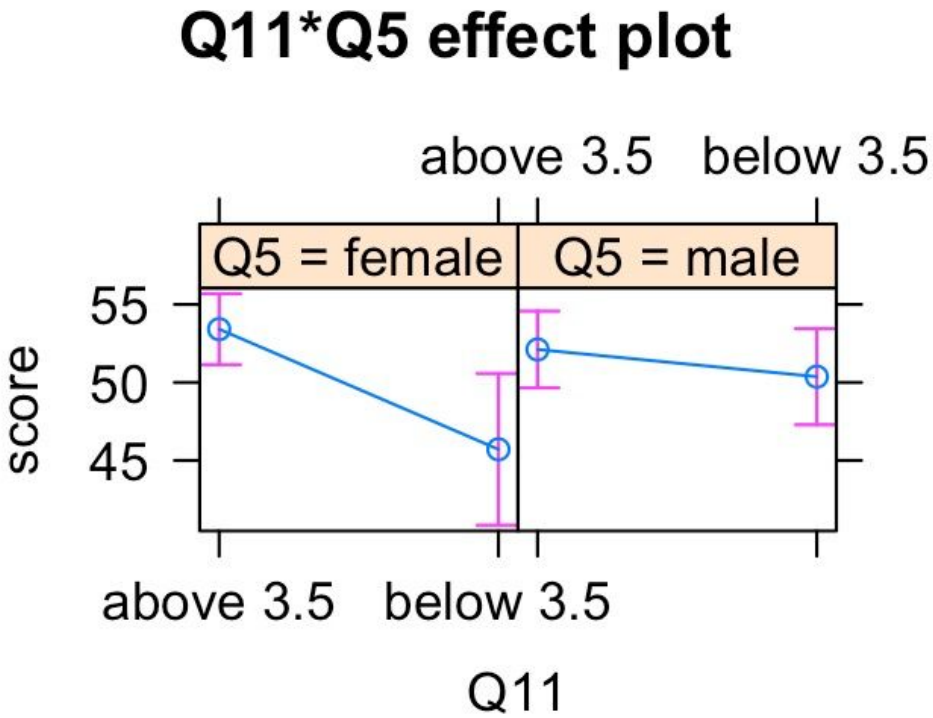
- Imposter syndrome disproportionately affect high achieving women (Clance & Imes, 1978).
- If we use GPA to define achievement, could there be a interaction between gender and GPA?

Model Output

	Estimate	Std. Error	t value	p value	Sig. Level
(Intercept)	28.9191	2.8364	10.196	4.02e-16	***
Q2_4	1.4127	0.4454	3.171	0.00215	**
Q2_6	2.1539	0.4967	4.336	4.19e-05	***
Q2_9	2.7455	0.4454	6.164	2.71e-08	***
Q11below 3.5	-7.7027	2.7065	-2.846	0.00562	**
Q5male	-1.3018	1.6853	-0.772	0.44212	
Q11below 3.5:Q5male	5.9582	3.2754	1.819	0.07264	.

R-squared = 0.7011; adjusted R-squared = 0.6787

Interpretation of Interaction



We can observe an increase in the mean imposter syndrome scores as we switch from below a 3.5 GPA to above 3.5 GPA. we can see that

On average female students tend to have higher imposter syndrome scores compared to males as GPA increases.

However, the interaction effect between GPA and gender is only marginally significant (**p-value = 0.073**), so we decided not to include the interaction term.

Model Output Interpretation

- Intercept: Imposter Syndrome Scores
- Q2_4: “I doubt if I can repeat my academic successes.”
  - Feelings of fraudulence.
- Q2_6: “I tend to compare myself with others.”
  - Tendency to compare.
- Q2_9: “I do not think I belong at UCLA.”
  - Feelings of estrangement.
- Q11: GPA (baseline = above 3.5)

Model Output

	Estimate	Std. Error	t value	p value	Sig. Level
(Intercept)	28.7288	2.6484	10.848	<2e-16	***
Q2_4	1.4507	0.4353	3.332	0.00129	**
Q2_6	2.0855	0.4990	4.179	7.28e-05	***
Q2_9	2.6895	0.4466	6.023	4.65e-08	***
Q11below 3.5	-3.7649	1.6351	-2.303	0.02384	*

R-squared = 0.6886; adjusted R-squared = 0.6735. Both measures improved after we removed the outlier and the bad leverage observations (R-squared = 0.6636; adjusted R-squared = 0.6475 from the previous model).

Model Output Interpretation

Keeping everything else constant:

- Q2_4: for every one unit increase in ratings of feelings of fraudulent, on average the imposter syndrome increases by 1.45 units.
- Q2_6: for every one unit increase in tendency to compare, on average imposter syndrome increases by 2.09 units.
- Q2_9: for every one unit increase in feelings of estrangement, on average imposter syndrome increases by 2.69 units.
- Q11: switching from the below 3.5 GPA to the above 3.5 GPA, on average imposter syndrome is going to increase by 3.76 units.
- Adjusted R-squared: After adjusting for the number of predictors in the model, we were able to explain 67.35% of the total variation in imposter syndrome scores.

Research Questions

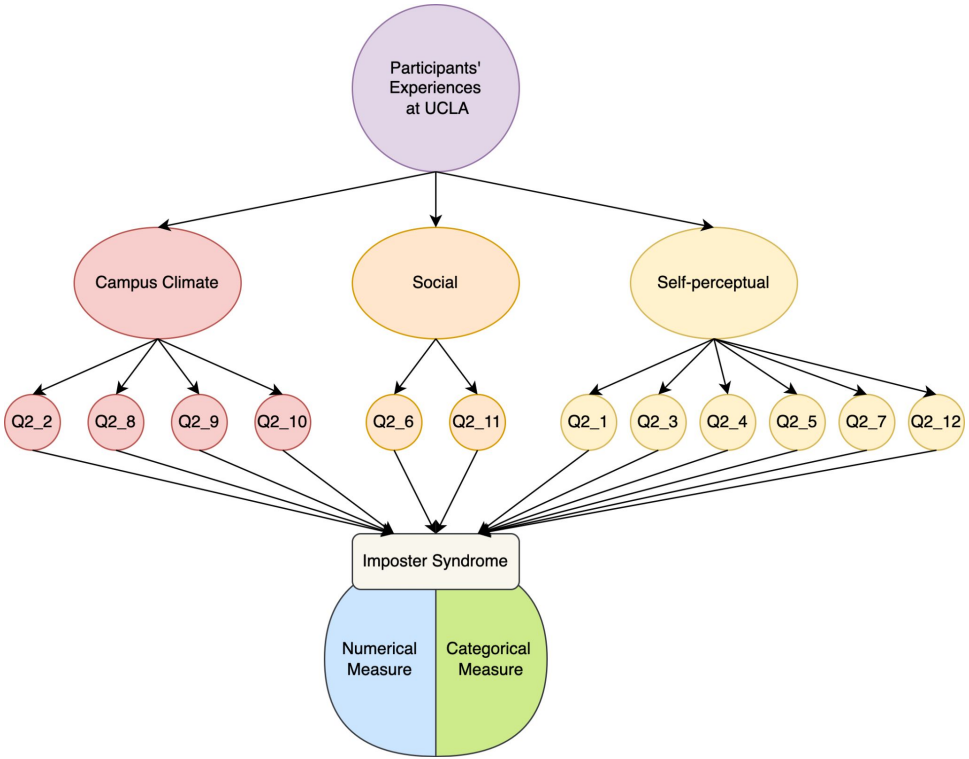
Can we predict a person's imposter syndrome score from existing scales using a newly developed scale that contains statements pertaining specifically to their experiences at UCLA?

Answer: Yes! Using feelings of academic fraudulence, tendency to compare, feelings of estrangement, and GPA, we are able to explain 67.35% of the total variation in imposter syndrome scores.

Part II:
Categorical Outcome

Rationale for Part II

- In additional to the multiple linear regression model we had in part I, we decided to also include a multiple logistic regression model.
- MLR provides a unit-based explanation of relationships between predictors and outcome variable.
- Since we also collected a categorical measure of imposter syndrome, a more intuitive approach is to use a multiple logistic regression to look at the odds of a person selecting a certain answer given their prior responses on other measures.
- To further simplify the model, we could also convert our predictors from numerical to binary.



Exploratory Data Analysis

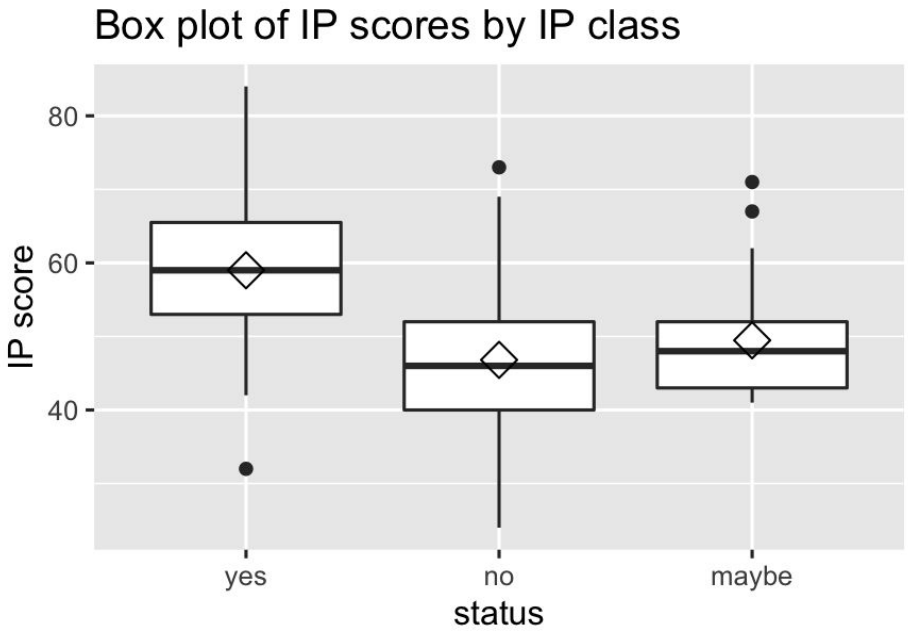
Categorical Outcome Variable

Frequency table of response

	yes	no	maybe
count	31	37	21

The frequencies are balanced, but to simplify the model and increase interpretability, we decided to convert the variable to binary (yes vs. no)

Box Plot with Group Mean



	diff	p adjusted
no - yes	-12.18919	0.0000118
maybe - yes	-9.52381	0.0037748
maybe - no	2.66538	0.6037405

- At least one pair of means contains a significant difference (**p-value = 1.37e-05; one-way ANOVA**).
- Tukey HSD results show that mean imposter syndrome score for people impacted by imposter syndrome is significantly higher than that of people who are not impacted by imposter syndrome or people who do not know whether they are impacted or not.

Categorical Outcome Variable

- On one hand, it makes sense to group “no” and “maybe” together, since the two groups have similar mean imposter syndrome scores, so we have partial evidence to support that these two groups responded to survey questions in a similar fashion.
- On the other hand, there may be systematic differences between people who chose the definite answer “no” and those who were not sure about whether they had imposter syndrome or not.
- We decided to drop all “maybe” responses from the data to ensure that people from one class of the binary outcome variable are distinctively different from those from the other class.

Categorical Outcome Variable

Frequency table of response

	yes	no	maybe
count	31	37	21

	yes	no
count	31	37

Balanced

- After we dropped all “maybe” responses, the new data set contains 68 samples in total.
- The outcome variable becomes a binary one with two balanced classes “yes” and “no”.

Binarize Predictors

	1	2	3	4	5	6	7
Q2_1	13	10	12	9	15	3	6
Q2_2	6	15	13	13	10	4	7
Q2_3	10	10	12	7	14	7	8
Q2_4	12	12	10	5	12	10	7
Q2_5	17	12	9	5	14	5	6
Q2_6	2	7	5	6	10	24	14
Q2_7	10	12	10	6	7	11	12
Q2_8	15	16	9	12	7	3	6
Q2_9	18	23	2	8	8	2	7
Q2_10	11	16	14	11	7	5	4
Q2_11	6	18	11	11	12	6	4
Q2_12	8	8	15	15	10	8	4

If rating < 4, convert to low.

If rating >= 4, convert to high.

	low	high
Q2_1	35	33
Q2_2	34	34
Q2_3	32	36
Q2_4	34	34
Q2_5	38	30
Q2_6	14	54
Q2_7	32	36
Q2_8	40	28
Q2_9	43	25
Q2_10	41	27
Q2_11	35	33
Q2_12	31	37

Mostly balanced

Variable Selection

- LASSO
- Best subset selection
- Model selection criteria
  - AKaike's Information Criteria (AIC)
  - Bayesian Information Criteria (BIC)
- Given our data set is shrunk to 68 observations, we decided to remove all imbalanced grouping variables from the data set before performing any variable selection.

	s1
Intercept	1.054
Q2_1low	.
Q2_2low	0.214
Q2_3low	-0.061
Q2_4low	-0.921
Q2_5low	-0.121
Q2_6low	-1.187
Q2_7low	-1.176

	s1
Q2_8low	.
Q2_9low	.
Q2_10low	.
Q2_11low	.
Q2_12low	.
Q2_5male	-0.121
Q2_10stem	.
Q11below 3.5	.

Interpretation of LASSO coefficients

- We removed all imbalance grouping variables.
  - The only ones being retained are gender, major, and GPA.
  - Major is not technically a balanced grouping variable, so the inclusion of major is an arbitrary choice. Variable selection should be able to handle the question of whether we should include it or not.
  - Age was removed because other methods for variable selection used for Part II would output results with low interpretability. And it was previously shown not to be an important predictor.
- Q2_2, Q2_3, Q2_4, Q2_5, Q2_6, Q2_7, and gender are the most important predictors selected by LASSO.
  - However, the coefficients of Q2_3, Q2_5, and gender are relatively low compared to those of other important predictors.
  - We decided to drop Q2_3 and Q2_5 because of their low importance.
  - We retained gender for its potential interaction with age.

Variable Selection Continued

AIC selection

	Estimate	Std. Error	z value	p value
(Intercept)	1.392558	0.5562825	2.503330	0.012303
Q2_2low	2.555847	1.2256978	2.085218	0.037050
Q2_4low	-1.751027	0.8294555	-2.111056	0.034767
Q2_6low	-2.607491	1.2197435	-2.137737	0.032538
Q2_7low	-1.552439	0.7925527	-1.958784	0.050138
Q2_8low	-1.723294	1.1995878	-1.436572	0.150840

Only Q2_2, Q2_4 , Q2_6, Q2_7, and Q2_8 are selected, but Q2_8 is not significant.

Variable Selection Continued

BIC selection

	Estimate	Std. Error	z value	p value
(Intercept)	1.263595	0.4281335	2.951405	0.003163
Q2_4low	-2.876361	1.1374537	-2.528772	0.011446
Q2_6low	-2.343506	0.6288221	-3.726820	0.000193

Only Q2_4 and Q2_6 are selected, and both are significant.

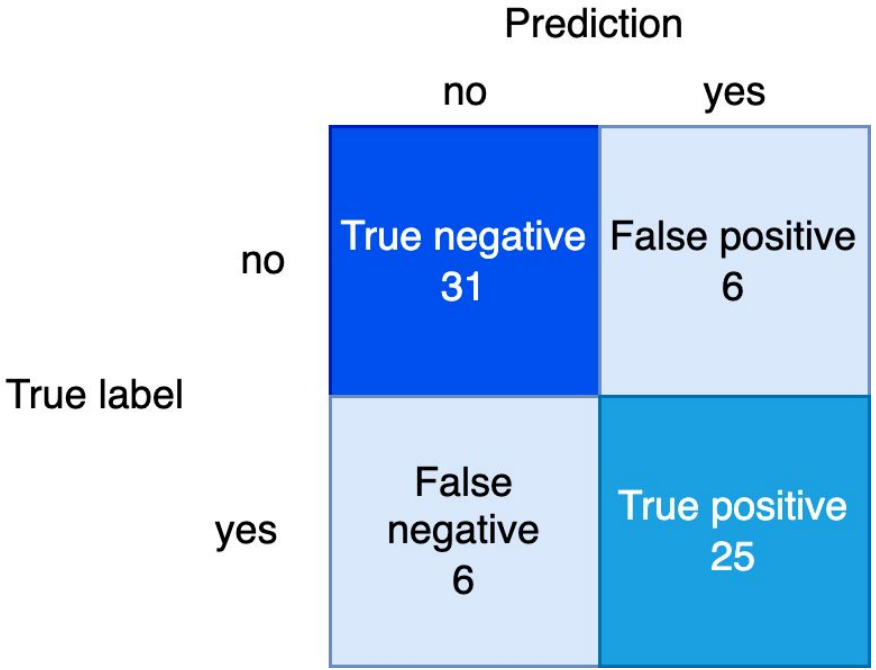
Model Output

	Estimate	Std. Error	z value	p value	Sig. Level
(Intercept)	1.1947	0.5145	2.322	0.02024	*
Q2_2low	1.4448	0.7953	1.817	0.06927	.
Q2_4low	-1.8132	0.8063	-2.249	0.02453	*
Q2_6low	-2.7583	1.2478	-2.211	0.02707	*
Q2_7low	-1.9234	0.7369	-2.610	0.009505	**

Model Output Interpretation

- Q2_2: “ I feel like people like myself are underrepresented at UCLA”
  - Lack of diversity.
  - Not statistically significant, but close to the significance level.
- Q2_4: “I doubt if I can repeat my academic successes.”
  - Feelings of academic fraudulence.
- Q2_6: “I tend to compare myself with others.”
  - Tendency to compare.
- Q2_7: “I think most people are more academically competent than myself.”
  - Feelings of academic inferiority.

Model Accuracy



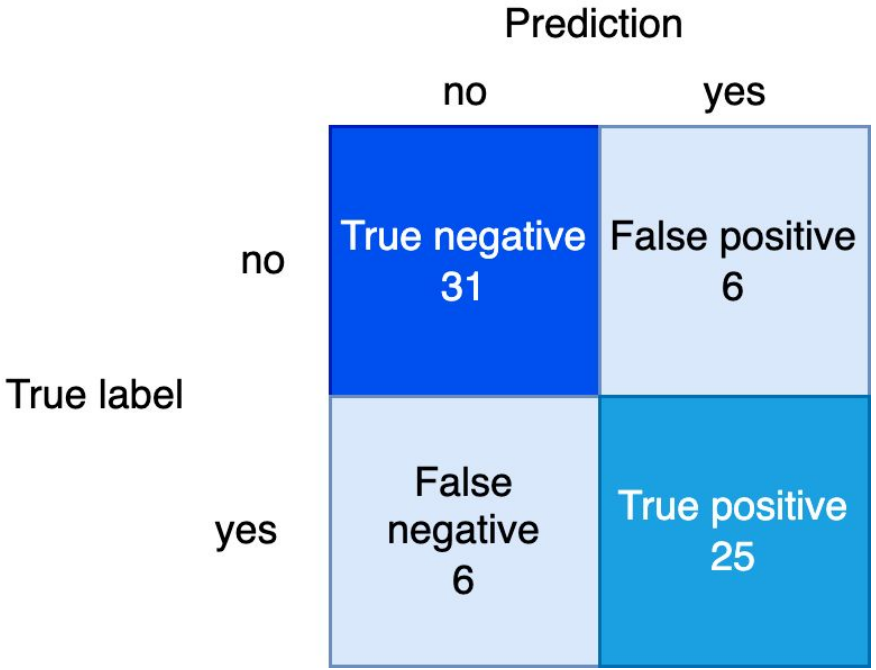
Error rate = (6 + 6) / (31 + 25 + 6 + 6) = 0.1764706

Model accuracy = 1 - error rate = 0.8235294

Model Output with Interaction

	Estimate	Std. Error	z value	p value	Sig. Level
(Intercept)	1.62707	0.69375	2.345	0.01901	*
Q2_2low	1.51709	0.84666	1.792	0.07316	.
Q2_4low	-1.80376	0.83456	-2.161	0.03067	*
Q2_6low	-2.80850	1.23344	-2.277	0.02779	*
Q2_7low	-2.00401	0.77117	-2.599	0.00936	**
Q5male	-0.89261	0.76231	-1.171	0.24163	
Q11below 3.5	0.07276	1.44944	0.050	0.95996	
Q5male:Q11 below 3.5	0.20236	1.86577	0.108	0.91363	

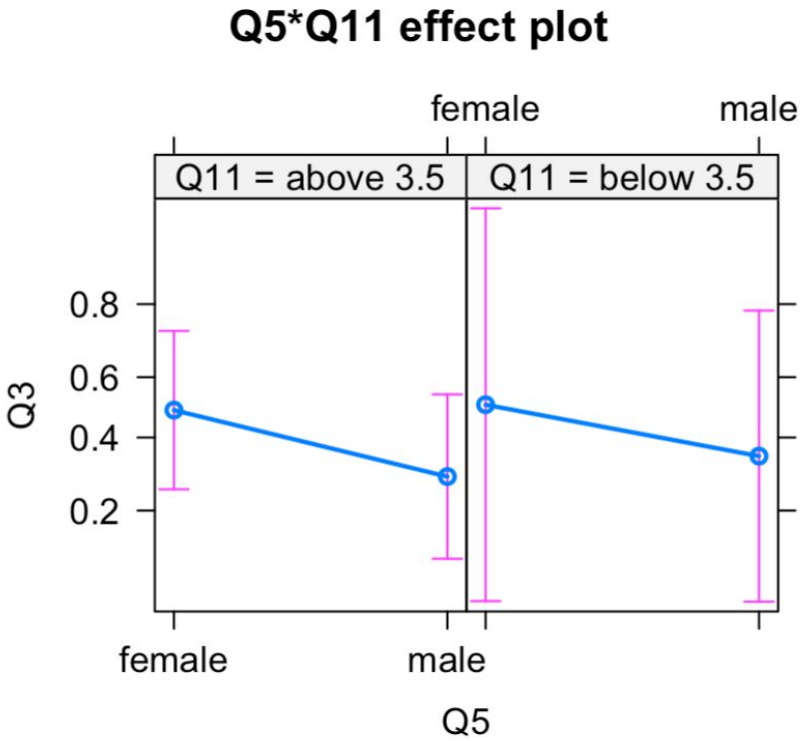
Model Accuracy with Interaction



Error rate = (6 + 6) / (31 + 25 + 6 + 6) = 0.1764706

Model accuracy = 1 - error rate = 0.8235294

Model Accuracy with Interaction



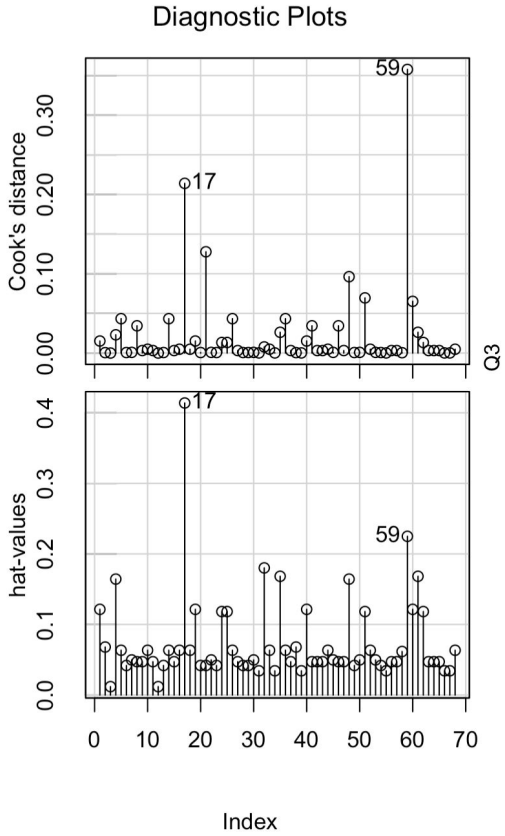
The interaction term is not significant, and the two parallel lines in the interaction plot also indicates no interaction. Since adding the interaction effect to the model has no influence on the model's accuracy, we should probably go with the simpler model without interaction term.

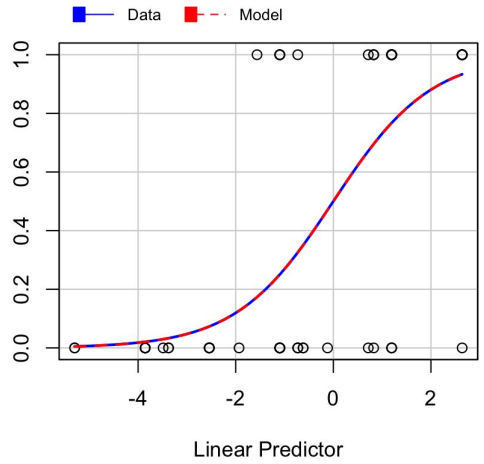
Model Output

	Estimate	2.5% CI	97.5% CI	Std. Error	z value	p value	Sig. Level
(Intercept)	1.1947	0.2410	2.2911	0.5145	2.322	0.02024	*
Q2_2low	1.4448	0.0060	3.2041	0.7953	1.817	0.06927	.
Q2_4low	-1.8132	-3.5543	-0.2948	0.8063	-2.249	0.02453	*
Q2_6low	-2.7583	-5.9241	-0.6763	1.2478	-2.211	0.02707	*
Q2_7low	-1.9234	-3.4700	-0.5248	0.7369	-2.610	0.009505	**

Classification accuracy = 0.8235294

Influential Plots and Marginal Model Plot





- The influential plots show that observations 17 and 59 are high leverage points.
- However, removing these two observations changed the standard error of the coefficient of Q2_6 in the multiple logistic regression model from 1.2478 to 1730.6908, which by no means is a normal pattern.
- Without removing the two leverage points, the marginal model plot shows that our multiple logistic regression model seems to capture the patterns in the data set.
- In the end, the observations 17 and 59 were not removed from the model.

Model Output Exponentiated

	Estimate	2.5% CI	97.5% CI	p value	Sig. Level
(Intercept)	3.3026	1.2726	9.8860	0.02024	*
Q2_2low	4.2410	1.0061	24.6339	0.06927	.
Q2_4low	0.1631	0.0286	0.7447	0.02453	*
Q2_6low	0.0634	0.0027	0.5085	0.02707	*
Q2_7low	0.1461	0.0311	0.5917	0.009505	**

Model Output Interpretation

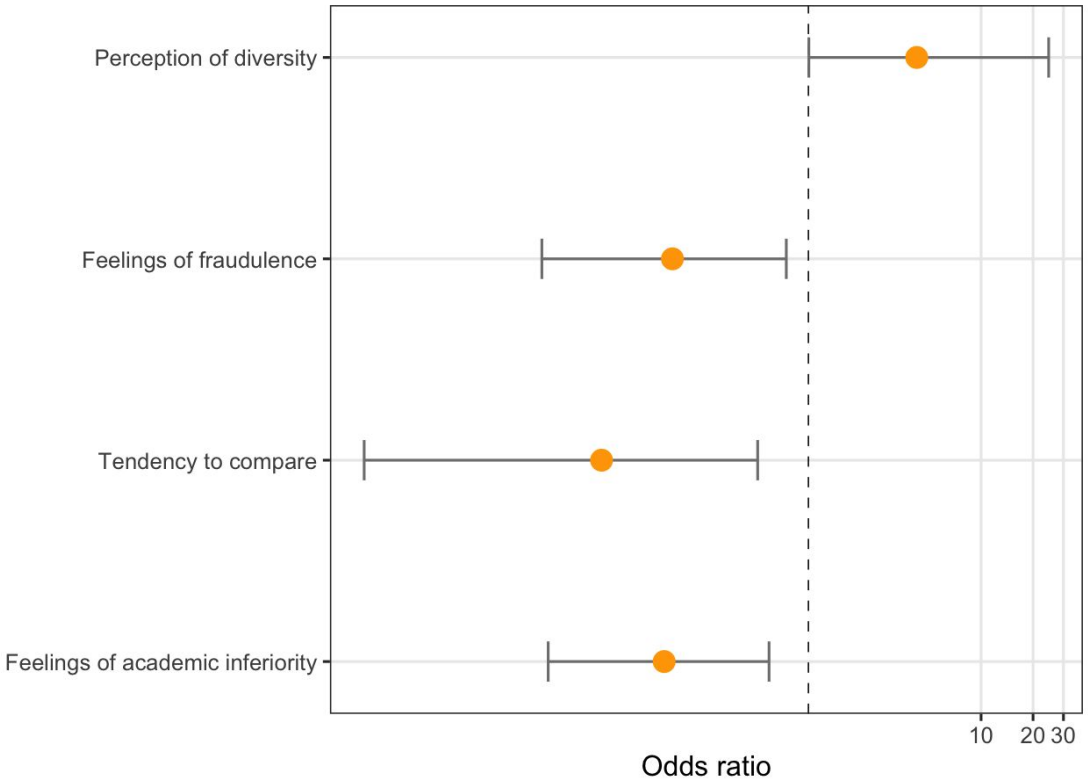
- Q2_2: “ I feel like people like myself are underrepresented at UCLA”
  - Perception of diversity.
  - Not statistically significant, but close to the significance level.
- Q2_4: “I doubt if I can repeat my academic successes.”
  - Feelings of academic fraudulence.
- Q2_6: “I tend to compare myself with others.”
  - Tendency to compare.
- Q2_7: “I think most people are more academically competent than myself.”
  - Feelings of academic inferiority.

Model Output Interpretation

Keeping everything else constant:

- The odds of having imposter syndrome is 83.69% less for people with low perception of academic fraudulence compared to people with high perception of academic fraudulence.
- The odds of having imposter syndrome is 93.66% less for people with low tendency to compare themselves with other compared to people with high tendency to compare.
- The odds of having imposter syndrome is 85.39% less for people with low perception of academic inferiority compared to people with high perception of academic inferiority .

Visualization of the Model Output



The dotted line indicates a odds ratio of 1, or even odds. All of the significant predictors have confidence intervals that do not capture 0. Because these intervals are on the left of the even odds line, they show that the odds of having imposter syndrome will be less for those who have low perceptions of these measures.

Research Questions

Can we predict the odds of a person being impacted by imposter syndrome using statements from the newly developed scale that contains statements pertaining specifically to their experiences at UCLA?

Answer: Yes! Using feelings of fraudulence, tendency to compare, and feelings of academic inferiority, we are able to make correct predictions about whether a person has imposter syndrome or not 82.35% of the time.

Part III:
Text Analysis

Method

We took the text data from Q4, which asked respondents to describe a time when they were impacted by imposter syndrome, and to identify two potential reasons that made them feel like an imposter in the situation they just described.

We grouped the data based on whether the respondent answered “yes” “no” and “maybe” to Q3, which asked if they think they were impacted by impostor syndrome. We then generated a word cloud for each.

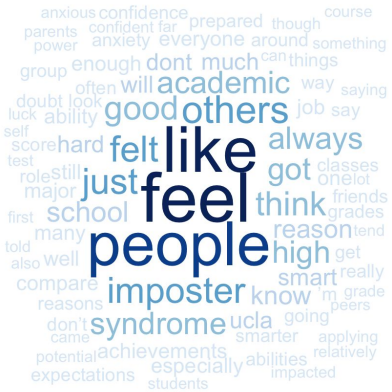
Text Clouds by Imposter Syndrome Status

YES

NO

MAYBE

Text Clouds by Imposter Syndrome Status



YES

NO

MAYBE

All 3 word clouds demonstrate roughly the same words as the most frequent. However, we noticed that the word “others” was included in the “yes” cloud but not in any of the other word clouds. This finding suggests that among those who believe they have been impacted by impostor syndrome, thoughts regarding other people may be more frequent among those who believe they have been impacted by imposter syndrome than in other groups. This is consistent with our finding that comparing oneself to others is one of the most important predictors of impostor syndrome.

Summary of Findings from Word Clouds

The word clouds shares the following similarities:

- The most frequent words are the same: “feel”, “like”:
  - The responses focuses on respondent’s emotions and perceptions.
- All mentioned “UCLA”:
  - No matter how the respondents feel about whether they are affected by imposter syndrome, they all use UCLA as a criterion.

And they are different in terms of:

- The word cloud for people who identify as impacted by imposter syndrome contains relatively-high frequency words like “compare”, “around”, and “achievements” that the other two graphs does not:
  - People who are constantly comparing themselves with other people near them will tend to suffer more from imposter syndrome.
- In the word cloud for “yes”, there are also words like “anxiety”, “confidence”, “role”, and “doubt” that do not appear on the other two graphs:
  - We interpret this as accounts of how students are likely to feel when experiencing imposter syndrome.

Research Questions

Does the design of our project have high ecological validity? More specifically, do the results from our study match with what people think of imposter syndrome in reality?

Answer: Yes! We found people who thought they had imposter syndrome mentioned comparisons and thoughts regarding other people more often than those from the other two groups. This result matches our statistical analysis that comparison related predictors tend to be more important.

Summary

- People seem to have a good judgment about whether they have imposter syndrome or not.
  - People reported to have imposter syndrome scored significantly higher than people who didn't.
- Using feelings of academic fraudulence, tendency to compare, feelings of estrangement, and GPA as our numerical predictors, we are able to explain 67.35% of the total variation in imposter syndrome scores.
- Using feelings of fraudulence, tendency to compare, and feelings of academic inferiority as our categorical predictors, we are able to make correct predictions about whether a person has imposter syndrome or not 82.35% of the time.
- The interaction between gender and GPA is marginally significant. More specifically women with high achievements tend to have higher imposter syndrome scores.

Zero-In

- Feelings of estrangement or lack of sense of belonging contributes to imposter syndrome.
  - Educate faculty and students about the significance of diversity and inclusion and create diverse and inclusive environments to cultivate sense of belonging.
  - Introduce students to support services offered on campus and make these services accessible.
  - Foster a sense of community in classes through promoting student-student and student-faculty interactions.

Zero-In

- The tendency to compare oneself with others and the resulting perception of academic inferiority are the key components of imposter syndrome.
  - Shift the focus of comparison to personal growth. Grading systems can include a “personal improvement”component that is specific to each individual. For instance, writing courses take into account each student’s starting point and mainly use personal progress to evaluate performance.
  - Using absolute grading scales instead of the curving method to avoid direct ranking of students.
  - Promote collaborative interaction or group work between students to loosen the competitive atmosphere.
  - Students can learn to look at comparisons from a different len. For instance, instead of using comparisons as a means to degrade our achievements, focus on how we can improve and learn from people who we think are better than us.

Zero-In

- Among different measures of feelings of fraudulence, the most important element is whether one can repeat their academic successes.
  - Assessment standards need to yield consistent results of one's academic abilities and minimize the role of luck in scoring. For example, quizzes and exams should maintain similar levels of difficulty and contain similar type of questions.
  - Offer more opportunities for feedback and assessment might be helpful. For instance, the midterm-final system can be broken down into more frequent and less stressful assessments. On one hand, students will have more opportunities to repeat their academic successes. On the other hand, students can keep track of their personal progress and effort and avoid arbitrary attributing successes to luck.

Limitations

- The study overall suffered from small sample size.
- The data came mostly from Asian and STEM-majoring participants, which limits the study's generalizability.
- Models were constructed exclusively to understand the data in-hand, and prediction accuracy was not the focus of the study. Therefore, the results may have high variability.
- Overemphasizing interpretability may lead to bias in variable selection.
- Voluntary response allowed participants to skip questions, which resulted in loss of observations that may possess specific characteristics.