Introduction

Every four years, the Hennepin County public health department conducts a survey about its residents mental and physical health along with their social determinants of health. This analysis uses a cleaned version of the county's 2022 survey; focusing on if two distinct characteristics (insecurity score and discrimination in healthcare) are associated with or predictors of a resident's number of bad mental health days. This dataset is from a cross-sectional study that includes 4539 residents with responses in 8 categories. The categories of interest in this study include the number of bad mental health days (MHdays), a yes/no of if the respondent had been discriminated against when seeking medical care within the past year (Discrim2), and a numerical measure of overall insecurity which was compiled from a number of questions (Insecurity2).

Demographics and Characteristics

A previous analysis explored if demographic factors such as gender identity, transgender status, or race were considered confounding factors for the analysis at hand. Confounding factors for insecurity score include a respondent's gender, trans status, and race/ethnicity (table 1). Race/ethnicity was found to be a confounding factor for discrimination in healthcare (table 2). Confounding factors for the number of not good mental health days include a respondent's gender, trans status, and race/ethnicity (table 3). These are all denoted as being confounding factors as an ANOVA analysis found there to be significant differences in response of one or more demographic groups.

Model 1

The first part of this analysis is focused on the relationship between a respondent's average number of not good mental health days and whether or not they faced discrimination in health care. Looking at figure A, it is clear that those who faced medical discrimination had, on average, a greater number of not good mental health days than those that did not. This is supported by comparing the means, as those who did not face medical discrimination had an average of 7.8 not good mental health days whereas those who did had an average of 13.8 (table 4). There is also a lot more variability in the number of not good mental health days for those who experienced medical discrimination with a standard deviation of 10.14 compared to a standard deviation of 7.59 for those who did not (table 4). In order to determine if this difference is statistically significant, a two sample t-test was performed. The three assumptions need to be met in order to perform this test are independent sampling, normal distribution, and equal variance. Independent sampling is met by study design, normal distribution is met as the sample sizes of each group are large enough for the central limit theorem to hold, and equal variance is met as twice the smallest standard deviation is larger than the largest standard deviation. The null hypothesis for this test was that there was no difference between the mean number of not good mental health days of those who faced discrimination in a medical setting and those that did not. The alternative hypothesis states that there is a difference in the mean number of mental health days faced by those who faced medical discrimination and those that did not. These could be otherwise written as H_0: $\mu_1 = \mu_2$ and H_α: $\mu_1 \neq \mu_2$ where μ_1 is the mean of those who faced medical discrimination and μ_2 is the mean of those who did not. The test returned a p-value of $<2.20\times10^{-16}$, which is statistically significant. This means that there is a significant difference between the number of not good mental health days for those who faced discrimination in a medical setting and those that did not. The 95% confidence interval for this output is [8.33, 8.82]; in context, this means the true population means are likely to fall within this range. On average & rounding up, using the trimmed mean, those who did not face medical discrimination had 7 not good mental health days as opposed to those who did with 14.

Model 2

The second model focused on the relationship between a respondent's basic needs insecurity score and the number of not good mental health days. Beginning with a visual representation of the

data in figure B, there is no clear linear relationship seen within the data. These assumptions were not met and thus the variable of not good mental health days was log transformed. The first assumption is linearity, which is assessed using the 'Residuals vs Fitted' plot in figure C. After transformation, the data is approximately linear. The next assumption is independence, which is met in the study design as mental health days and insecurity scores do not have overlap in their questions. Normality is the next assumption, which is assessed using the 'Normal Q-Q' plot in figure C. The data follows the line enough to be considered normal. Equal variance is the last assumption to be met, which is assessed using both the 'Residuals vs Fitted' and 'Scale-Location' plots; the lack of cone-shaped data implies that the assumption is met. Since the necessary assumptions are met, a linear regression model model could be created. A model was fitted and found to be $\widehat{y} = 5.62 + 0.669x$ where x is the insecurity score. This means there is a shallow, positive relationship between insecurity score and number of bad mental health days. This can be visualized in figure B, where a red line represents this model. A hypothesis test was performed to see if insecurity score is a significant predictor of not good mental health days. The null hypothesis was that the slope is equal to zero and the alternative hypothesis was that the slope was not equal to zero. These are written respectively as $H_0: \beta_1 = 0$ and $H_\alpha: \beta_1 \neq 0$ where β_1 is the slope. The p-value of this test was $<2.20\times10^{-16}$, causing the rejection of the null hypothesis and accepting that the slope is not equal to zero. This means that there is evidence that the two variables are correlated and that insecurity scores are a significant predictor of a resident's number of not good mental health days. The next test performed had the goal of testing to see if the model created was as good as or better than predicting y than just the intercept. This is written as $H_0: Y = \beta_0 + \beta_1 x + \epsilon$ is the same as $Y = \beta_0 + \epsilon$ and $H_\alpha: Y = \beta_0 + \beta_1 x + \epsilon$ is better than $Y = \beta_0 + \epsilon$ where β_0 is the intercept, β_1 is the coefficient of x, and X is the insecurity score. The p-value of this test was $<2.20\times10^{-16}$, meaning the null hypothesis is rejected and the model with insecurity score is better than the model without. The R^2, which quantifies how well the model fits the data, is 0.107. This number is near zero, meaning $\hat{y}$ is not a good predictor of y. This is likely due to confounding factors, such as the respondent's gender, trans status, and race/ethnicity (table 1), are also impacting the number of not good mental health days. Using the model above to predict the number of not good mental health days for a person who has an insecurity score of 5, the number of not good mental health days would be 4.69, which could be rounded to five. Doing the same for a person with an insecurity score of 20 results in 16.0 not good mental health days.

Model 3

The third model focused on the relationship between a resident's number of not good mental health days, their insecurity score, and whether or not they faced medical discrimination. figure D shows these relationships, which seem to have an upward linear trend for both those who faced medical discrimination and those who do not. From the previous two models, it is known that necessary assumptions to create a linear model are met. A multiple linear regression model was created in the style of $\widehat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ where $\hat{y}$ is the predicted value 'number of bad mental health days', β_0 is the y-intercept, β_1 is the slope for insecurity score, x_1 is the insecurity score, β_2 is the slope for medical discrimination, and x_2 is rather or not they did face medical discrimination. This model can be written as $\widehat{y} = 1.348 + 0.076 \cdot x_1 + 0.149 \cdot x_2$.

To see if the number of not good mental health days can be predicted from the insecurity score, controlling for the discrimination in health care status, an ANOVA F-test was used. The null hypothesis was that the model with predictors was as good as that without, the alternative hypothesis was that the model with predictors is better than the model without. This can be written as

H_0: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ is the same as $Y = \beta_0 + \epsilon$ and H_α: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ is better than $Y = \beta_0 + \epsilon$ where β_0 is the intercept, β_1 is the insecurity score slope, x_1 is the insecurity score, β_2 is the medical discrimination slope, and x_2 is if they faced medical discrimination or not. The p-value of this test was $<2.20\times10^{-16}$, meaning that because the value is under 0.05 the null hypothesis was rejected. There is evidence that the model with predictors performs better than the one without. The model is $Y = 1.35 + 0.08 \cdot insecurity\ score + 0.140 \cdot medical\ discrimination$ .The adjusted R^2 value, which explains how much variation is explained by the model, is 0.1238.

To see if the number of not good mental health days can be predicted from whether or not a person faced discrimination in health care, controlling for their insecurity score, an ANOVA F-test was used. The null hypothesis was that the model with predictors was as good as that without, the alternative hypothesis was that the model with predictors is better than the model without. This can be written as H_0: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ is the same as $Y = \beta_0$ and H_α: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ is better than $Y = \beta_0$ where β_0 is the intercept, β_1 is the medical discrimination slope, and x_1 is if they faced medical discrimination or not, β_2 is the insecurity score slope, and x_2 is the insecurity score. The p-value of this test was $<2.20\times10^{-16}$ meaning that because the value is under 0.05 the null hypothesis was rejected. There is evidence that the model with predictors performs better than the one without. better than the one without. The adjusted R^2 value, which explains how much variation is explained by the model, is 0.124. This model has a slightly higher R^2 than the former, so this is the model that should be used for predictions.

Using the model above to predict the number of not good mental health days for someone who experienced discrimination in health care and had a basic needs insecurity score of 5, the number of not good mental health days would be 3.56, which could be rounded to four. Doing the same for a person with an insecurity score of 20 results in 2.63 not good mental health days, which could be rounded to three.

Conclusion

This analysis has focused on the usability of both discrimination in healthcare and insecurity score as predictors of the number of not good mental health days a resident has. It was found that, on their own, both a resident's status of having faced medical discrimination and their insecurity score are both significant predictors for a resident's number of bad mental health days. Together, they also act as a statistically meaningful multiple linear regression model. The best model for predicting a resident's number of bad mental health days is the model that utilizes both medical discrimination data and insecurity score. This is because, of the three explored today, this model accounts for the most confounders and thus allows the number of not good mental health days to be most accurately predicted. This model is still far from perfect as, as was discussed at the beginning of this analysis, there are a number of confounding factors for both predictors that consist of gender identity, transgender status, and race/ethnicity. My critique of this analysis is that there are a number of confounding factors that were addressed in a previous report that were not incorporated into the models of this analysis. Such models were beyond the scope of this analysis, but there are a number of confounding variables that would create a better fitted model and allow a more accurate prediction of a resident's number of not good mental health days. This would, of course, require a far more complex model that is beyond the range of the interest of this analysis.

Appendix

Figure A

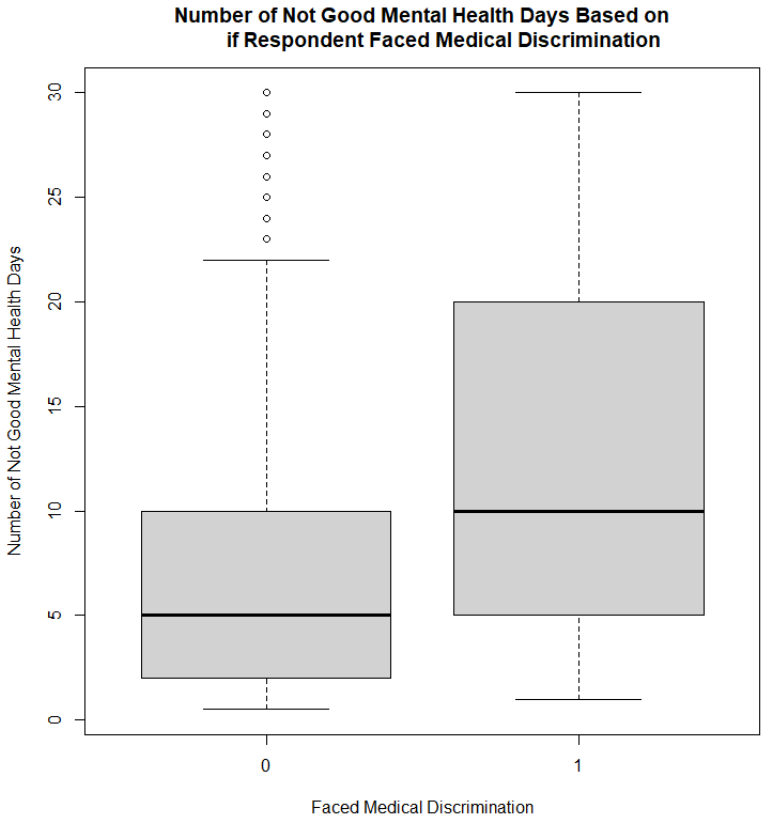

Figure B

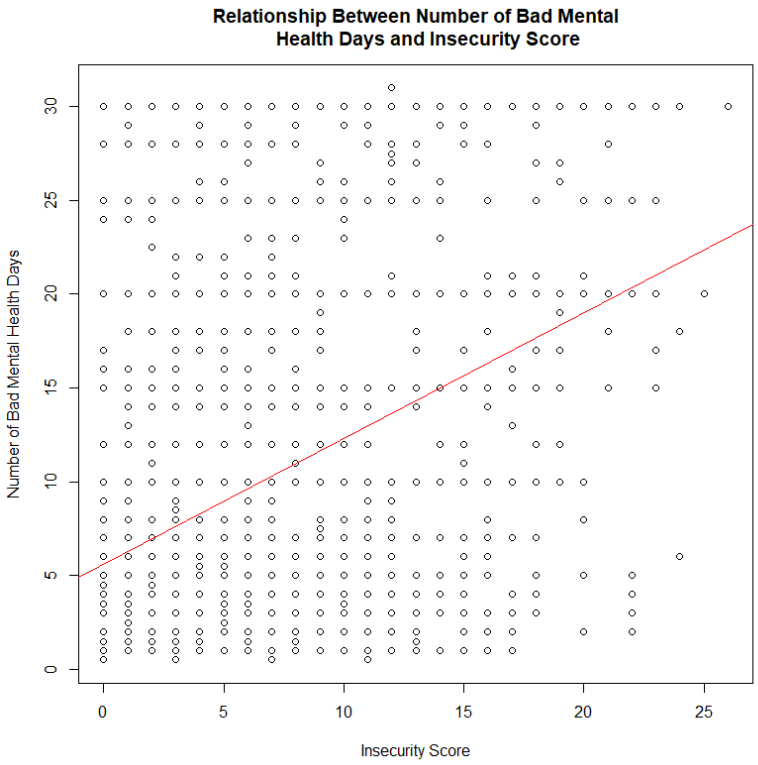

Figure C

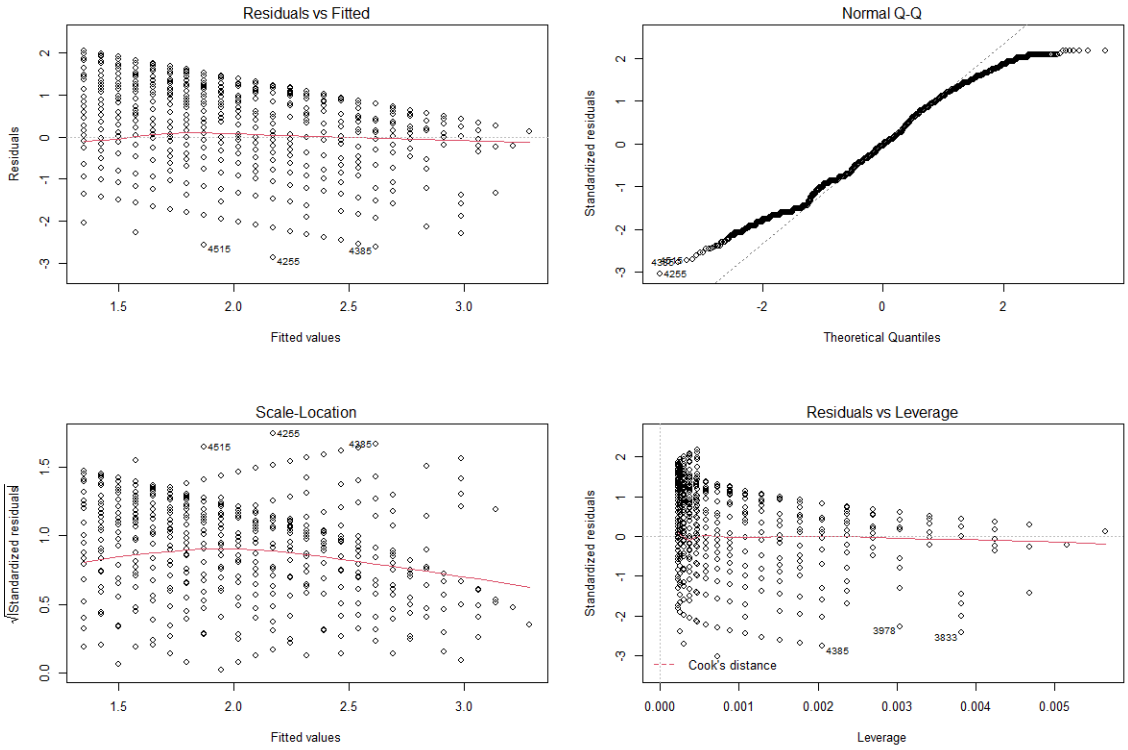

Figure D

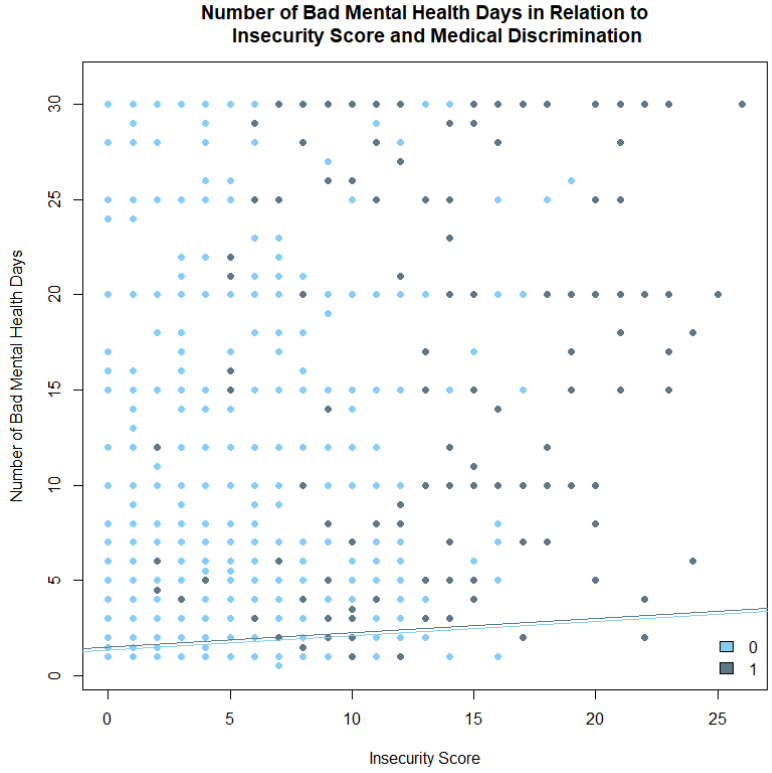

Table 1: Demographics versus Insecurity Score

Variable/Categories	n	Mean	SD	p-value	Differences
Gender				7.96×10^{-3}	
Male	1336	4.57	4.54		**Males < Agender/Genderqueer /Nonbinary**
Female	2967	4.44	4.34		**Females < Agender/Genderqueer /Nonbinary**
Agender/Genderqueer/Nonbinary	87	5.9	3.9		
Trans status				4.24×10^{-4}	
Cis	4338	4.49	4.39		**Cis < Trans**
Trans	62	6.47	4.44		
Race/Ethnicity				$< 2.2 \times 10^{-16}$	
Native 2022 inclusive	193	10.05	5.7		**Native > Hispanic** **Native > Asian** **Native > White** **Hispanic > Asian** **Hispanic < Black** **Hispanic > White** **Asian < Black** **Asian < White** **Black < White**
Hispanic	218	6.39	4.37		
Asian-NH	190	5.32	4.52		
Black-NH	404	9.28	5.4		
White-NH	3158	3.34	3.25		

Table 2: Demographic risk of discrimination in health care

Variable/Categories	Relative Risk[2]	95% CI for RR	p-value
Gender2			0.513
Agender/Genderqueer/ Nonbinary	1.272	[0.731, 2.212]	
Male/Female	1	N/A	
Trans Status			0.091
Trans	1.716	[1.003, 2.935]	
Cis	1	N/A	
BIPOC/AI			0
BIPOC/American Indian	4.569	[3.72, 5.611]	
white	1	N/A	

[2] The relative risk of being treated unfairly or discriminated against in medical, mental, or dental care

Table 3: Demographics versus number of not good mental health days

Variable/Categories	n	Mean	SD	p-value	Differences
Gender				2.63×10^{-13}	
Male	1381	8.37	8.36		**Male < Agender/Genderqueer/ Nonbinary** **Female < Agender/Genderqueer/ Nonbinary**
Female	3057	8.64	8.3		
Agender/Genderqueer/Nonbinary	90	15.27	8.67		
Trans status				4.14×10^{-14}	
Cis	4471	8.58	8.33		**Cisgender < Transgender**
Trans	68	16.31	9.3		
Race/Ethnicity				4.80×10^{-12}	
Native 2022 inclusive	200	11.57	9.27		**Native > Asian** **Native > White** **Hispanic > White** **Asian < Black** **Asian > White**
Hispanic	229	9.77	8.67		
Asian-NH	200	8.71	8.72		
Black-NH	427	10.9	9.36		
White-NH	3239	8.06	7.96		

Table 4:

Medical Discrimination	n	Mean	Trimmed Mean	SD
No	2208	7.8	6.44	7.59
Yes	309	13.8	13.31	10.14

Project 1 Write-Up