Victor Xu
Professor Okonofua & Mendoza-Denton
Psych 167AC: Stigma & Prejudice

Data Science Project: Investigating the Relationship Between Racial Bias (Implicit) & Percent of Population in Poverty

VARIABLES:

- Manipulated (Project Implicit Data): Implicit Bias of Race (D_WhiteGood) across counties within Washington State
- Responding (Outcome Data): Percent of Population in Poverty
- Covariate: Explicit bias of Race (tblacks, on a scale of 0-10)

:

INTRODUCTION

Socioeconomic status (SES) is defined by the American Psychological Association as the "social standing or class of an individual or group; often measured as a combination of education, income and occupation." ("Socioeconomic") Closer examinations of socioeconomic inequalities are inseparable from concepts of 'privilege, power and control'. Thus, low SES and its correlates, such as poverty levels, are entwined with the identification of race and ethnicity. ("Ethnic") Research shows that "minority racial groups are more likely to experience multidimensional poverty than their White counterparts." (Reeves, Rodrigue, Kneebone) On a broader scope, American Indian, Hispanic, and Pacific Islander families are more likely than Caucasian and Asian families to live in poverty. (U.S. Census) Furthermore, African-American men "working full-time earn only 72 percent of the average earnings of comparable Caucasian men." (Rodgers)

My rationale for this project is to explore whether implicit bias of race (limited to African Americans through the Project Implicit Data), is indeed correlated with the percentage of population in poverty. Implicit bias might be perceived as one starting point of SES issues, ultimately reflecting in larger social phenomenons like poverty. These variables could create a negative feedback loop, which negatively disrupt social order. The social phenomena I predict to find across county-level data, is greater implicit racial bias linked with an increase in the percent of the population in poverty. The covariate, explicit racial bias, explores whether *recognized* bias actually has a stronger link with poverty rates. Aside from investigating these variables to confirm/contradict existing research literature and critical theories, I'm curious to ponder on the ramifications of potential correlations. This includes potential interventions towards implicit racial bias, and whether concepts like racial anxiety or stereotype threats are contributors to the process. (Godsil) However, these interests are mostly secondary (out-of-scope for this particular project), and will just be briefly touched upon towards the end of this report.

METHOD

Datasets & participants: Data for the implicit bias of race (and explicit bias covariate) were drawn from Project Implicit* directly. The explicit bias are thermometer measures, obtained by asking participants to rate how warm they felt towards a target group (for this particular dataset, whites) and a contrast group (African Americans) on scales from 0 (coldest) to 10 (warmest). To simplify the data, I'll be looking specifically at tblacks (ethnic minority) in relation to poverty statistics. For implicit bias, values were obtained through a calculated Implicit Association Test score centered around zero. For this project, when D_WhiteGood is higher, there's a greater implicit preference towards whites (and perhaps by extension, greater bias towards minorities). These implicit + explicit data values were merged with the outcome dataset (percent of population in poverty) through county-level identification, during the pre-processing stage. Note that outcome dataset actually contains statistics ranging from year 2000-2013. But because the racial bias data doesn't go into the same level of granularity, the newest year/data, 2013, was chosen. As a Seattle native, I've decided to further limit this project to include Washington state counties only. The reason is twofold: Firstly, to prevent the merged dataset from being overly large and vague in the analysis stage. Secondly, to gauge the outcomes with first-hand observations about places I'm familiar with.

Below is my procedure for merging and pre-processing the data:

Step 1: Importing the Race data

In [15]:
```
my_data = Table.read_table('Implicit-Race_IAT.csv')
my_data
```

Out[15]:

FIPS	twhites	tblacks	D_WhiteGood
1001	6.89714	6.09714	0.399913
1003	7.28256	6.32678	0.401266
1005	6.74286	6.92857	0.2871
1007	7.23913	7.04348	0.267628
1009	7.15152	6.40909	0.409505
1011	7.75	7.65	0.120589
1013	7	6.75	0.199301
1015	7.1955	7.02595	0.291929
1017	7.10853	7.46512	0.233866
1019	7.46154	6.76923	0.314446

... (3177 rows omitted)

Step 2: Importing the Poverty (outcome) data

In [16]:
```
my_data2 = Table.read_table('Outcome-Poverty.csv')
my_data2
```

Out[16]:

State	FIPS	County	Year	Percent_of_Population_in_Poverty	Stability
Alabama	1001	Autauga	2000	10.5	1
Alabama	1001	Autauga	2001	10.8	1
Alabama	1001	Autauga	2002	10.3	1
Alabama	1001	Autauga	2003	10.4	1
Alabama	1001	Autauga	2004	11.6	1
Alabama	1001	Autauga	2005	10.4	1
Alabama	1001	Autauga	2006	12.5	1
Alabama	1001	Autauga	2007	10.4	1
Alabama	1001	Autauga	2008	10.7	1
Alabama	1001	Autauga	2009	11.2	1

... (43964 rows omitted)

Step 3: Merging and eliminating unnecessary columns (please observe code comments)

```
# work on subsetting and joining your data here

#In this step, I am merging the 2 data tables together
combined_data = my_data.join("FIPS", my_data2)

#In this step, I am filtering out the explicit bias data because I am focusing on implicit for this project
select_combined_data = combined_data.select("FIPS", "D_WhiteGood", "State", "County", "Year", "Percent_of_Population_in_Poverty")

#To simplify the table, I'll just be taking the most current percent poverty data (year 2013). This is also because
#the implicit bias data doesn't have the granularity of years, so it wouldn't make sense to have the same D_WhiteGood
#implicit bias while percent poverty changes per year (for reach county)

new_census_data = select_combined_data.where("Year", are.equal_to(2013))

#This is still too much data, so I decided to just focus on counties within Washington State.
#The output is much more managable and easy to plot + experiment with correlations
new_census_data2 = new_census_data.where("State", are.equal_to('Washington'))

#Finally, I'm dropping the columns State and Year - since we know all the retained data is Washington State and 2013
final_census_data = new_census_data2.drop("State", "Year")
final_census_data
```

Step 4: The final output (merged + processed dataset) for Washington State Counties, 2013:

FIPS	D_WhiteGood	County	Percent_of_Population_in_Poverty
53001	0.359445	Adams	17.6
53003	0.325221	Asotin	16.8
53005	0.306126	Benton	13.2
53007	0.341816	Chelan	16
53009	0.339222	Clallam	17.5
53011	0.328837	Clark	12.6
53013	0.332785	Columbia	17
53015	0.344823	Cowlitz	15.9
53017	0.296598	Douglas	14.4
53019	0.38021	Ferry	22.1

... (29 rows omitted)

Note: Dataset, stored in my Jupyter notebook (which I used to complete the regression + plotting analysis) is complete. The output here has been visually truncated for neatness.

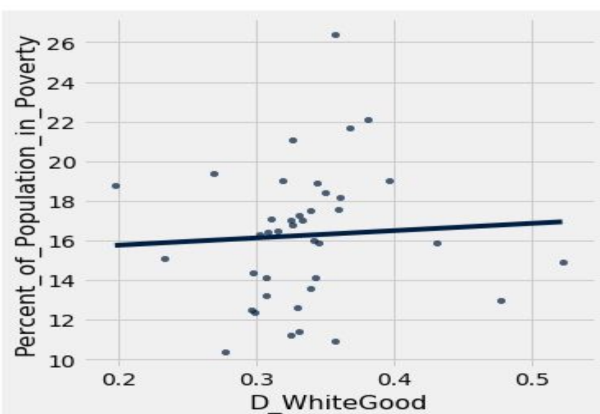
Analysis: With the merged dataset of racial bias and population in poverty, I ran my analysis through Jupyter, a cloud-based data science programming tool. First, I plotted the implicit bias of race (D_WhiteGood, x-axis) with poverty statistics (y-axis) in a scatterplot to observe general patterns. Then, I ran a correlation between these two variables, to find the correlation coefficient (0 meaning no linear correlation, 1 or -1 representing perfect correlations).

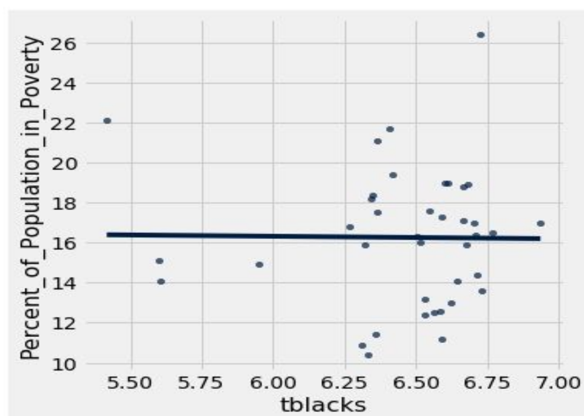
Next, I performed the linear regression (best fit for graph), and found summary statistics like the p-value (where a number < .05 is generally considered as statistically significant). Lastly, I went back and plotted the explicit bias of race (tblacks, x-axis), against poverty statistics (y-axis). The computational process for correlation and regression were repeated once more.

RESULTS

For correlation, I found that for **implicit** racial bias (against percent of population in poverty), there was a correlation coefficient of 0.06044946. This indicates a very weak positive correlation between higher D_WhiteGood and percent of population in poverty. Looking at the graph, datapoints are the most dense in the range of 0.3-0.4 for D_WhiteGood. What is interesting is that when D_WhiteGood is below 0.4, there seems to be a much stronger correlation (linear, upwards) between implicit racial bias and poverty statistics. However, when D_WhiteGood increases beyond 0.5, the percent of population in poverty actually drops. For the **covariate** (explicit bias) and poverty, the correlation coefficient is -0.01205063. This indicates an even weaker, negative linear relationship between tblacks and poverty statistics. For regression of implicit bias, we have 0.714694 as the p-value. The R-square is 0.004, whereas the coef is 0.0010. For the covariate, we have a p-value of 0.941957, R-square of 0.000 (not displayed within 3 significant digits), and coef of -0.0011.

Figure 1: Poverty Stats vs DWhiteGood (implicit)   Figure 2: Poverty Stats.vs. tblacks (covariate)





DISCUSSION

The findings suggest several things: Firstly, that the covariate is demonstrating an even weaker relationship with the output variable than the manipulated (implicit bias). We can observe this from the best-fit line having a flatter slope, weaker correlation coefficient, and larger p-value (getting further from being statistically significant). Each unit increase in tblacks also corresponds to less change in the percent of population in poverty (coef value), and the R-squared value is almost zero (meaning the regression line does a poor job of fitting the data). Still, the covariate best-fit having an overall negative slope, shows that in counties with less explicit bias towards African Americans, there's a reduction in the poverty percentages. For the implicit bias, we observe a stronger relationship; where counties having higher implicit bias, also have a higher percentage of poverty. With this being said, the results are still not demonstrating a strong correlation, with the p-value far from being statistically significant.

Because of this, my original hypothesis was not confirmed statistically, although the covariate was also rejected. The result does raise implications like why in counties with higher tblack scores, the poverty percentage isn't significantly lower. Possible limitations include having a relatively smaller/homogenous dataset (Washington state) to observe trends (though I initially thought of it as a positive), or perhaps choosing twhites instead for the covariate (although the contrast would be less interesting between D_WhiteGood). Finally, SES inequalities are highly complex - with perhaps other hidden covariates and confounding variables. Potential directions for future research might include: replacing the covariate with other variables (i.e. education level, demographic percentages of counties - whether having more minorities would change the biases of a particular place) to test alternative links, replication with other minority groups (Hispanic, Asian), and interventions (whether decreasing implicit racial bias through education/video/tutorial would significantly affect the poverty percentage, etc.).

APPENDIX: Brief Description of Project Implicit* Dataset

From Project Implicit: "Project Implicit was founded as a multi-university research collaboration in 1998 by three scientists - Tony Greenwald (University of Washington), Mahzarin Banaji (Harvard University), and Brian Nosek (University of Virginia), and was incorporated as a non-profit in 2001 to foster dissemination and application of implicit social cognition.
There are 14 IAT studies on the Project Implicit Demonstration site:

- Age IAT
- Arab-Muslim IAT
- Asian IAT
- Disability IAT
- Gender-Career IAT
- Gender-Science IAT
- Native IAT
- Presidents IAT
- Race IAT
- Religion IAT
- Sexuality IAT
- Skin-tone IAT
- Weapons IAT
- Weight IAT

Most of the studies have been collecting data online for around 10 years. This OSF project will provide the archived data sets with codebooks for researchers who are interested in the PI demo site data to do their own analysis. Most of the data have not been well analyzed for publication. Given the length and consistence of the PI data collection, researchers may answer interesting and novel questions with the data sets.

Each separate OSF sub-project will have data sets year by year and for all years, codebooks associated with data sets, experiment materials used in the current version, and skeleton syntax to facilitate analysis.”

Work Cited (Referenced Literature):

[1] "Ethnic and Racial Minorities & Socioeconomic Status." American Psychological Association, apa.org/pi/ses/resources/publications/minorities

[2] Godsil, Rachel "Breaking the Cycle: Implicit Bias, Racial Anxiety, and Stereotype Threat." *Poverty & Race: Poverty & Race Research Council*. 24.1 (2015): 1-9. Web. 28 Nov. 2017.

[3] Reeves, R., Rodrigue, E., & Kneebone, E. (2016). *Five evils: Multidimensional poverty and race in America.* Retrieved from brookings.edu/wp-content/uploads/2016/06/ReevesKneeboneRodrigue_MultidimensionalPoverty_FullPaper.pdf

[4] Rodgers, W. M. (2008). Understanding the Black and White earnings gap: Why do African Americans continue to earn less despite dramatic gains in education? Retrieved from prospect.org/cs/articles?article=understanding_the_black_white_earnings_gap

[5] "Socioeconomic Status." American Psychological Association, American Psychological Association, http://www.apa.org/topics/socioeconomic-status/.

[6] U.S. Census Bureau. (2014). U.S. Poverty Report. Retrieved from https://www.census.gov/population/projections/data/national/2014.html