

HW 3: Correlation and Regression

1. In this assignment I will be analyzing three different potential correlations for the belief in contemporary US discrimination towards African Americans. Historical, institutionalized, and codified accounts and occurrences of racial discrimination have been both insidious and overt aspects of American society since the implementation of the 13th Amendment, and thus this is both an interesting and important area of study for political and legal pursuits.

- All of the data analyzed is sourced from the American National Election (ANES) 2016 Pre-Test/Pilot Data

-The dependent variable (DV) of interest in this analysis will be the raw unrecoded index from HW 2, which, as aforementioned, measures the concept of belief in contemporary discrimination towards African Americans in the US context, measured by the perceived notions of police treatment on the basis of race, the general level of perceived discrimination that African Americans receive, as well as the perceived frequency with which police officers stop African Americans as opposed to other races.

-In a previous analysis of this raw index, it was determined that there was an acceptable level of reliability between the three indicators used to construct the index, with a **Cronbach's Alpha score of .637**.

-Below is a tabular depiction of the raw index:

	Frequency	Percent	Valid Percent	Cumulative Percent
.00	6	.8	.8	.8
.50	27	3.9	3.9	4.7
1.00	94	13.7	13.8	18.6
1.50	134	19.6	19.7	38.3
2.00	123	18.1	18.2	56.4
2.50	124	25.2	18.2	74.7
3.00	172	99.4	25.3	100.0
Total	680	99.4	100.0	
Missing System	4	.6		
Total	684	100.0		

-In terms of univariate statistics, the raw index produced the following:

There is a large amount of variance for the raw index, and in terms of the mean, the most accurate in terms of predicting the response of the respondents, most held a moderate belief in the existence of contemporary discrimination of African Americans, the median, or mid-point in the data reflected this result as well, but the mode, or the highest selected response demonstrates that the respondents strongly believed in the contemporary discrimination towards African Americans.

Mean	2.0325
-------------	---------------

Median	2.0000
Mode	3.00
Std. Deviation	.77920
Variance	.607

-Hypotheses:

a. H₁ Shorthand Equation:

Likelihood of Involvement in Local Political/Social Meeting → ↑ Raw Index

- IV: meet: Involvement in Local Meetings
- “In the future, how likely are you to attend a meeting to talk about political or social concerns (extremely likely, very likely, moderately likely, a little likely, or not at all likely)?”
- The concept this independent variable and subsequent hypothesis is meant to measure is essentially: the more likely someone is to attend a local meeting about political concerns, the higher their belief will be in the contemporary discrimination of the African American population in the US. This rationale behind this hypothesis is that the more involved someone is in local political and social concerns, the more in tune they would be with national concerns, thus making it more likely that they would believe in the occurrence of racial discrimination towards African Americans.
- Missing Values/Recodes: I assigned the 8/9 values of ‘don’t know/no response’ as missing, and I then recoded the indicator into ordinal level data with three categories of high, medium, and low according to their previous responses; the high scores received a score of 1 and combined the groups ‘extremely’ and ‘very likely,’ the medium responses were assigned the score .5 and combined the ‘moderately’ and ‘a little likely’ responses, and the low scoring group were labeled as 0 and included the ‘not at all likely’ responses. The indicator was renamed localpol.

b. H_{2/3} Shorthand Equation:

H₂: Democrats → ↑ Raw Index

H₃: Republicans → ↓ Raw Index

- IV: pid1d & pid1r: Party Identification
- “Generally speaking, do you usually think of yourself as a republican, democrat, an independent, or what?”
- This IV measures party identification. H₂, or the first of the above mentioned hypotheses pertaining to party, states that respondents who self-select as democrat will be more likely to answer in accordance with a stronger belief in the racial discrimination occurring in the US against the African American population. This is based upon the general correlation between a liberal political and ideological outlook reflected in the democratic voters, and their historic support of policies to redress racial discrimination in the US. The latter of the referenced hypotheses, H₃ states that the respondents who will self-select as affiliated with the republican party will be less likely to have a strong belief of the issue of racial discrimination of

African Americans; in other words their responses to the questions contained in the index will point towards a lack of belief in the severity of the issue of racial discrimination towards African Americans. This speaks to again the correlation of ideology with party affiliation, republicans tend to be more conservative, and thus in recent policy solutions have reflected the belief that we are inculcating a colorblind society that does not need to deal with issues of this nature in any significant way in terms of redress.

- Missing Values/Recodes: As in other ANES data, I declared the 8/9 answers to be missing, as well as the 'something else' response, which only accounted for 17 responses or 2.4% of the responses. The reason why these two hypotheses are contained within one subsection of this report reflects how the indicator was originally coded and then recoded; beyond the fact that both hypotheses regard party identification, I recoded the two questions via a conditional transformation in order to account for all but 2.4% of the respondents in the same variable. I then decided to recode this nominal level data that included democrats, independents, and republicans into a dummy variable with the reference category of independents since this category had the highest N and was easy to compare to the other categories (meaning later on during the regression analysis, independent was eliminated as a category for regression).

c. H₄ Shorthand Equation: Acknowledgment of white privilege → ↑ Raw Index

- IV: wad1b: whiteness as a societal advantage.
- "Does being white help you, hurt you, or make no difference for you personally in today's society?"
- This IV measures whether or not a white responder, (I will discuss this distinction in the recode section below), believes that their respective whiteness is a societal advantage. In terms of the hypothesis, I state above that the stronger the belief that their whiteness privileges them, the higher the likelihood that they will believe in the occurrence of racial discrimination of the African American population in the US. I believe this has a strong correlation because if a white individual is able and willing to acknowledge the privilege they maintain in society writ large, the more aware they will be of other racial groups' relative disadvantage.
- Missing Values/Recodes: This question, as was previously mentioned, is only asked if the respondent self identifies as white, and thus the N is much lower than the other IVs, (at 257 to their relative ~684 range). The missing values of 8/9 were declared as such, and the data was recoded into the ordinal data configuration of low, (whiteness hurts them), medium, (makes no difference), and high, (helps them significantly), each with the respective score of 0, 0.5, and 1.

2. Correlation Matrix:

Correlation among Belief in Contemporary Racial Discrimination and Three Variables

*Note: Cell entries are Pearson's Correlation coefficients with statistical significance indicated in parentheses.

	Racial discrim. belief	Dem	Rep	Whitepriv
Racial discrim. belief				
dem	.385(0.00) N=680			
rep	-.251(0.00) N=680	-.405(0.00) N=680		
whitepriv	.360(0.00) N=257	.166(0.00) N=257	-.080(.204) N=257	
localpol	.067(0.00) N=680	.091(0.017) N=684	.023(.541) N=684	.123(.049) N=257

3. Discussion of Correlation Matrix:

a. H₁: Likelihood of Involvement in Local Political/Social Meeting → ↑ Raw Index

- There is a very weak but positive correlation between the likelihood of involvement in a local political meeting and belief in racial discrimination. (0.067). It is statistically significant, meaning this very weak, positive correlation is not due to chance. The r^2 of this correlation is .0044, meaning the independent variable in this case only explains .44% of the variation in the dependent variable. This data analysis would indicate that there is no real correlation between these two factors, disproving the hypothesis in terms of strength, but not direction. This IV does seem to be positively correlated with acknowledgment of white privilege, meaning that the more likely you are to attend local political meetings, the higher the chance you agree with the advantages of whiteness in society. The r or correlation coefficient in this case was .123, meaning that 1.5% of the dependent variable's variation would have been explained if it were statistically significant, which it was not.

b. H₂: Democrats → ↑ Raw Index

- There is a strong positive correlation between the IV and the DV, proving this hypothesis. The correlation coefficient is .38, a strong qualification, and is statistically significant meaning that this correlation is not due to chance, and the r^2 is then .144, in turn indicating that 14.4% of the variation in the dependent variable can be explained by the independent variable. Democrats are more likely to believe that racial discrimination exists in the US. This IV does not have a strong correlation with any other IVs.

c. H₃: Republicans → ↓ Raw Index

- There is a moderately strong negative correlation between republicans and belief in racial discrimination in the US, proving the hypothesis to be true. The correlation coefficient is .251 and is statistically significant, meaning that the r^2 would be .063 or in other terms that the IV of republican party identification would explain 6.3% of the variation in the dependent variable. Not surprisingly, the IV of party identification as republican has a very strong and statistically significant negative correlation to the party identification variable of democrats, the r is -.405.

d. H₄: Acknowledgment of white privilege → ↑ Raw Index

- There is a strong positive correlation between this IV and the raw index, showing that, as indicated by the hypothesis, as there is increasing acknowledgement of the societal privilege imbued by whiteness, the greater the belief in racial discrimination in the current US climate. The correlation coefficient is .360 and is statistically significant; the r^2 value would be .129 meaning that almost 13% of the variance of the index can be explained by the dependent variable. This IV has no statistically significant or otherwise correlation with any of the other IVs.

4. Regression:

Prediction of Belief In Racial Discrimination with Party, Political participation, and Acknowledgment of White Privilege

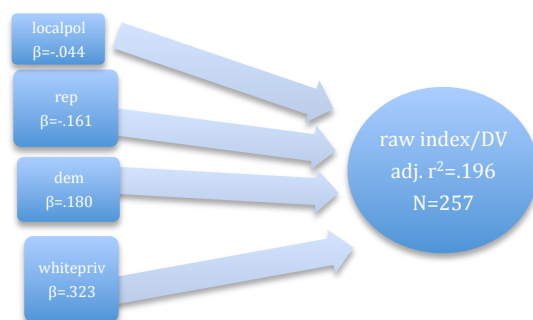
	B	Std. error	Beta (β)	Tol
dem	.310(.004)	.107	.180	.832
rep	-.279(.008)	.105	-.161	.857
whitepriv	.727(0.00)	.129	.323	.961
localpol	-.096(.440)	.124	-.044	.977
constant	1.493(0.00)	.107		
N	257			
R ²	.208	.692		
Adjusted R ²	.196	.692		

Note: Cell entries in the unstandardized b or regression coefficient column have statistical significance indicated in parentheses.

- Regression Equation:

Raw Index = 1.493 + .310(dem) - .279(rep) + .727(whitepriv) - .096(localpol)

- Pictorial Depiction of Relative Strength in Beta scores (standardized results):



- Overall, from this regression analysis we can conclude that the regression equation is statistically significant, as indicated in the ANOVA results of the regression, the

score was 0.000. In terms of individual IVs, all but the localpol can be considered statistically significant.

- Dem:(.004)
- Rep: (.008)
- Whitepriv: (.000)
- There was glaring statistical insignificance for localpol, (.440) so in the results in the regression analysis for it may be due to chance.
- The regression equation explained 19.6% of the variance in the regression model, which is a relatively weak regression model. However, as displayed by the diagram, there were some strong standardized regression coefficients for both the IVs of Democratic party and acknowledgment of white privilege, further proving the conclusions pertaining to the efficacy of the hypotheses from the earlier correlation analysis.

5. In terms of unstandardized regression coefficients, the most effective in terms of explaining the variation in the dependent variable was by far the whitepriv independent variable at a level of .727, which indicates an almost suspiciously strong positive correlation. In other words, the acknowledgment of ones white privilege was the single highest predictor of positive change on beliefs pertaining to racial discrimination against African Americans occurring in the contemporary US context. The next highest unstandardized b was the democratic party, as predicted in terms of a strong positive correlation with a strong score of .310. The republican party IV was the next highest with a moderate negative score of -.279 followed by the regression coefficient which explained the least amount of change to the DV, localpol, with a negative score of -.096. With regard to standardized beta scores, the diagram above orders them from least (at the top) to most variation explained by the beta score, and although the results are within a closer range of each other via the metric standardization by standard deviation, the most predictive IV is still whitepriv, followed still by dem, then rep, and then localpol. The regression results differ slightly from the earlier correlation results: the high scorer for correlation was the democrat IV, followed by the whitepriv, and then localpol. The other notable difference is that the dem and rep were higher scoring in terms of Pearson's r and depicting an obvious correlation between the IVs and the indexed DV.

6. The adjusted R^2 for this regression analysis was .196, meaning that 19.6% or roughly 20% of the variance in the DV was explained by the regression model. This is a moderate score in terms of explanatory power, and if time permitted I would exchange the localpol IV for an IV that would further improve the efficacy of the model. According to the Tol or tolerance column in the regression table, there is a high tolerance for all of the IVs, (not a single IV came close to the .2 lower limit/barrier), and thus I can confidently say that there was no collinearity collapsing the regression equation/model. The interrelationship between the dem and rep IV in the correlation analysis were not replicated in the regression analysis.

Syntax:

* Encoding: UTF-8.

weighting the data

Weight by weight_spss.

dependent variable

```
fre var disc_police.
missing values disc_police (8,9).
recode disc_police (1,2,3=1) (4=.5) (5,6,7=0) into policetreat.
value labels policetreat 0 'treat blacks better' .5 'same' 1 'treat whites better'.
fre var policetreat
  /statistics mode median mean stdev skew kurtosis.
```

```
fre var disc_b.
missing values disc_b (8,9).
recode disc_b (1,2=1) (3=.5) (4,5=0) into discrim.
value labels discrim 0 'little to none' .5 'moderate' 1 'a lot'.
fre var discrim
  /statistics mode median mean stdev skew kurtosis.
```

```
fre var stopblack.
missing values stopblack (8,9).
recode stopblack (1,2=0) (3=.5) (4,5=1) into racialprof.
value labels racialprof 0 'rarely, if ever' .5 'sometimes' 1 'often'.
fre var racialprof
  /statistics mode median mean stdev skew kurtosis.
```

reliability analysis

```
reliability
  /variables= policetreat discrim racialprof
  /scale (policetreat) all
  /summary = all.
```

Index

```
compute RawIndex = policetreat + discrim + racialprof.
fre var RawIndex
  /statistics = mean median mode stddev var.
```

IV prep

party

```
missing values pid1d (8,9).
```

missing values pid1r (8,9).

```
if (pid1d=1) or (pid1r=1) pidboth =1.
if (pid1d=2) or (pid1r=2) pidboth =2.
if (pid1d=3) or (pid1r=3) pidboth=3.
if (pid1d=4) or (pid1r=4) pidboth=4.
value labels pidboth 1 'democrat' 2 'republican' 3 'indep' 4 'other'.
fre var pidboth.
```

```
*dummies*
*party dummy*
```

```
missing values pidboth (0,4,8,9).
fre var pidboth.
recode pidboth (1=1) (else=0) into dem.
recode pidboth (3=1) (else=0) into indep.
recode pidboth (2=1) (else=0) into rep.
fre var dem indep rep.
```

```
*white privledge*
```

```
fre var wad1b.
missing values wad1b (8,9).
recode wad1b (1 thru 3=1) (4=.5) (5 thru 7=0) into whitepriv.
value labels whitepriv 1 'high' .5 'med' 0 'low'.
fre var whitepriv.
```

```
*local political organizing*
```

```
missing values meet (8,9).
recode meet (1,2=1) (3,4=.5) (5=0) into localpol.
value labels localpol 1 'high' .5 'med' 0 'low'.
fre var localpol.
```

```
*correlation analysis*
```

```
correlations RawIndex dem rep whitepriv localpol.
```

```
*regression*
```

```
regression variables = RawIndex dem rep whitepriv localpol
  /statistics = anova coeff r tol
  /descriptives = n
  /dependent =RawIndex
  /method = enter.
```