**Examining the Effects of Race on College Graduation**

**Sameeullah Babar**

**SDS 358**

****

**Introduction**

**Objectives:**

* This project will examine the effects race has in moderating High School graduation rates among individuals from underserved communities.

**Hypotheses:**

* I believe that Race will have a significant interaction with High School graduation rates.

**Methods**

**Sample:**

* My Data comes from surveys done by the My Brother’s Keeper program over the years 2000-2013.
* The data was divided into race for every year. Race is made of 4 categories (White, Black, Hispanic, Asian) with White as the reference group.
* The result variable is percent of college graduation for the sampled race, and Race will be interacting with percentage of High School Graduates.
  + Other variables include Math Scores (based on a standardized test score administered biennially), percent of violent crime and drug use

**Analysis Method:** A Multivariate Regression with a quantitative by categorical interaction will be used to analyze this data

**Descriptives**

**Response Variable:**

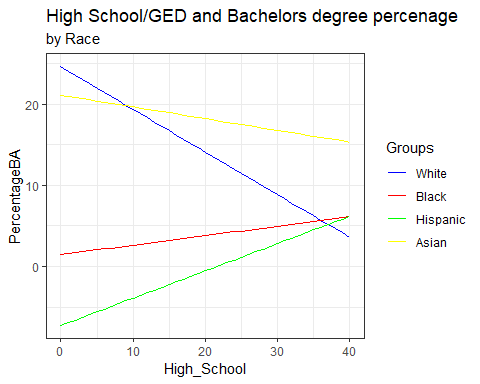
|  |  |  |
| --- | --- | --- |
|  | **Mean** | **Standard error** |
| PercentageBA | 9.316 | 5.951 |

**Explanatory Variables:**

|  |  |  |
| --- | --- | --- |
|  | mean | Standard Error |
| Math | 238.75 | 15.902 |
| High\_School | 27.923 | 6.871 |
| violent\_crime | 4.641 | 1.849 |
| Drugs | 20.121 | 5.023 |

**Results**

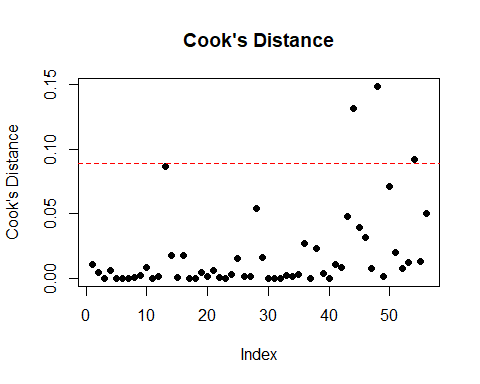
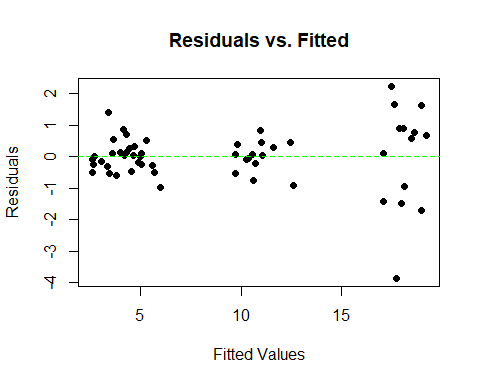
## lm(formula = PercentageBA ~ Race + Math + violent\_crime + Drugs +   
## High\_School + Race \* High\_School, data = Bach)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.8594 -0.3578 0.0396 0.4717 2.2265   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 13.82949 9.48192 1.459 0.15164   
## RaceBlack -23.16088 8.22032 -2.818 0.00716 \*\*  
## RaceHispanic -31.90625 11.08481 -2.878 0.00610 \*\*  
## RaceAsian -3.49702 5.97977 -0.585 0.56160   
## Math 0.03422 0.02374 1.441 0.15648   
## violent\_crime 0.05500 0.17582 0.313 0.75584   
## Drugs 0.11784 0.07039 1.674 0.10108   
## High\_School -0.52538 0.20760 -2.531 0.01495 \*   
## RaceBlack:High\_School 0.64170 0.26550 2.417 0.01977 \*   
## RaceHispanic:High\_School 0.86217 0.35791 2.409 0.02016 \*   
## RaceAsian:High\_School 0.38075 0.23954 1.590 0.11895   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.01 on 45 degrees of freedom  
## Multiple R-squared: 0.9765, Adjusted R-squared: 0.9712   
## F-statistic: 186.6 on 10 and 45 DF, p-value: < 2.2e-16



Anova Table (Type III tests)  
##   
## Response: PercentageBA  
## Sum Sq Df F value Pr(>F)   
## (Intercept) 2.168 1 2.1273 0.151645   
## Race 18.027 3 5.8958 0.001752 \*\*  
## Math 2.117 1 2.0768 0.156479   
## violent\_crime 0.100 1 0.0979 0.755841   
## Drugs 2.856 1 2.8021 0.101078   
## High\_School 6.527 1 6.4044 0.014950 \*   
## Race:High\_School 7.616 3 2.4909 0.072268 .   
## Residuals 45.865 45   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Race = White:  
## contrast estimate SE df t.ratio p.value  
## 1 - 0 -0.5253818 0.2076048 45 -2.531 0.0150  
##   
## Race = Black:  
## contrast estimate SE df t.ratio p.value  
## 1 - 0 0.1163204 0.1523105 45 0.764 0.4490  
##   
## Race = Hispanic:  
## contrast estimate SE df t.ratio p.value  
## 1 - 0 0.3367835 0.2514278 45 1.339 0.1871  
##   
## Race = Asian:  
## contrast estimate SE df t.ratio p.value  
## 1 - 0 -0.1446333 0.1272529 45 -1.137 0.2617

**Assumptions**

The model is homoscedastic and has no values with a Residual X Fitted value greater than 3 \* Cook’s Distance( 3 \* 0.089)

**Discussion**

**Interpretation:** The overall model was significant with an F(10,45) = 186.6, p < .05 and had an R-Squared of 0.9765. The coefficients for Race\*High\_School were significant for the Black and Hispanic groups(t(45) = 2.417, t(45) = 2.409 respectively, p < 0.05), but the interaction itself was not significant (F(3,45) = 2.491, p > 0.05). This is an occurrence commonly known as a Crossover Interaction, where an interaction is significant depending on what level the categorical variable is on. This can be further seen in the simple slopes graph, as White and Asian(-0.525, t(45) = -2.531, p < 0.05 and -0.144, t(45) = -1.137, p > .05) have slopes that are almost opposite to those of Black and Hispanic(0.166, t(45) = 0.764 p > 0.05 and 0.337, t(45) = 1.339, p > 0.05.

**Limitations:**

* The data was collected and aggregated from around the country, excluding many potential regional variables and leading to larger error
* Mixed race was not factored in due to large chunks of missing data

**Implications:** Seeing as White and Asian have drastically different results to Black and Hispanic, it is a good idea to examine other factors such as cultural effects and family situations. Additionally, it seems that Black and Hispanic groups need to be more thoroughly investigated to see why these groups in particular seem to have such a significant impact on the response variable.

**References**:

* My Brother’s Keeper data: <https://catalog.data.gov/dataset/my-brothers-keeper-key-statistical-indicators-on-boys-and-men-of-color/resource/05614937-d0ba-44a4-91f5-2ceeb87bfef1>
* Overview of Crossover Interaction: <https://www.theanalysisfactor.com/interactions-main-effects-not-significant/>