**Sarah Roach**

**Geo 310**

**Final Project Report**

**Introduction**

For my final project, I examined the differences between Black and White life expectancies by county in Kentucky and its surrounding states, specifically Indiana, Tennessee, Ohio, Illinois, and Virginia. I decided on these states, as they generally have similar populations and socio-economics conditions due to their shared geographic location in the Midwest. Life expectancy is a good measure to use when examining differences and change over time between groups of people in an area and is easily understandable. Life expectancy data and its overall trends are indicative of the overall health of a given population. It can also help to identify potential problems that shorten life expectancy. For example, according to the Hamilton Project, overall life expectancy has dramatically increased from 46 years for men and 48 years for women in 1900 to 76 years for men and 81 years for women in 2015 due to increase technology and health care, but disparities among different groups still exist due to inequalities.

It is important to examine differences between Black and White people in the United States, as they can shed light on structural racism and inequalities. These economic and social disparities and inequities culminate and ultimately negatively impact the life expectancy of Black people. This ties back to the practices of slavery, Jim Crow policies, and other racist practices such as redlining that have subjugated many Black Americans to a stressful life of poverty within racially segregated urban centers across the country, often promoting and exacerbating health issues. Making the connection between the legacy of structural racism and its ties to overall health and life expectancy/quality will help to solve these problems. If being a Black American automatically shaves a few years of life compared to a White American, then there is an obvious and widespread problem that must be addressed to promote equity.

My research questions include:

How does race correlate with life expectancy between Black and White populations in each state?

Does a higher segregation rate correlate with lower life expectancies for Black populations?

Do urban areas have a wider disparity between life expectancies for Black and White populations?

Which county has the widest disparity between the life expectancy of Black and White populations and how does this compare to their segregation and urban/rural rates?

Which county has the smallest disparity between the life expectancy of Black and White populations how does this compare to their segregation and urban/rural rates?

My hypothesis is that in most counties, Black life expectancy will be lower than White life expectancy. I also think that higher segregation rates, as well as higher urban living rates, will correlate with lower Black life expectancy. This is due to the aforementioned racist policies that contribute to disproportionately higher rates of disease and poverty among Black Americans.

**Data Sources and Methods**

See Google sheet for TSDOM Tests:

<https://docs.google.com/spreadsheets/d/1PsIv6CXMiBSBrcmrcFnKrYh9NEKktKI5_vsLPp_V4ng/edit?usp=sharing>

Data Table: <https://docs.google.com/spreadsheets/d/124T6L6WSKnC67-G6FKcg5nhK-eD2o3FdF5BU8nwrgdY/edit?usp=sharing>

To address my research questions, I had to obtain life expectancy data by race for each of my selected counties. The data source I used for this is [www.countyhealthrankings.org](http://www.countyhealthrankings.org). They organize the data by counties for each state and includes overall county life expectancy, as well as data for Asian, Black, Hispanic, and White residents in each county . This data is from 2017-2019 and they explain under their data sources, “Data on deaths and births were provided by NCHS and drawn from the National Vital Statistics System (NVSS). These data are submitted to the NVSS by the vital registration systems operated in the jurisdictions legally responsible for registering vital events”. Under its methods explanation, County Health Rankings define this dataset, “Life Expectancy takes into account the number of deaths in a given time period and the average number of people at risk of dying during that period, allowing us to compare data across counties with different population sizes”. Unfortunately, this data is skewed by deaths of the elderly, as well as infant and child mortality. It also tends to overestimate life expectancy for smaller populations. There is also missing data for the majority of counties in each state, and often does not include estimates for all race groups. To remain consistent, I only recorded data for a county if it included life expectancy estimates for both Black and White populations. I used this original list of counties and deleted counties not included in the other data I collected. I removed these missing counties manually, which resulted in some errors like deleting name of county, but not the data which shifted the data. I then corrected this error and ensured it did not occur again.

Next, I collected data for residential segregation rates between Black and White populations in each county. I used the same data source as the life expectancy data, being [www.countyhealthrankings.org](http://www.countyhealthrankings.org). They explain their data measurement on their website, “Racial/ethnic residential segregation refers to the degree to which two or more groups live separately from one another in a geographic area. The index of dissimilarity is a demographic measure of the evenness with which two groups are distributed across the component geographic areas that make up a larger area. The residential segregation index ranges from 0 (complete integration) to 100 (complete segregation)” Their data sources are ACS 5-Year Estimates from 2015-2019 on various social and economic measures. Next, they explain why they included this measure and its effects on health. “Residential segregation remains prevalent in many areas of the country and may influence both personal and community well-being. Residential segregation of Black and White residents is considered a fundamental cause of health disparities in the US and has been linked to poor health outcomes, including mortality, a wide variety of reproductive, infectious, and chronic diseases, and other adverse conditions. Structural racism is also linked to poor-quality housing and disproportionate exposure to environmental toxins. Individuals living in segregated neighborhoods often experience increased violence, reduced educational and employment opportunities, limited access to quality health care and restrictions to upward mobility”. I then recorded this data and removed the same counties as before. The vast majority of counties that had Black and White life expectancies also had a segregation score, with a few exceptions.

Next, I recorded 2010 Census Bureau data of the total counts of residents living in either urban or rural areas and can be found here, <https://data.census.gov/cedsci/>. Unfortunately, there is no 2020 census data or ACS estimates available, so the data is outdated and may not be reflective of 2021 trends. The Census Bureau defines what urban and rural are, “To qualify as an urban area, the territory identified according to criteria must encompass at least 2,500 people, at least 1,500 of which reside outside institutional group quarters. The Census Bureau identifies two types of urban areas: Urbanized Areas (UAs) of 50,000 or more people; Urban Clusters (UCs) of at least 2,500 and less than 50,000 people. “Rural” encompasses all population, housing, and territory not included within an urban area”. I converted the total counts of residents living in urban or rural areas into percentages using Google Sheets (=urban count/total population \* 100). I then removed counties with missing data as before.

For statistical analysis, I will do a Two Sample Difference of Means to find if there is a significant difference between Black and White life expectancy by examining the one tailed p value and rejecting/accepting the null hypothesis. It will tell me the percent chance of making a type 1 error, which is when the null hypothesis is rejected, even when it is true. This will allow me to tell if there is a significant correlation between race and life expectancy and compare this score to other states. I will input all of the results from the TSDOM tests for Kentucky, Indiana, Tennessee, Ohio, Illinois, and Virginia into a summary table.

I also made scatterplots to show the co-variation and correlation between Black or White life expectancy and its segregation score, as well as urban/rural living rates. This will allow me to compare their r squared scores, which indicates the strength of co-variation and correlation and if it is significant or not. I made bar charts that compares overall county life expectancy, white life expectancy, and black life expectancy to serve as a visual aid.

Lastly, I made a map of the difference between Black and White life expectancy to see where White life expectancy is higher than Black life expectancy and vice versa. By adding the FIPS code to my life expectancy data, I was able to do a table join to a Census shapefile of state boundaries and made a choropleth map. I subtracted the difference between White life expectancies and Black life expectancies. If it is a positive value, it means that White life expectancy is higher than Black life expectancy and is assigned the color red. If it is negative, Black life expectancy is higher than White life expectancy and is assigned blue.

**Findings**

Table 1

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Kentucky** | | | | |  | **Indiana** | | | | | |
|  | **Mean** | **Standard Deviation** | | **n** |  | **Mean** | **Standard Deviation** | | | **n** |
| **Black Life Expectancy** | 75.27 | 3.81 | | 23 | **Black Life Expectancy** | 74.79 | 4.59 | | | 24 |
| **White Life Expectancy** | 76.58 | 1.20 | | 23 | **White Life Expectancy** | 76.95 | 1.87 | | | 24 |
|  | | | | |  | | | | | |
| **T-Test Score** | -1.57 | | | | **T-Test Score** | -2.13 | | | | |
| **P-value Two Tail (chance of Type I Error)** | 12.91% | | | | **P-value Two Tail (chance of Type I Error)** | 4.11% | | | | |
| **P-value One Tail (chance of Type I Error)** | 6.45% | | | | **P-value One Tail (chance of Type I Error)** | 2.06% | | | | |
|  | | | | |  | | | | | |
| **Tennessee** | | | | | **Illinois** | | | | | |
| **Black Life Expectancy** | 73.51 | 2.29 | 43 | | **Black Life Expectancy** | 74.57 | | 3.24 | 38 | |
| **White Life Expectancy** | 75.63 | 1.92 | 43 | | **White Life Expectancy** | 78.08 | | 1.93 | 38 | |
|  | | | | |  | | | | | |
| **T-Test Score** | -4.65 | | | | **T-Test Score** | -5.74 | | | | |
| **P-value Two Tail (chance of Type I Error)** | 0.0013% | | | | **P-value Two Tail (chance of Type I Error)** | 0.00003% | | | | |
| **P-value One Tail (chance of Type I Error)** | 0.0006% | | | | **P-value One Tail (chance of Type I Error)** | 0.00002% | | | | |
| **Virginia** | | | | | **Ohio** | | | | | |
| **Black Life Expectancy** | 76.06 | 3.28 | 95 | | **Black Life Expectancy** | 75.12 | | 4.10 | 41 | |
| **White Life Expectancy** | 77.89 | 3.50 | 95 | | **White Life Expectancy** | 77.01 | | 1.70 | 41 | |
|  | | | | |  | | | | | |
| **T-Test Score** | -3.72 | | | | **T-Test Score** | -2.74 | | | | |
| **P-value Two Tail (chance of Type I Error)** | 0.03% | | | | **P-value Two Tail (chance of Type I Error)** | 0.84% | | | | |
| **P-value One Tail (chance of Type I Error)** | 0.01% | | | | **P-value One Tail (chance of Type I Error)** | 0.42% | | | | |
|  | | | | |  |  | | | | | |
| **All Counties** | | | | |
| **Black Life Expectancy** | 75.10 | 3.54 | 265 | |
| **White Life Expectancy** | 77.21 | 2.66 | 265 | |
|  | | | | |
| **T-Test Score** | -7.78 | | | |
| **P-value Two Tail (chance of Type I Error)** | 0.00000% | | | |
| **P-value One Tail (chance of Type I Error)** | 0.00000% | | | |

**Figure 1 (Kentucky).**Chart, bar chart

Description automatically generated

**Figure 2Chart, scatter chart

Description automatically generated**

**Figure 3Chart, scatter chart

Description automatically generated Figure 4 (Indiana)Chart, bar chart

Description automatically generated**

**Figure 5**

**Chart, scatter chart

Description automatically generated**

**Figure 6** **Chart, scatter chart

Description automatically generated**

**Figure 7 (Tennessee)**

Chart

Description automatically generated**Figure 8** Chart, scatter chart

Description automatically generated **Figure 9**Chart, scatter chart

Description automatically generated **Figure 10 (Illinois)Chart, bar chart

Description automatically generated Figure 11Chart, scatter chart

Description automatically generated Figure 12Chart, scatter chart

Description automatically generated Figure 13 (Virginia)Chart

Description automatically generated Figure 14Chart, scatter chart

Description automatically generated Figure 15Chart, scatter chart

Description automatically generated Figure 16 (Ohio)Chart

Description automatically generated Figure 17Chart, scatter chart

Description automatically generated Figure 18Chart, scatter chart

Description automatically generated Figure 19 (All counties)Shape, map

Description automatically generated Figure 20Chart, scatter chart

Description automatically generated Figure 21Chart, scatter chart

Description automatically generated**

For Kentucky, the one tail p-value is 6.45% (Table 1),meaning there a 6.45% chance of making a type 1 error. Because it is more than 5%, I do not reject the null hypothesis that there is no difference between Black and White life expectancy and reject the alternative hypothesis. My alternative hypothesis states that Black life expectancy is lower than White life expectancy in Kentucky

For the scatterplots, there is a weak correlation (0.063%) between urban living rates and segregation score (Figure 2). There is also a weak correlation for Black life expectancy and segregation score (0.048%) (Figure 3).

In the bar chart (Figure 1), major outliers where Black life expectancy is higher than White life expectancy are Campbell, Oldham, and Scott counties. Campbell’s Black population is only 2290/90336, Oldham’s is 2620/60316, and Scott’s is 2468/47173. All of these counties have below 3000 Black residents, so it may be skewed.

Overall, Kentucky is the only state where I was unable to reject the null hypothesis, and I hypothesize this is due to only 23/120 counties having data available. There were also stark outliers which may have also skewed the data.

For Indiana, the one tail p-value is 2.06%, (Table 1) meaning there a 2.06% chance of making a type 1 error. Because it is less than 5%, I reject the null hypothesis that there is no difference between Black and White life expectancy and accept the alternative hypothesis. My alternative hypothesis states that Black life expectancy is lower than White life expectancy in Indiana.

For the scatterplot, there was a weak correlation (0.009) (Figure 5) between urban living rates and Segregation Score, and also a weak correlation for Black life expectancy and segregation score (0.003%) (Figure 6)

In the bar chart (Figure 4), major outliers where Black life expectancy is higher than White life expectancy are Hamilton, Johnson, and Miami counties. Hamilton’s Black population is only 9603/274569, Johnson’s is 1578/139654, and Miami’s is 1673/36903. These low rates skew the life expectancy data.

For Tennessee, the one tail p-value is 0.0006%, meaning there a very low (0.0006%) (Table 1) chance of making a type 1 error. Because it is less than 5%, I reject the null hypothesis that there is no difference between Black and White life expectancy and accept the alternative hypothesis. My alternative hypothesis states that Black life expectancy is lower than White life expectancy in Tennessee

In the scatterplots, there is an overall weak correlation (0.22) (Figure 8) between urban living and segregation score. Overall, it has the strongest correlation of any single state though. There is also a weak correlation for Black life expectancy and segregation score (0.043%) (Figure 9).

In the bar chart (figure 7), the outliers with a higher Black life expectancy are Lake and Montgomery counties. Lake has a population of 2171 Black residents out of 7832 total, and Montgomery’s is 32982/172331. These are higher than the prior states population, but still may have been small enough to skew the data.

For Illinois, the one tail p-value is 0.00002% (Table 1), meaning there a very low (0.00002%) chance of making a type 1 error. Because it is less than 5%, I reject the null hypothesis that there is no difference between Black and White life expectancy and accept the alternative hypothesis. My alternative hypothesis states that Black life expectancy is lower than White life expectancy in Illinois

For the scatterplots, there is again a weak correlation between urban (0.006) living and segregation score (figure 11). There is also a weak correlation for Black life expectancy and segregation score (0.268%) (figure 12), which is the highest score among all states

For the bar chart (figure 10), the major outliers are Pulaski and Williamson counties. Pulaski has a Black population of 1994 out of a total of 6161. Williamson’s Black population is 2540/66357. Pulaski has a relatively high percentage of Black residents, but the overall low population may skew the data. Williamson’s Black population is low, so that also may have skewed the data.

For Virginia, the one tail p-value is 0.01% (table 1), meaning there a low (0.01%) chance of making a type 1 error. Because it is less than 5%, I reject the null hypothesis that there is no difference between Black and White life expectancy and accept the alternative hypothesis. My alternative hypothesis states that Black life expectancy is lower than White life expectancy in Virginia.

For the scatterplots, there is a weak correlation between urban (0.195) (figure 14) living and segregation score. There is also a weak correlation for Black life expectancy and segregation score (0.002%) (figure 15).

In the bar chart (figure 13), the major outliers are Augusta, Pittsylvania, and Westmoreland counties. Augusta’s Black residents come to 2930 out of a total of 73750. Pittsylvania’s is 14018/63506, and Westmoreland’s is 4891/17454. All three counties have low rates of Black residents and could skew the data.

For Ohio, the one tail p-value is 0.42% (table 1) meaning there a low (0.42%) chance of making a type 1 error. Because it is less than 5%, I reject the null hypothesis that there is no difference between Black and White life expectancy and accept the alternative hypothesis. My alternative hypothesis states that Black life expectancy is lower than White life expectancy in Ohio

For the scatterplots, there is a weak correlation between urban (0.006) (figure 17) living and segregation score. There is also a weak correlation for Black life expectancy and segregation score (0.008%) (figure 18).

For the bar chart (figure 16), the major outliers are Fairfield, Scioto, and Union counties. Fairfield has 8702 Black residents out of a total 146156, with Scioto’s being 2129/79499 and Union’s being 1231/52300. All of these counties have rates of Black residents, which may have skewed the data.

When looking at all counties measured, the one tail p-value is 0.00000% (table 1), meaning there a very low (0.00000%) chance of making a type 1 error. Because it is less than 5%, I reject the null hypothesis that there is no difference between Black and White life expectancy and accept the alternative hypothesis. My alternative hypothesis states that Black life expectancy is lower than White life expectancy for all counties where data was available.

With the scatterplots, there were overall weak correlation between urban (0.252) (figure 20) living and segregation score, but it is high compared individual state’s r squared scores. There was very weak correlation for Black life expectancy and segregation score (0.053%) (figure 21) for all counties, which is smaller than what I hypothesized.

In my map of all counties (figure 19), it shows the large amount of missing data for all states that I researched. There are many more counties with higher Black life expectancies than I originally hypothesized. Most of these counties are not major cities though. Urban areas generally have higher White life expectancies.

**Conclusions**

Based on the results of my statistical tests, my hypothesis that Black life expectancy is lower than White life expectancies in most counties was proven right overall. In all states except Kentucky, I was able to reject the null hypothesis that there is no difference between Black and White life expectancies. This was due to their low p-values obtained from the TSDOM test, which was particularly low for all states. I believe Kentucky’s lack of data contributed to this rather than no correlation, as only 23/120 counties had data available. Many of the counties were clustered in central Kentucky, which leaves out western and eastern Kentucky. The county with the largest difference White and Black life expectancies is Jefferson County, Ohio, with 64.2 years for Black residents and 75.5 for White residents. It has a segregation score of 55 and an urban living rate of 61%. On the opposite end, Union County has a Black life expectancy rate of 87.6 compared to 78.7 for White residents. Their segregation score is also 55 and urban living rates are 47%.

My hypothesis was disproven with low correlation rates as shown by the R squared value between urban living rates and segregation scores among all states. I thought there would be a clear positive correlation between increasing urban living rates and segregation scores, but as seen in figure 5, Indiana counties had a slight negative relationship. This is the same outcome for correlation among segregation scores and Black life expectancy, as all R-squared values were low.

This relationship and disparity between Black and White life expectancies is most likely due to structural racism that leads to inequalities in the United States. This inequality, both socially and economically, negatively affects the health and overall life expectancy of Black people as a result of the racist legacies of policies such as redlining and Jim Crow. As the Harvard School of Public Health explain, “For example, blacks have higher rates of diabetes, hypertension, and heart disease than other groups, and black children have a 500% higher death rate from asthma compared with white children. Williams and Lavizzio-Moruey write that geography plays a large role in all of this because, “where we live determines opportunities to access high-quality education, employment, housing, fresh foods or outdoor space – all contributors to our health.”There needs to more research done on the topic of the health effects of poverty, stressful environments, environmental pollution, etc., that are disproportionately affecting Black communities in the United States.

**Citations**

“Census Data Tables.” Explore Census Data, Census Bureau, data.census.gov/cedsci/.

“How Healthy Is Your County?: County Health Rankings.” County Health Rankings &amp; Roadmaps, Population Health Institute, [www.countyhealthrankings.org/](http://www.countyhealthrankings.org/).

“Health Disparities between Blacks and Whites Run Deep.” Harvard, Harvard School of Public Heath, 15 Apr. 2016, [www.hsph.harvard.edu/news/hsph-in-the-news/health-disparities-between-blacks-and-whites-run-deep/](http://www.hsph.harvard.edu/news/hsph-in-the-news/health-disparities-between-blacks-and-whites-run-deep/).